KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

INSTITUTE OF DISTANCE LEARNING

OPTIMAL PIPELINE CONNECTION FOR THE WATER SUPPLY SYSTEM IN THE KWAHU SOUTH DISTRICT OF GHANA

BY

BAMFO ERNEST FREMPONG

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE OF MASTER OF SCIENCE IN INDUSTRIAL MATHEMATICS

APRIL, 2013

DECLARATION

I hereby declare that this submission is my own work towards Master of Science degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made in the text.

BAMFO ERNEST FREMPONG ……………… ………………
(PG 6318611) Signature Date
Student's Name and ID

Certified by:

PROF. SAMUEL K. AMPONSAH ……………….. ……………..
Supervisor Signature Date

Certified by:

MR. KWAKU DARKWAH …….………… ………….……
Head of Department Signature Date

Certified by:

PROF. I. K. DONTWI …………..…… ………..………
Dean, IDL Signature Date

DEDICATION

To The Glory of God

This work is dedicated to the following personalities: my mother Gladys Mirekua, my friends Anim Rexford Wiafe, an Environmental Consultant and Kissi Samuel Scanty, Computer Networking Specialist, for their support.

ACKNOWLEDGEMENT

I am most grateful to the Almighty God for guiding and sustaining me throughout this course.

His everlasting love and Grace has made me see the light of day and subsequent completion of this work.

My profound gratitude to my supervisor, Professor S. K. Amponsah, whose guidance has enabled me to produce this work.

Special thanks to my parents and my siblings, for their prayers and support throughout my education.

While I share the credit of this Masters' thesis with all the above mentioned people, responsibility for any errors, shortcomings or omissions in this thesis is solely mine.

ABSTRACT

Water supply is becoming more than a challenge with increasing demand due to the increase in population. Various interventionist approaches have been adopted including boreholes. In this thesis, a mathematical model for finding a minimum spanning tree length was used to determine an optimal pipeline connection for the water supply system in the Kwahu South District of Ghana. The study found out the possible implications of the minimum spanning tree length on the cost of water supply in the Kwahu South District. The distances between the nodal towns were thus determined to estimate the lengths of the pipelines required. The data was arranged in a matrix form and the Prim's algorithm applied to the data. The study revealed that the total minimum length of pipelines required to supply the captured nodal towns is fifty-two (52) kilometres. The study also showed that with this minimum spanning tree length, the total cost of procuring pipelines would be significantly reduced. This consequently will make resources available for other productive ventures in the water treatment sector.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AB – Abetifi

ACM - Association for Computing Machinery

ACO - Ant Colony Optimization

AD - Aduamoa

AS - Asakraka

AT – Atibie

BE - Bepong

BL - Bethe Lattice

BSP - Bulk Synchronous Parallel

CMST – Cost Minimum Spanning Tree

CPU – Central Processing Unit

DCMST - Diameter - Constrained Minimum Spanning Tree

EMST - Euclidean Minimum Spanning Tree

GLA - Generalized Linear Approximation

GMDMCST – Geometric Minimum Diameter Minimum Cost Spanning Tree

GMST - Generalized Minimum Spanning Tree

H – MCOP – Heuristic Multiple Constrained Optimization Problem

KHMST - K – Hop Minimum Spanning Tree

KO – Kotoso

MATLAB – Matrix Laboratory

MDMCST - Minimum Diameter Minimum Cost Spanning Tree

MRMCST - Minimum Radius Minimum Cost Spanning Tree

MST - Minimum Spanning Tree

MSTEAM - Minimum Spanning Tree Based Energy Aware Multicast Protocol

MRCST - Minimum Routing Cost Spanning Tree

MP – Mpraeso

MSCT - Minimum Spanning Clustering Tree

NK – Nkwatia

NP-hard - Non-Deterministic Polynomial-Time hard

OB – Obo

OG – Obomeng

PE – Pepease

PERT - Program Evaluation and Review Technique

QoS - Quality – of – Service

SMT - Steiner Minimal Tree

TA - Tafo

TI – Texas Instruments

VLSI/CAD - Very-Large-Scale Integration/Computer-Aided Design

WAGP - West African Gas Pipeline

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION

Human lives hugely depend on clean water. The human body is composed of at least seventy percent of water. Water supply to our domestic settings and even our enterprises is not in the best of shapes. It is the quest for clean and treated water that is high on the agenda of the water supply agencies. Their workings toward this desire have never been a smooth run. The pipelines required to be laid to carry treated water to their destination towns from the treatment site are not readily available due to the high cost associated with their acquisition. It is in this light that we require the optimal pipeline connections to be able to supply treated water in the Kwahu South District of Ghana. In this chapter, we considered the background, problem statement and objectives of the study. The methodology, significance and limitations of the study were also put forward.

1.1 BACKGROUND TO THE STUDY

Water is a chemical substance with the chemical formula H_2O. Water molecule contains one oxygen atom and two hydrogen atoms connected by covalent bonds. Water is a liquid at temperature above 0^0C (273.15K or 32^0F) at sea level, but it often co-exists on earth with its solid state, ice, and gaseous state (water vapour or steam).Water also exists in a liquid crystal state near hydrophilic surfaces. Water covers 71% of the earth`s surface and is vital for all known forms of life (www.wikipedia.com).

On earth, 96.5% of the planet's water is found in oceans, 1.7% in ground water, 1.7% in glaciers and the ice caps of Antarctica and Greenland, a small fraction in other large water bodies and 0.001% in the air as vapour, clouds (formed of solid and liquid water particles suspended in air) and precipitation. Only 2.5% of the earth's water is fresh water and 98.8% of that water is in ice and ground water. Less than 0.3% of all fresh water is in rivers, lakes and the atmosphere, and an even smaller amount of the earth's freshwater (0.003%) is contained within biological bodies and manufactured products. Water, on earth, moves continually through the hydrological cycle of evaporation and transpiration (evapotranspiration), condensation precipitation and runoff, usually reaching the sea. Evaporation and transpiration contribute to the precipitation over land (www.wikipedia.com).

Safe drinking water is essential to humans and other life forms. Access to safe drinking water has improved over the last decades in almost every part of the world, but approximately one billion people still lack access to safe water and over 2.5 billion lack access to adequate sanitation (www.wikipedia.com). There is clear correlation between access to safe water and gross domestic product per capita however; some observers have estimated that by 2025, more than half of the world population will be facing water – based vulnerability. Water plays a very critical role in human lives. All known forms of life depend on water. It is vital both as a solvent in which many of the body's solutes dissolve and as an essential part of many metabolic processes within the body.

Water is fundamental to photosynthesis and respiration. Water is also central to acid-based neutrality and enzymes function. Nearly all fish live exclusively in water and there are many types of marine mammals such as dolphins, whales and amphibians spend portions of their lives in water and portions on land. Plants such as kelp and algae grow in the water and are the basis for some underwater ecosystems. Plankton is generally the foundation of the ocean food chain (www.wikipedia.com).

Civilization has historically flourished around major rivers and major waterways; Mesopotarnia the so-called cradle of civilization, was situated between the major rivers Tigris and Euphrates; the ancient society of the Egyptians depended entirely upon the Nile. Large metropolises like Rotterdam, London, Tokyo, Paris, New York City, Buenos Aires, Shanghai, Chicago, and Hong Kong owe their success in part to their easy accessibility via water and the resultant expansion of trade. Islands with safe water ports like Singapore, have flourished for the same reason. In places such as North Africa and the Middle East, where water is scarcer, access to clean drinking water was, and is still a major factor in human development (www.wikipedia.com).

The most important use of water in agriculture is for irrigation which is a key component to produce enough food. Irrigation takes up to 90% of water withdrawn in some developing countries and significant proportions more economically developed countries (United States, 30% of fresh water usage is for irrigation). It takes around 3000 litres of water, converted from liquid to vapour, to produce enough food to satisfy one person’s daily dietary need. This is a considerable amount, when compared to that required for drinking, which is between two and five litres. To produce food for the over 7 billion or so people who inhabit the planet today requires the water that would fill a canal ten metres deep, 100 metres wide and 7.1 million kilometres long-that’s enough to circle the globe 180 times (www.wikipedia.com).

Water is used as a scientific standard. On 7th April 1795, the gram was defined in France to be equal to the ‘absolute weight of a volume of pure water equal to a cube of one hundredth of metre, and to the temperature of the melting ice (www.wikipedia.com).

The Kelvin temperature scale of the SI system is based on the triple point of water, defined as exactly 273.16k or 0.01^0C.

The human body contains from 55% to 78% of water, depending on body size. The body requires between one and seven litres of water per day to avoid dehydration; the precise amount depends on the level of activity, temperature, humidity, and other factors. Most of this is ingested through food or beverages other than drinking straight water. It is not clear how much water intake is needed by healthy people, though most advocates agree that approximately 2 litres (6-7 glasses) of water daily is the minimum to maintain proper hydration. Medical literature favours a lower consumption, typically 1 litre of water for average male, excluding extra requirements due to fluid loss from exercise or warm weather (www.wikipedia.com).

For those who have healthy kidneys, it is rather difficult to drink too much water, but (especially in warm humid weather and while exercising) it is dangerous to drink too little. People can drink far more water than necessary while exercising, however, putting them at risk of water intoxication (hyper hydration), which can be fatal. The popular claim that ’a person should consume eight glasses of water per day’ seems to have no real basis in science. Similar misconceptions concerning the effect of water on weight loss and constipation have also been dispelled. The single largest (by volume) freshwater resource suitable for drinking is Lake Baikal in Siberia (www.wikipedia.com).

The propensity of water to form solutions and emulsions is useful in various washing processes. Washing is also an important component of several aspects of personal body hygiene.

The use of water for transportation of materials through rivers and canals as well as the international shipping lanes is an important part of the world economy.

Water is widely used in chemical reactions as a solvent or reactant and less commonly used as a solute or catalyst.

Water and steam are used as heat transfer fluids in diverse heat exchange systems; due to its availability and high heat capacity, both as a coolant and for heating. In the nuclear power industry, water can be used as a neutron moderator. In most nuclear reactors, water is both a coolant and a moderator. Water is used for fighting wildfires.

Humans use water for many recreational purposes, as well as for exercising and for sports, including swimming, water skiing, boating, surfing and diving. Lakesides, beaches and water parks are popular places for people to go and relax and enjoy recreation. People keep fish and other life in aquariums or ponds for show, fun and companionship (www.wikipedia.com).

Water is used in power generation. Hydroelectricity is electricity obtained from hydropower. Hydroelectric power comes from water driving a water turbine connected to a generator. Typically, a dam is constructed on a river, creating an artificial lake behind it. Water can be used to cook foods such as noodles and rice. Water hardness is a critical factor in food processing. It can dramatically affect the quality of a product as well as playing a role in sanitation. Water hardness is measured in grains; 0.064g calcium carbonate is equivalent to one grain of hardness. Water is classified as soft if it contains 1 to 4 grains, medium if it contains 5 to 10 grains and hard if it contains 11 to 20 grains. The hardness of water also affects its pH balance which plays a critical role in food processing. For example, hard water prevents successful production of clear beverages. Water hardness also affects sanitation; with increasing hardness, there is a loss of effectiveness for its use as a sanitizer. Water is used for dishwashing (www.wikipedia.com).

Water is also used for religious and philosophical purposes. Water also has a use related to literature. Water is considered a purifier in most religions and also it is a symbol of purification (www.wikipedia.com).

1.2 STATEMENT OF THE PROBLEM

There are many sources of water in the Kwahu South District of Ghana including bore holes, rivers and stagnant waters and treated tap water. Bore holes and rivers are the commonest sources of water for human consumption. Rivers and stagnant waters are the most common sources of water for animal consumption. These sources have rivers as the leading source of water for both human and animal consumption though not treated posing a great deal of threat to both human and animal health. This is the main reason why treated tap water is the preferred choice. The supply of treated water is highly erratic and not in the best of shapes. All the aforementioned sources of water serve sections of the populace but quality is of the essence. Tap water is widely favoured, though proving to be a mirage in the Kwahu South District. Since each person cherishes his or her good state of health, the question then arises as to whether we shall have access to treated pipe-borne water even when we are ready to bear the cost of it to make sure water flows in our taps. It is obviously against this backdrop that this research is being undertaken to determine the optimal pipeline connection for the treated water supply system, with the Kwahu South District as the reference district.

1.3 OBJECTIVES OF THE STUDY

The objectives of the study are:

(i) to determine the optimal pipeline connections for the water supply system in the Kwahu South District of Ghana using Prim's algorithm and based on the road network model.

(ii) to determine possible implications for water supply, of the resulting optimal connection.

1.4 METHODOLOGY

Data for the research was obtained from the District Assembly Offices at Mpraeso and the Ghana water company office at Mpraeso and also through internet searches. The information obtained concerned distances of the road network from one town to another, in the district. The towns are the nodes. The pipelines are mostly laid along the road networks. In essence, the length of a road network fits the length of a pipeline required. Physical maps of the district indicating distances between the towns, MATLAB software and TI calculators aided computational efforts for highly accurate results for the Minimum Spanning Tree. The Prim's and Kruskal's algorithms were implemented.

1.5 SIGNIFICANCE OF THE STUDY

In the quest to supply treated water to the various towns and subsequently to the individual homes, there is the urgent need to minimize operational cost, which has the cost of laying pipelines as a major component .The study will help the water supply authorities to procure only the optimal amount of pipelines thereby reducing waste to the barest minimum. The study will also make costing of water supply in the district much easier depending on the pipeline required in such a project. It is worthy of note that this research is a significant cost cutting exercise, which will inure to the benefit of the state. With clean treated water supplied, quality health is guaranteed.

1.6 LIMITATIONS OF THE STUDY

The entire Kwahu South District is under study but not every inhabited location would be captured in study since the small localities obtain their clean water under a different scheme. They are mostly bailed out by the Community Water and Sanitation Agency with bore-holes.

Again, though the data obtained are very credible, field trips to the various towns for verification of some of the distance of the road networks would have been in good taste but financial constraint was a major setback.

1.7 ORGANIZATION OF THE STUDY

The thesis is in five chapters. The first chapter covers the introductory part of the study under which the background of the study, emphasizing water and its importance, a brief on the Kwahu South District, statement of the problem and objectives of the study, methodology to be used in the study and the significance of the study. This chapter also includes the limitations of the study and the organization of the study. The second chapter covers the review of related literature. The focus is on work done by other researchers. The third chapter covers the methodology, the mathematical model for solving the shortest paths using the Prim's algorithm. The fourth chapter covers discussion of results; the focus would be on results obtained. The fifth and final chapter shall focus on the summary of the findings, conclusions and recommendations.

CHAPTER 2

LITERATURE REVIEW

2.0 INTRODUCTION

A wide variety of research works have been undertaken on shortest paths and water supply in Ghana and the world, where clean water for domestic use and water for industrial use is of grave concern. The research works include areas such as the ‘Optimal Pipeline Connection for The West Africa Gas Pipeline Project in Ghana and other West African countries’ whiles others delved into the Minimum Spanning Tree Route for Tourist Centres in the Brong Ahafo Region of Ghana. This research shall focus entirely on the optimal pipeline connection for the water supply system in the Kwahu South District of Ghana.

2.1 REVIEW OF RELATED WORKS

Agarwal et al., (1998) considered the parametric minimum spanning tree problem, in which a graph with edge weights that are linear functions of a parameter λ was given. The authors computed the sequence of minimum spanning trees generated as λ varies. The kinetic minimum spanning tree problem, in which λ represented time and the graph was subject in addition to changes such as edge insertions, deletions, and modifications of the weight functions as time progressed. The authors solved the problems in time O ($n^{2/3}$ log $^{4/3}$ n) per combinatorial change in the tree (or randomized O ($n^{2/3}$ log n) per change). The time bounds reduced to O($n^{2/3}$ log $^{3/2}$ n) per change (O (n $^{1/2}$ log n) randomized) for planar graphs or other minor –closed families of graphs, and O (n $^{1/4}$ log $^{3/2}$ n) per change (O(n $^{1/4}$ log n) randomized) for planar graphs with weight changes but no insertions or deletions.

Quoc (2012) proposed that Minimum Routing Cost Spanning Tree (MRCST) is one of the spanning tree optimization problems having many applications in network design and in general cases, the problem was proved as NP – hard. The author proposed two heuristic-based algorithms used for MRCST, with the general idea of those algorithms being to start at a spanning tree and step by step improve edges to obtain better spanning tree. The author conducted experiment implementations for those proposals and obtained better result than the result of current approximate algorithms.

Golden et al., (2004) proposed that the Heuristic Search for the Generalized Minimum Spanning Tree (GMST) problem occurred in telecommunications network planning, where a network of node clusters needs to be connected via a tree architecture using exactly one node per cluster. The authors proposed that the problem was NP-hard. The authors presented two heuristic search approaches for the GMST problem: local search and genetic algorithms. The authors' computational experiments showed that these heuristics rapidly provide high quality solutions for the GMST and outperformed some previously suggested heuristics for the problem. In the computational tests on 211 test problems, their local - search heuristic found the optimal solution in 179 instances and their genetic algorithm procedure found the optimal solution in 185 instances (out of the 211 instances, the optimal solution was known in 187 instances).

Rico (2011) considered the minimum spanning tree problem under the restriction that for every vertex, v, the edges of the tree that are adjacent to v satisfy a given family of constraints. A famous example thereof was the classical degree - bounded Minimum

Spanning Tree problem, where for every vertex, v, a simple upper bound on the degree was imposed. Iterative rounding/relaxation algorithms became the tool of choice for degree - constrained network design problems. A cornerstone for that development was the work of Singh and Lau; who showed that the degree-bounded Minimum Spanning Tree problem; one could find a spanning tree violating each degree bound by at most one unit and with cost at most the cost of an optimal solution that respected the degree bounds. Previous iterative rounding approaches were not suited for degree constraints where some edges are in a super-constant number of constraints. The author presented an algorithm for the degree-constrained minimum spanning tree problem where for every vertex, v, the edges adjacent to v have to be independent in a given matroid. The algorithm returned a Spanning Tree of optimum cost, such that for every, v, it suffices to remove at most 8 edges from the spanning tree to satisfy the matroidal degree constraint at v.

Jackson and Read, (2009) took interest in the theory of minimum spanning trees, with emphasis on mean-field theory and strongly disordered spin-glass model. The authors considered the random Minimum Spanning Tree, in which the edge costs were (quenched) independent random variables. There is a strongly – disordered spin- glass model due to Newman and Stein (Phys. Rev. Lett.72, 2286(1994)}, which maps precisely onto the random Minimum Spanning Tree. The authors studied scaling properties of random Minimum Spanning Trees using a relation between Kruskal’s greedy algorithm for finding the Minimum Spanning Tree, and bond percolation. The authors solved the random minimum spanning tree problem on the Bethe Lattice (BL) with appropriate wired boundary conditions and calculated the fractal dimension D = 6 of the connected components. Viewed as a mean – field theory, the result implied that on a lattice in Euclidean space of dimension, d, there were of order W^ (d-D) large connected components of the random MST inside a window of size W, and that d = d – c = D = 6 was a critical dimension. That differed from the value eight (8) suggested by Newman and Stein. The authors also critiqued the original argument for eight (8), and provided an improved scaling argument that again yielded d – c = 6. The result implied that the strongly-disordered spin-glass model had many ground states for d > 6, and only of order one below six. The results for Minimum Spanning Trees also apply on the Poisson-weighted infinite tree, which is a mean-field approach to the continuum model of Minimum Spanning Trees in Euclidean space, and is a limit of the BL.

Cheong and Changryeol, (2012) took a dive into the Single - Source Dilation- Bounded Minimum Spanning Trees problem and it was noted that given a set S of points in the plane, a geometric network for S was a graph with vertex set S and straight edges. The authors considered a broad casting situation, where one point r in S was a designated source. Given a dilation factor δ, they sought for a geometric network G such that for every point v in S, there was a path from r to v in G of length at most δ |rv|, and such that the total edge length was minimized. The authors showed that finding such a network of minimum total edge length was NP-hard, and gave an approximation algorithm.

Assad and Wu, (1992) considered the Quadratic Minimum Spanning Tree problem that introduced a new optimization problem that involved searching for the spanning tree of minimum cost under a quadratic cost structure. That was proven to be NP- hard. A technique for generating lower bounds for this problem was discussed and incorporated into branch-and –bound schemes for obtaining exact solutions. Two heuristic algorithms were also developed. Computational experience with both exact and heuristic algorithms was reported.

Dae et al., (2009) took interest in the Geometric Minimum Diameter Minimum Cost Spanning Tree Problem. The authors considered bi-criteria geometric optimization problems, in particular, the minimum diameter minimum cost spanning tree problem for a set of points in the plane. The graph-theoretic Minimum Diameter Minimum Cost Spanning Tree (MDMCST) problem and the Minimum Radius Minimum Cost Spanning Tree (MRMCST) problem were shown to be NP-hard. The authors showed that the geometric version of those two problems, GMDMCST and GMRMCST problems were also NP-hard. The authors also gave two heuristic algorithms, one MCST-based and the other MDST-based for the GMDMCST problem and presented some experimental results.

Sörensen and Janssens, (2005) presented that Minimum Spanning Tree of an undirected graph could be easily obtained using classical algorithms by Prim or Kruskal. Good time and space complexities are the major concerns of those algorithms. Most algorithms generate spanning trees using some fundamental cut or circuit. In the generation process, the cost of the tree was not taken into consideration. The paper presented an algorithm to generate spanning trees of a graph in order of increasing cost. In that way, it was possible to determine the second smallest or, in general, the k-th smallest spanning tree. The authors' algorithm was based on an algorithm by Murty (1967), which enumerated all solutions of an assignment problem in order of increasing cost.

Claus (2008) described the optimal minimum spanning tree algorithm given by Pettie and Ramachandran (in Journal of the ACM, 2002). The algorithm presented found a minimum spanning tree of a graph with n vertices and m edges deterministically in time O ($T^*(m, n)$, where T^* is the minimum number of edge –weight comparisons needed to determine the solution of the problem. Even though the algorithm ran in optimal time, the exact function describing the running time was not known and was still an open problem. A trivial lower bound is $T^*(m, n) = \Omega (m)$. A deterministic upper bound is $T^* (m, n) = O (m.\alpha (m, n)$, due to Chazelle (in Journal of the ACM, 2000). Here α is the very slow growing inverse of the Ackermann function. The optimal algorithm used hardwired optimal decision trees for very small graphs compared to the input graph, but for input graph instances with practical size, the decision trees were not required to obtain the optimal running time. Because the optimal algorithm was relatively advanced, the hidden Big - Oh constant of its running time was relatively large. The results of the experiments showed that some of those algorithms were significantly faster than the optimal algorithm for practical input graph instances. The overall experiment winner was the algorithm originally discovered by Jarnik, and later by Prim and Dijkstra. However, for a special graph family the optimal algorithm was faster than the earliest MST algorithm known, namely Boruvka's algorithm.

Chi-Chih (1982) discussed the problem of finding a minimum spanning tree connecting n points in a k-dimensional space under three common distance metrics- Euclidean, Rectilinear and L∞. The author employed a subroutine that solved the post office problem, they showed that, for fixed $k \geq 3$, such a minimum spanning tree could be found in time O $[n^{2-a\,(k)}\ (\log n)^{1-a\,(k)}]$, where $a\,(k) = 2^{-(k+1)}$. The bound could be improved to O $((\log n)^{1.8})$ for points in the 3-dimensional Euclidean space. The author also obtained O (n^2) algorithms for finding a farthest pair in a set of n points and for other related problems.

Althaus et al., (2004) presented that given a complete graph on n nodes with metric edge costs, the minimum – cost K – Hop Minimum Spanning Tree (KHMST) asked for a spanning tree of minimum total cost such that the longest root – leaf path in the tree had at most k edges. The authors presented an algorithm that computed such a tree of total expected cost O (logn) times that of a minimum – cost k – hop spanning tree.

Martin and Gunther, (2009) posited that the bounded diameter minimum spanning tree problem was an NP – hard combinatorial optimization problem arising in particular in network design. The authors presented a method based on hierarchical clustering to guide the construction process of a diameter constrained tree. Solutions obtained were further refined using a greedy randomized adaptive search procedure, especially on large Euclidean instances with a tight diameter bound, the results were excellent. In that case the solution quality could also compete with that of a leading Meta heuristic.

Ghodsi et al., (2007) found a spanning tree of G where the maximum weighted degree of its vertices, is minimum, given a metric graph G. In a metric graph (or its spanning tree), the weighted degree of a vertex was defined as the sum of the weights of its incident edges. The authors proposed a 4.5 – approximation algorithm for this problem. The authors also proved it was NP – hard to approximate this problem within a 2 – ε factor.

Zhou (2003) postulated that Steiner Minimal Tree (SMT) problem was a very important problem in VLSI CAD. Given $n$points in a plane, a Steiner minimal tree connects these points through some extra points (called Steiner points) to achieve a minimal total length. The author recorded an implemental of an efficient Steiner minimal tree algorithm that has a worst case running time of O (nlogn) and a similar performance as the iterated 1-steiner algorithm. The algorithm efficiently combines Borah et al’s edge substitute concept with Zhou et al’s spanning graph.

Holm et al., (1998) assumed no edges in a graph with n vertices. The amortized operation costs are O ($\log^5$ n) for connectivity, O ($\log^4$n) for Minimum Spanning forest, 2-edge connectivity, and O ($\log^5$n) for biconnectivity.

Peter et al., (2010) detected clusters with irregular boundaries. In the paper, the authors proposed MST based clustering algorithm. The algorithm produced k clusters with Minimum Spanning Clustering Tree (MSCT), a new data structure which could be used as search tree. The algorithm worked in two phases. The first produced a sub tree (cluster) and the second converted the sub tree into binary tree called MSCT.

Katajainen et al., (1982) proposed an application of the bucket sort in Kruskal's minimal spanning tree algorithm. The modified algorithm was very fast if the edge costs were from a distribution which is close to uniform, due to the fact that the sorting phase then takes for an m edge graph an O (m) average time. The O (mlogm) worst case occurred when there was a strong peak in the distribution of the edge costs.

Gustavo et al., (1982) posited that the Boruvka's algorithm which computes the minimum cost spanning tree was used to define a rule to share the cost among the nodes (agents). The authors showed that the rule coincided with the folk solution, a very well-known rule of that literature.

Welch et al., (1988) conceded that highly-optimized concurrent algorithms were often hard to prove correct because they had no natural decomposition into separately provable parts. The authors presented a proof technique for the modular verification of such non-modular algorithms. It generalized existing verification techniques based on a totally – ordered hierarchy - that is a lattice of different views of the algorithm. The technique was applied to the well-known distributed minimum spanning tree algorithm of Gallager, Humblet and Spira, which had until recently lacked a rigorous proof.

Deo et al., (2000) postulated that a Minimum Spanning Tree with a small diameter was required in numerous practical situations as in distributed mutual exclusion algorithms in order to minimize the number of messages communicated among processors per critical section. The authors stated the diameter - constrained Minimum Spanning Tree

(DCMST) problem as; given an undirected edge-weighted graph G with n nodes and a positive integer k, find a spanning tree with the smallest weight among all spanning trees of G which contain no path with more than k edges. This problem was known to be NP-complete, for all values of k; $4 \leq k \leq (n-2)$.Therefore, one has to depend on heuristics and live with approximate solutions. The authors explored two heuristics for the DCMST problem: first, they presented a one-time-tree-construction algorithm that constructed a DCMST in a modified greedy fashion, employing a heuristic for selecting edges to be added to the tree at each stage of the tree construction. This algorithm was fast and easily parallelizable and particularly suited when the specified values of k are small-independent of n. The second algorithm started with an unconstrained Minimum Spanning Tree and iteratively refined it by replacing edges, one by one, in long paths until there was no path left with more than k edges .This heuristic was found to be better suited for larger values of k. The authors discussed convergence, relative merits and parallel implementation of these heuristics on the MasPar MP-1 – a massively parallel SIMD machine with 8192 processors. The authors' extensive empirical study showed that the two heuristics produce good solutions for a wide variety of inputs.

Achuttan and Caccetta, (1992) assumed G to be a simple graph with non-negative edge weights. The problem arises in network design and as a sub problem in many combinatorial optimization problems such as vehicle routing. The authors postulated that in some applications, it was necessary to restrict the diameter of the spanning tree and thus one is interested in the problem: Find, in a given weighted graph G, a minimum weight spanning tree of diameter at most D. The problem was known to be NP- complete for D ≥ 4. The paper presented a mixed integer linear programming formulation.

Frey et al., (2007) presented a Minimum Spanning Tree based energy aware multicast protocol (MSTEAM), which was a localized geographic multicast routing scheme designed for ad hoc and sensor networks which used locally-built minimum spanning tree as an efficient approximation of the optimal multicasting backbone. The authors said that using minimum spanning tree is highly relevant in the context of dynamic wireless networks since its computation had a low time complexity O (nlogn). However, their protocol was fully localized and required nodes to gather information only on 1- hop neighbours, which is a common assumption in existing work. The authors provided a theoretical analysis proving that MSTEAM was loop-free and always achieved delivery of the multicast message, as long as a path existed between the source node and the destinations. The authors' experimental results demonstrated that MSTEAM was highly energy-efficient, outperformed the best existing localized multicast scheme and was almost as efficient as a centralized scheme in high densities.

Golden et al., (2006) introduced the prize-collecting generalized minimum spanning tree problem. In that problem a network of node clusters needed to be connected via a tree architecture using exactly one node per cluster. Nodes in each cluster compete by offering a payment for selection. The problem was NP-hard, and the authors described several heuristic strategies, including local search and a genetic algorithm. Further, the authors presented a simple and computationally efficient branch – and – cut algorithm, which found optimal solutions for networks with up to two hundred nodes within two hours of CPU time, while the heuristic search procedures rapidly found near-optimal solutions for all of the test instances.

Vikas (2010) discussed an algorithm for minimum spanning tree apart from the traditional Kruskal’s and Prim’s algorithms. That was the Vicky algorithm. Initially, the algorithm formed a forest and then converted the forest into the minimum spanning tree.

Takahashi et al., (1979) considered an O (kn^2) time algorithm finding an approximate solution for the Steiner Problem in Graphs, where n was the number of vertices in a given graph and k was the number of vertices that must be connected. The worst case cost-ratio of the obtained solution to the optimal solution was tightly 2(1-1/k).

Pettie et al., (2002) found that for many fundamental problems there exist randomized algorithms that are asymptotically optimal and are superior to the best known deterministic algorithm. The authors developed some general methods for reducing exponentially the consumption of random bits in comparison based algorithms. In some cases the methods were able to reduce the number of random bits from linear to nearly constant without affecting the expected running time. Most of the results of the authors were obtained by adjusting existing randomized algorithms to work well with a pair wise or O (1) – wise independent sampler. The prominent exception - and the main focus of that paper – was a linear – time randomized Minimum Spanning Tree algorithm that was not derived from the well known Karger – Klein – Tarjan algorithm. With the authors’ algorithm as a guide, they presented a unified view of the existing ‘non - greedy’ minimum spanning tree algorithms. Concepts from the Karger – Klein – Tarjan algorithm, such as F – lightness, minimum spanning tree verification and sampled graphs, are related to the concepts of edge corruption, sub graph contractibility, and soft heaps, which were the basis of the deterministic Minimum Spanning Tree algorithms of Chazelle and Pettie – Ramachandran.

Feltkamp et al., (1994) presented two cost allocation rules. The authors used the proportional and decentralized rules. The authors proved that those rules were refinements of the irreducible core, as defined in Feltkamp, Tijs and Muto (1994b). The authors axiomatically characterized the proportional rule.

Adler et al., (1999) presented lower and upper bounds for finding a minimum spanning tree in a weighted undirected graph on the BSP model. Let p denote the number of processors, n be number of nodes of the input graph, and m the number of edges of the input graph. The authors showed that in the worst case a total of Ω (k.min (m, pn)) bits needed to be transmitted in order to solve the minimum spanning tree problem, where k is the number of bits required to represent a single edge weight. This implied that that if a message could contain at most O (k) bits, any BSP algorithm for finding a minimum spanning tree required communication time Ω (g.min (m/p, n)), where g is the gap parameter of the BSP model. The authors presented two algorithms, the first of which could employ at most m/n processors efficiently and the second was a randomized algorithm that performed linear work with high probability, provided that m ≥ nlogp.

That was the first linear work BSP algorithm for finding a minimum spanning tree in sparse graphs.

Ben – Zwi (2006) stated that the Euclidean Minimum Spanning Tree problem was to decide whether a given graph G = (P, E) on a set of points in two dimensional plane was a minimum spanning tree with respect to the Euclidean distance. The author showed that every non – adaptive (not necessarily 1 – sided - error) property – tester for that task had a query complexity of Ω ($\sqrt{n}$). The author further proved that every adaptive property – tester had query complexity of Ω ($n^{1/3}$). Those lower bounds held even when the input graph was promised to be a bounded degree tree.

March et al., (2010) proposed a new, fast, general EMST algorithm which was motivated by the clustering and analysis of astronomical data. Large – scale astronomical surveys, including the Sloan Digital Sky Survey, and large simulations of the early universe, such as the Millennium Simulation, could contain millions of points and fill terabytes of storage. The authors stated that traditional EMST methods scale quadratically, and more advanced methods lacked rigorous runtime guarantees. The authors presented a new dual – tree algorithm for efficiently computing the EMST, used adaptive algorithm analysis to prove the tightest (and possibly optimal) runtime bound for the EMST problem to – date, and demonstrated the scalability of their method on astronomical data sets.

Patterson et al., (2000) used Cost Minimum Spanning Tree (CMST) problem to develop and demonstrate a hybrid neural network methodology that incorporated heuristic methods into the neural network topological design. The heuristic procedure was embedded into the neural network topological design, and an iterative improvement process was performed using the neural network. The semi – relaxed energy function of the problem was used to develop a neural network weight adjustment procedure that modified the problem cost. In 75% of their experiments, the hybrid neural networks produced better results than any of the traditional procedures tested.

Raj et al., (2012) found that swarm intelligence has the capability to recover path with minimum complexity. It only needed small amount of some special purpose information aside from probability information; they could have the value of alpha equal to zero and beta equal to one. So pheromone distribution although was a key feature for discovering minimum path by ants. As there could be cases in which there was a MST with only single edge originating from a vertex, an ant got its final position at the last vertex and in that case from the idea of ‘ant colony optimization’, they could reach Prim’s but with limitation and hence the name that ACO is limited case of Prim’s algorithm.

Gorad et al., (2007) presented an application which found the minimum and maximum spanning trees of a connected weighted system of systems based on a new method. The integral part of the new method was a model to define combined weights of a pair of nodes and a link between them. The minimum and maximum spanning trees could represent the upper or lower operational boundaries of a connected weighted system of systems, depending on categories and measures for each category. The authors demonstrated the proposed method using nine – node cluster of a weighted system of systems.

Gloor et al., (1993) described a system for visualizing correctness proofs of graph algorithms. The system was demonstrated for Prim's algorithm for finding a minimum spanning tree of an undirected, weighted graph. While an example was nota proof, their system provided concrete examples to illustrate the operation of the algorithm. Those examples could be referred to by the user interactively and alternatively with the visualization of the proof where the general case was portrayed abstractly.

Korkmaz et al., (2001) proposed that providing quality – of – service (QoS) guarantees in packet networks gave rise to several challenging issues including how to determine a feasible path that satisfied a set of constraints while maintaining high utilization of network resources. The latter objective implied the need to impose an additional optimality requirement on the feasibility problem, done through a primary cost function (e.g. administrative weight, hop - count) according to which the selected feasible path is optimal. In general, multi – constrained path selection, with or without optimization, is an NP – complete problem that could not be exactly solved in polynomial time. Those problems were dealt with using heuristics and approximation algorithms with polynomial - and pseudo – polynomial – time complexities. The authors proposed an efficient heuristic algorithm for the most general form of that problem. They first showed that multiple constraints could be dealt with using a non – linear cost function whose minimization provided a continuous spectrum of solutions ranging a generalized linear approximation (GLA) to an asymptotically exact solution. The authors then introduced their heuristic algorithm (H - MCOP), which attempted to minimize both the non – linear cost function (for the feasibility part) and the primary cost function (for the optimality part). The authors proved that H – MCOP guaranteed at least the performance of GLA and often improved upon it. H – MCOP had the same order of complexity as Dijkstra's algorithm. The authors used extensive simulations on random graphs with correlated and uncorrelated link weights to show that under the same level of computational complexity, H – MCOP outperformed its contenders in its success rate in finding feasibility and in the cost of such paths.

Resende et al., (2003) suggested that computational algorithms for the solution of network flow problems were of great practical significance. In the last decade, a new class of computationally efficient algorithms, based on the interior point method, has been proposed and applied to solve large scale network flow problems. The authors reviewed interior point approaches for network flows, with emphasis on computational issues.

Donkor et al., (2010) combined both Prim's and Steiner Tree algorithms with factor rating method to solve the single source shortest path offshore/onshore pipeline problem. Data on the West African Gas Pipeline (WAGP) project was collected and analyzed. The authors used Prim's algorithm to find the minimum spanning tree of length 712.30km, a reduction over the original 788.90km WAGP project design. The authors also used factor rating method to find an alternative path of length 723.29km. The authors again used Steiner Tree algorithm and geometry to obtain an optimal pipeline length of 707.75km, a 10.3% reduction of the WAGP length. The authors' solution was shown to be topologically equivalent to the WAGP network and hence optimal in pipeline distance and project cost.

Subadra et al., (2011) considered graph theory applications to find out the path of a real world problem and the most adequate algorithm had to be chosen in order to solve a given real world problem. The authors applied four algorithms separately for the problem and found out the shortest path by using those algorithms. Lastly, the authors compared the paths that they got by those algorithms with the solution they got by assignment problem. The main purpose of that research was to find out which algorithm was suitable out of all the algorithms. The algorithms included that of Prim, Kruskal, Dijkstra and the assignment problem (Hungarian method).

Zhan et al., (1998) suggested that most of the computational testing on shortest path algorithms had been based on randomly generated networks, which might not have the characteristics of real world networks. The authors provided an objective evaluation of fifteen shortest path algorithms using a variety of real road networks. Based on the evaluation, a set of recommended algorithms for computing shortest paths on real road networks was identified. The evaluation should be particularly useful to researchers and practitioners in operations research, management science, transportation and geographic information systems.

2.2 SUMMARY

In this chapter, we reviewed relevant literature in spanning trees.

In the next chapter, we shall put forward the research methodology of the study.

CHAPTER 3

METHODOLOGY

3.0 INTRODUCTION

In this chapter, we shall concentrate solely on the appropriate methodologies that shall be required in this study. We shall analyze the various mathematical theories and algorithms required to solve the problem. We shall first put forward the profile of the Kwahu South District.

3.1 THE KWAHU SOUTH DISTRICT

The Kwahu South District is located on the Kwahu Mountain, with some of its villages beneath the mountain. Its administrative capital is Mpraeso. The district has a vast green belt. The Kwahu South District has the river Asuboni, a source of fish and other resources. The land in the Kwahu South District is a little bit rocky in some areas. The Kwahu South District has the famous Odweanoma Mountain where the paragliding festival is held annually during the Easter festivities. The weather is relatively cool. The people of Kwahu mostly migrated from the Ashanti region of Ghana to the mountains to seek protection from persecutors. The people of Kwahu undertake numerous economic activities including trading, farming and making of pottery. Pottery making is a preserve of the women. The major towns in the district include but not limited to Atibie, Asakraka, Bepong, Mpraeso, Obomeng and Obo.

3.2 MATHEMATICAL MODELS

In dealing with problems of such relatively large sizes, we need to have a formulation of the mathematical principles underpinning the combinatorial problem. In that manner, a solution to the real – world problem, such as in this study shall correspond to a solution of the mathematical model formulated. We shall have a graph depicting the problem under study, which shall be points connected in pairs. In this problem, each and every town under consideration is taken to be a point or node. We connect each pair of points by a line segment. This happens if there is a direct link between the towns under consideration. We associate the line segments with lengths or weights, which are the same as the distances between the two points that the line segments connect.

3.2.1 GRAPH THEORY

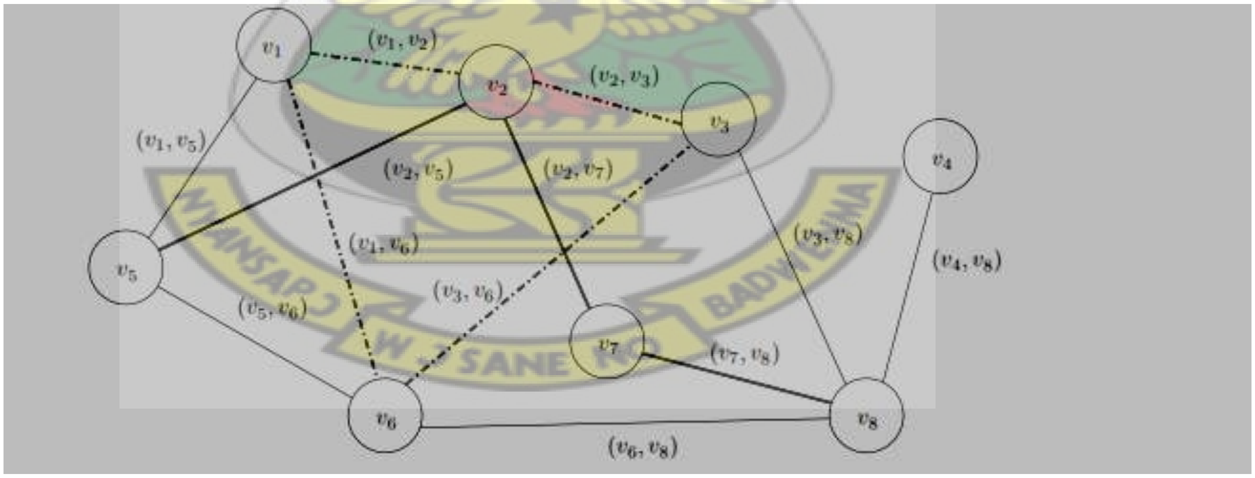


Figure 3.1: A simple connected graph G with n = 8 and m = 12, defined by V = {v1, . . . , v8} and E = {(v_1, v_2), (v_2, v_5). . . (v_6, v_8), (v_7, v_8).

A sequence as (v5, (v2, v5), v2, (v2, v7), v7, (v7, v8), v8) (bold edges) forms a simple path.

A sequence as(v2, (v2, v3), v3, (v3, v6), v 6, (v1, v6), v1, (v1, v2), v_2) (dash-dotted edges) forms a simple cycle.

An undirected (multi) graph G is an abstract data type defined by an ordered pair G = (V, E), where V is a set of vertices and E is a multiset of edges, which are unordered pair of vertices. By definition n = |V| and m = |E|, so we have vertices V = {v_1, v_2, . . . ,v_n} and edges E = {e_1, e_2, . . . , e_m}. The class of graphs with n vertices and m edges is denoted by G(m, n). The vertex set of a specific graph G is denoted by V (G), and the edge set is denoted by E (G).

An edge e_i connects two vertices, say v_a and v_b, and is denoted (va, v_b). By definition of an undirected graph, an edge is an unordered pair, so (v_a, v_b) = (v_b, v_a). A visual example of a graph is given in Figure 3.1. Each edge e_i is assigned a positive weight w (e) that is the cost of "using" the edge. If w (e_i) < w (e_j) for $i \neq j$, then e_i is said to be lighter than ej. Similarly, e_j is said to be heavier than e_i. For simplicity, all example graphs in this thesis will have edge-weights corresponding to the Euclidean distance between endpoints. This is a common assumption in real-world examples, such as road networks and networks of power or data wires.

The degree of a vertex deg (v) is the number of edges that connect to it. This is also referred to as the number of edges incident to vertex, v.

A path in a graph is a sequence of alternating vertices and edges that starts at a vertex and ends at a vertex, such that each edge is incident to its predecessor and successor vertex. That is a sequence (v_{start}, (v_{start}, v_a), v_a, (v_a, v_b), v_b, . . . , v). A cycle is a path with the same start and end vertex. That is $v_{start} = v_{end}$. A simple path is a path where each vertex is distinct. Similarly, a simple cycle is a cycle where each vertex is distinct, except for the start and end vertex. See Figure 3.1 for an example of a simple path and a simple cycle.

A graph is connected if there exists a path between any pair of vertices. That is, if it is possible to get from one vertex to any other vertex in the graph. A self loop is an edge that connects a vertex to itself, which is $e = (v_i, v_i)$. Such edges will be counted twice in deg (v). Two distinct edges, say e and e' are parallel if their endpoints are the same. That is, if $e = (v_i, v_j) = e'$. A graph containing parallel edges is called a multi graph, because E is a multiset as allowed in the first definition of a graph. A simple graph is an undirected graph with no self-loops and no parallel edges. By definition of a simple graph, E is a 'real' set as opposed to a multiset. The degree of a vertex in a simple graph is at most n - 1, and the degree is the same as the number of neighbouring vertices. The graph in Figure 3.1 is simple and connected. If a graph is not connected, its maximal connected sub graphs are called connected components.

We will show a lower bound on the number of edges in a simple connected graph by induction. For the base case, consider a trivial simple connected graph G consisting of one vertex (n = 1), and thus no edges (m = 0). Hence the expression m = n - 1 = 0holds. Then inductively consider a simple connected graph with n - 1 vertices where m = (n - 1) - 1 = n - 2 holds. Adding a new vertex to G requires one edge incident to the new vertex for the graph to be connected, so m = (n - 2) + 1 = n - 1. This proves that any simple connected graph with n vertices has m = n - 1 edges. It is easy to see that connecting a graph to an extra isolated vertex with exactly one edge does not induce a cycle in the simple connected graph. As the graph is connected and has no cycles, there exists exactly one simple path between each pair of vertices. Hence, adding one more edge to the graph will add an extra simple path between some pair(s) of vertices, and thus a cycle. Hence, a connected simple graph with m = n - 1 has no cycles. By definition a tree is a connected graph without cycles, equivalent to that m = n - 1.

A forest is by definition the union of a set of one or more vertex disjoint trees. Another special case of a simple connected graph is the complete graph. A complete graph is a graph which contains all possible edges, that is, each pair of distinct vertices (v_i , v_j) where $i \neq j$ are connected by an edge. Every vertex in a complete graph has degree n -1, that is one incident edge to every other vertex in the graph. The total number of edges is the number of distinct vertex pairs: n (n - 1) / 2 or equivalently $(n^2\text{-}n) / 2$, which is O (n^2) edges. The complete graph with n vertices is denoted K_n. The number of edges m in a simple connected graph with n vertices is $n – 1 \leq m \leq (n^2\text{-}n) / 2$. That is m is Ω (n) and O (n^2).

A sparse graph is loosely defined as a graph where m / n is small. Similarly a dense graph is loosely defined as a graph where m/n is large. The actual definition of sparse and dense graph will be clear in the contexts where the terms are used.

Yet another special case of a simple connected graph is a planar graph. A planar graph is a graph which can be drawn in the plane, such that no edges intersect, except for endpoints at the vertices.

Theorem 3.1

Simple connected planar graphs with $n = 3$ vertices has $m = 3n-6$ edges.

Proof

Let f be the number of faces, which are regions bounded by edges including the outer infinitely large region. If $f = 1$ (the outer face is the only face), then the graph is a tree and $m = n - 1 = 3n - 6$ for $n = 3$.

For $f > 1$, a connected graph with $n = 3$ vertices has $m = 3$ edges. Euler's formula (Som58) states that for a connected planar graph, $n - m + f = 2 \Rightarrow m = n + f - 2$. Each face is bounded by at least 3 edges (a triangular face), and every edge touches at most 2 faces.

Hence, $f = 2m/3$, so $m = n + 2m/3 - 2 \Rightarrow m/3 = n - 2 \Rightarrow m = 3n - 6$.

Consequently, for a planar graph, the density is bounded by a constant and m is O (n).

Due to the lower bound of n - 1 edges for simple connected graphs, n is O (m), which results in O (n + m) = O (m). So the input size of any simple connected graph with m edges is O (m). Consequently, the time needed to build a simple connected graph with m edges is Ω (m).

3.2.2 NETWORKS

Network theory is an area of computer science and network science and part of graph theory. It has applications in many disciplines including statistical physics, particle physics, computer science, biology, economics, operations research, and sociology. Network theory concerns itself with the study of graphs as a representation of either symmetric relations or, more generally, of asymmetric relations between discrete objects. Applications of network theory include logistical networks, the World Wide Web, Internet, gene regulatory networks, metabolic networks, social networks, epistemological networks.

Network problems that involve finding an optimal way of doing something are studied under the name of combinatorial optimization. Examples include network flow, shortest path problem, transport problem, transshipment problem, location problem, matching problem, assignment problem, packing problem, routing problem, Critical Path Analysis and PERT (Program Evaluation & Review Technique).

Social network analysis

Social network analysis examines the structure of relationships between social entities. These entities are often persons, but may also be groups, organizations, nation states, web sites, scholarly publications.

Since the 1970s, the empirical study of networks has played a central role in social science, and many of the mathematical and statistical tools used for studying networks have been first developed in sociology. Amongst many other applications, social network analysis has been used to understand the diffusion of innovations, news and rumours. Similarly, it has been used to examine the spread of both diseases and health-related behaviors. It has also been applied to the study of markets, where it has been used to examine the role of trust in exchange relationships and of social mechanisms in setting prices. Similarly, it has been used to study recruitment into political movements and social organizations. It has also been used to conceptualize scientific disagreements as well as academic prestige. More recently, network analysis (and its close cousin traffic analysis) has gained a significant use in military intelligence, for uncovering insurgent networks of both hierarchical and leaderless nature.

Biological network analysis

With the recent explosion of publicly available high throughput biological data, the analysis of molecular networks has gained significant interest. The type of analysis in this context is closely related to social network analysis, but often focusing on local patterns in the network. For example network motifs are small sub graphs that are over-represented in the network. Similarly, activity motifs are patterns in the attributes of nodes and edges in the network that are over-represented given the network structure.

Link analysis

Link analysis is a subset of network analysis, exploring associations between objects. An example may be examining the addresses of suspects and victims, the telephone numbers they have dialed and financial transactions that they have partaken in during a given timeframe, and the familial relationships between these subjects as a part of police investigation. Link analysis here provides the crucial relationships and associations between very many objects of different types that are not apparent from isolated pieces of information. Computer-assisted or fully automatic computer-based link analysis is increasingly employed by banks and insurance agencies in fraud detection, by telecommunication operators in telecommunication network analysis, by medical sector in epidemiology and pharmacology, in law enforcement investigations, by search engines for relevance rating (and conversely by the spammers for spamdexing and by business owners for search engine optimization), and everywhere else where relationships between many objects have to be analyzed.

Network robustness

The structural robustness of networks is studied using percolation theory. When a critical fraction of nodes (or links) is removed the network becomes fragmented into small disconnected clusters. This phenomenon is called percolation and it represents an order-disorder type of phase transition with critical exponents.

Web link analysis

Several Web search ranking algorithms use link-based centrality metrics, including Google's Page Rank, Kleinberg's HITS algorithm, the Chei Rank and Trust Rank algorithms. Link analysis is also conducted in information science and communication science in order to understand and extract information from the structure of collections of web pages. For example the analysis might be of the interlinking between politicians' web sites or blogs. Another use is for classifying pages according to their mention in other pages.

Centrality measures

Information about the relative importance of nodes and edges in a graph can be obtained through centrality measures, widely used in disciplines like sociology. For example, eigenvector centrality uses the eigenvectors of the adjacency matrix corresponding to a network, to determine nodes that tend to be frequently visited. Formally established measures of centrality are degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, and Katz centrality. The purpose or objective of analysis generally determines the type of centrality measure to be used. For example, if one is interested in dynamics on networks or the robustness of a network to node/link removal, often the dynamical importance of a node is the most relevant centrality measure.

Assortative and Disassortative mixing

These concepts were made because of the nature of hubs in a network. Hubs are generally called nodes which have lots of links. If we see one link in the hub, there is no difference between the hubs, however, some differences are exited between those nodes; some hubs tend to link to the other nodes and other hubs avoid to connect to the other nodes. We call a hub is assortative when it tends to the other hubs. The opposite one is dissortative hub, which avoids connecting to other hubs. If some nodes have some connections with the expected random probabilities, the hubs are neutral. There are three methods to quantify degree correlations.

Spreading processes

Content in a complex network can spread via two major methods: conserved spread and non-conserved spread. In conserved spread, the total amount of content that enters a complex network remains constant as it passes through. The model of conserved spread can best be represented by a pitcher containing a fixed amount of water being poured into a series of funnels connected by tubes. Here, the pitcher represents the original source and the water is the content being spread. The funnels and connecting tubing represent the nodes and the connections between nodes, respectively. As the water passes from one funnel into another, the water disappears instantly from the funnel that was previously exposed to the water. In non-conserved spread, the amount of content changes as it enters and passes through a complex network. The model of non-conserved spread can best be represented by a continuously running faucet running through a series of funnels connected by tubes. Here, the amount of water from the original source is infinite. Also, any funnels that have been exposed to the water continue to experience the water even as it passes into successive funnels. The non-conserved model is the most suitable for explaining the transmission of most infectious diseases, neural excitation, information and rumors, etc.

Interdependent networks

Interdependent networks are a system of coupled networks where nodes of one or more networks depend on nodes in other networks. Such dependencies are enhanced by the developments in modern technology. Dependencies may lead to cascading failures between the networks and a relatively small failure can lead to a catastrophic breakdown of the system. Blackouts are a fascinating demonstration of the important role played by the dependencies between networks. A recent study developed a framework to study the cascading failures in an interdependent networks system.

3.2.3 MINIMUM SPANNING TREES

Definition: Let G be a connected graph. A spanning tree of G is a tree containing all the vertices and a subset of the edges in G. In other words a tree that spans over all vertices in G. Hence, each spanning tree of a connected graph with n vertices has exactly n – 1edges.

Definition: Let T be a spanning tree of a connected weighted graph. The weight of T is defined by the sum of edge weights in T:

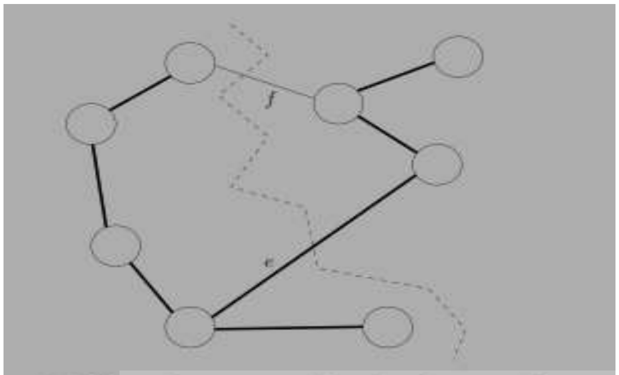
$$w(T) = \sum_{e \in E(T)} w\,(e)\ . \qquad (3.1)$$

Definition: Let G be a connected weighted graph. A minimum spanning tree (MST) of G is a spanning tree of G with minimum total edge weight. It is clear from the definition of a MST, that a solution to the problem requires comparisons of edge-weights.

A graph with equal edge weights may not have a unique minimum spanning tree, since the graph can have multiple spanning trees with equal minimum total edge weight. It will be come clear in the subsequent section that if a graph has distinct edges weights, then the graph has a unique minimum spanning tree.

Properties: The cycle property

Theorem 3.1: For each possible simple cycle in a connected weighted graph G with distinct edge weights, the heaviest edge in the cycle does not belong to a MST of G.



The cycle property. The fat edges are the assumed MST. The dashed line illustrates the splitting of the tree.

Figure 3.2a

Proof: See Figure 3.2a. Assume the contrary, namely that the heaviest edge e belongs to a MST. Deleting e from a MST would split the tree into two disjoint sub trees with the two endpoints of e in different sub trees. There exists some edge $f\neq e$ in the cycle with thetwo endpoints in different sub trees. As $w(f) < w(e)$, reconnecting the two sub trees with f will produce a spanning tree of smaller weight. Hence e does not belong to a MST.

The cut property

Definition: Let V (G) be the vertex set of a graph G such that $|V (G)| \geq 2$. Let V_1 (G) and V_2 (G) be nonempty disjoint partitions of the vertices in V (G). That is $V_1 (G) \neq Ø$, $V_2 (G) \neq Ø$, $V_1 (G) \cap V_2 (G) = Ø$ and $V_1 (G) \cup V_2 (G) = V (G)$. Let C be the set of edges with one endpoint in both V_1 (G) and V_2 (G). Then C forms a cut in the graph G.

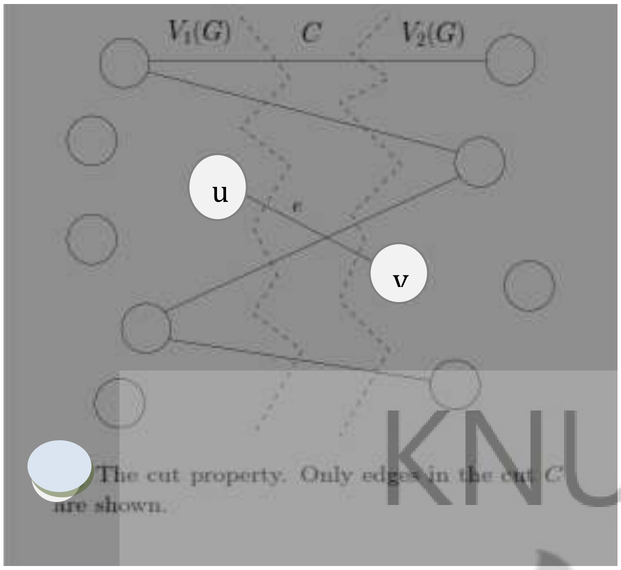


Figure 3.2b

Theorem 3.2

For each possible cut C in a connected weighted graph G with distinct edge weights, the lightest edge in C belongs to a Minimum Spanning Tree of G.

Proof

See Figure 3.2b. Let u and v denote the endpoint vertices of the lightest edge e.

If e is the only edge in C on any simple path between u and v, then the proof is obvious because all vertices in G must be connected in a minimum spanning tree. Otherwise, assume the contrary, namely that the lightest edge e does not belong to a minimum spanning tree. It is obvious that there is some edge, say f, in C on the path between u and v in the assumed minimum spanning tree otherwise the two vertices would not be connected in the minimum spanning tree. As w (e) < w (f), replacing f by e will produce a spanning tree of smaller weight. Hence e belongs to a minimum spanning tree.

Corollaries

Theorem 3.3

Let G be a connected weighted graph with distinct edge weights, let T be a MST of G, and let e be an arbitrary edge in E (G). If e ∈ T, then e is the lightest edge in some cut in G. If e ∉ T, then e is the heaviest edge in some cycle in G.

Proof

Due to the cycle property (Theorem 3.1) and the cut property (Theorem 3.2), we can prove this theorem by proving that any edge in E(G) is either the lightest edge in some cut in G, or the heaviest edge in some cycle in G.

Assume an edge (u, v) exists where this does not apply. If there is a simple path P in G with endpoints u and v, where every edge is proven to be in the minimum spanning tree by the cut property, then P ∪{(u, v)} forms a cycle, and thus (u, v) must be the heaviest edge in this cycle.

Otherwise, then the graph is not connected by minimum spanning tree edges, and thus there exists a cut without any edges proven to be in the minimum spanning tree, so we can add the lightest edge of this cut to the minimum spanning tree. Then repeat this procedure until (u, v) is either proven to be in the minimum spanning tree because it is the lightest edge in such cut, or proven to be the heaviest edge in a cycle.

This will clearly happen at some point, because we repeatedly add a new edge to the minimum spanning tree, so we have a contradiction.

3.2.3.1 PRIM’S ALGORITHM

In computer science, Prim's algorithm is a greedy algorithm that finds a minimum spanning tree for a connected weighted undirected graph. This means it finds a subset of the edges that forms a tree that includes every vertex, where the total weight of all the edges in the tree is minimized. The algorithm was developed in 1930 by Czech mathematician Vojtěch Jarník and later independently by computer scientist Robert C. Prim in 1957 and rediscovered by Edsger Dijkstra in 1959. Therefore it is also sometimes called the DJP algorithm, the Jarník algorithm, or the Prim–Jarník algorithm.

Other algorithms for this problem include Kruskal's algorithm and Borůvka's algorithm. However, these other algorithms can also find minimum spanning forests of disconnected graphs, while Prim's algorithm requires the graph to be connected.

Pseudocode for Prim’s Algorithm

Given a connected weighted graph $G = (V, E)$ with a weight function w and a minimum spanning tree T can be derived from the code below.

```
for any $v \epsilon V$
    cost [v] ← ∞
    parent [v] ← NULL
r ← arbitrary vertex of V
cost [r] ← 0
Q ← V
while $Q \neq \{ \}$
    u ← extract Min (Q)
```

```
        for each v $\epsilon$adja (u) do
            if v$\epsilon$ Q and w (u, v) < cost [v] then
                parent [v] ← u
                cost[v] ← w (u, v)
T ← {(v, parent [v]| v$\epsilon$V – {r}
Return T
```

FINDING THE SOLUTION BY NETWORK METHOD

Step 0: Choose any element r; set S = {r} and A = Ø. (Take r as the root of our spanning tree.)

Step 1: Find a lightest edge such that one endpoint is in S and the other is in V \ S. Add this edge to A and its (other) endpoint to S.

Step 2: If V \ S = Ø, then stop & output (minimum) spanning tree (S, A). Otherwise go to Step 1.

The idea: expand the current tree by adding the lightest (shortest) edge leaving it and its endpoint.

Example 1

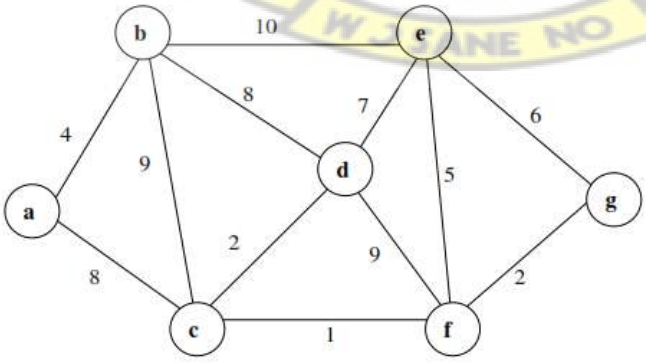


Figure 3.3 Undirected Graph

From Figure 3.3, the minimum spanning tree from Prim's algorithm can be obtained from the network method as follows:

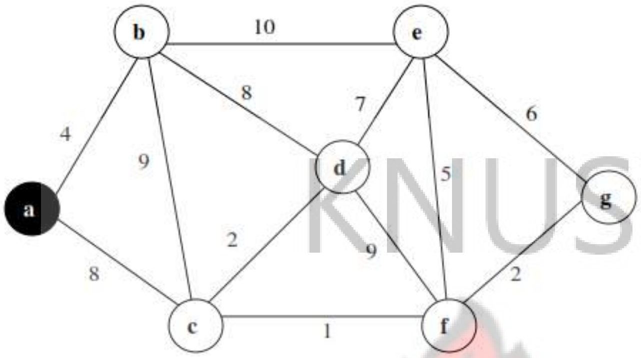


Figure 3.3a Undirected Graph 1

Step 0: $S = \{a\}$; $V \backslash S = \{b, c, d, f, g\}$; $A = \{\ \}$; $lightest\ edge = \{a, b\}$

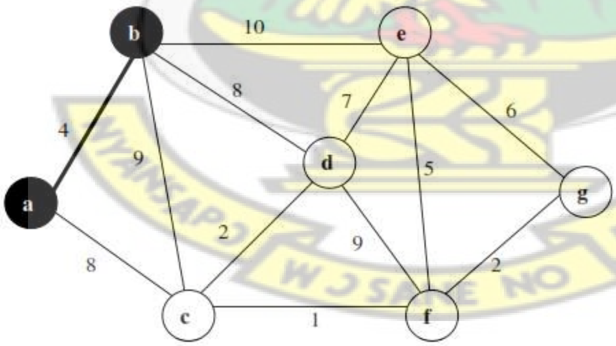


Figure 3.3b Undirected Graph 2

Step 1.0: $S = \{a, b\}$; $V \backslash S = \{c, d, e, f, g\}$; $A = \{\{a, b\}\}$; $lightest\ edge = \{b, d\}, \{a, c\}$

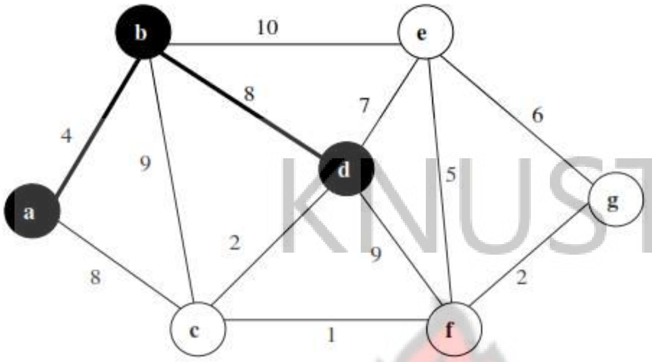


Figure 3.3c Undirected Graph 3

Step 1.1: $S = \{a, b, d\}$; $V \backslash S = \{c, e, f, g\}$; $A = \{\{a, b\},\ \{b, d\}\}$; $lightest\ edge = \{d, c\}$

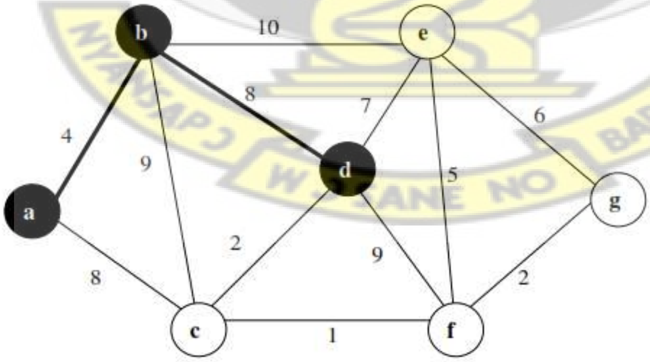


Figure 3.3d Undirected Graph 4

Step1.2:$S = \{a, b, d, c\}$; $V\backslash S = \{e, f, g\}$; $A = \{\{a, b\},\ \{b, d\}, \{d, c\}\}$; $lightest\ edge = \{c, f\}$

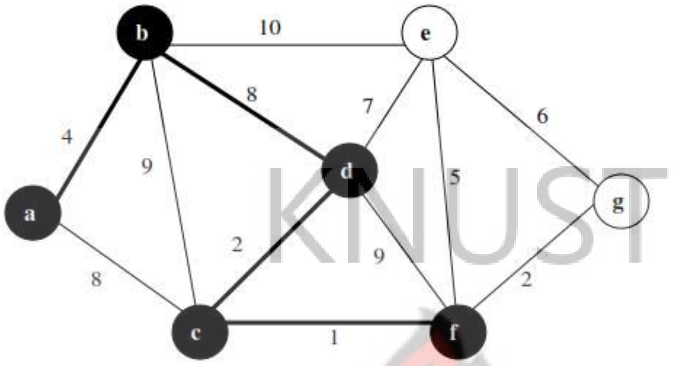


Figure 3.3e Undirected Graph 5

Step1.3:$S = \{a, b, d, c, f\}$; $V\backslash S = \{e, g\}$;

$A = \{\{a, b\},\qquad \{b, d\}, \{d, c\}, \{c, f\}\}$; $lightest\ edge = \{f, g\}$

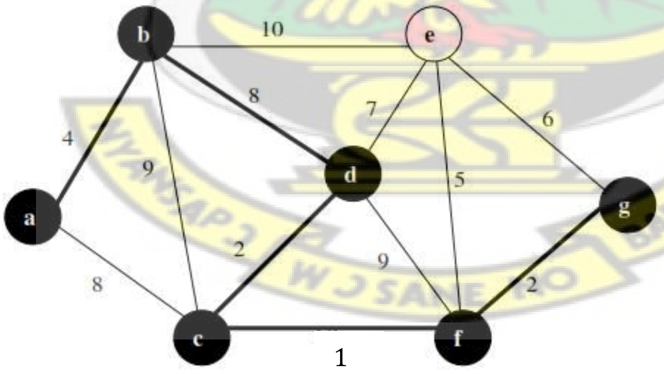


Figure 3.3f Undirected Graph 6

Step 1.4: $S = \{a, b, d, c, f, g\}$; $V\backslash S = \{e\}$;

$$A = \{\{a, b\}, \{b, d\}, \{d, c\}, \{c, f\}, \{f, g\}\}; \ lightest\ edge = \{f, e\}$$

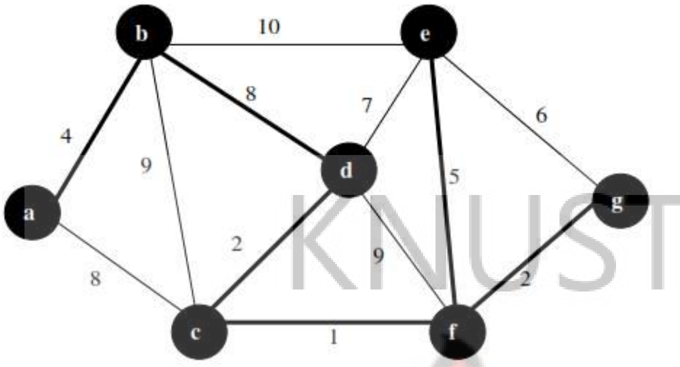


Figure 3.3g Undirected Graph 7

Step2.0: $S = \{a, b, c, d, e, f, g\}$; $V\backslash S = \{\ \}$; $A = \{\{a, b\}, \{b, d\}, \{d, c\}, \{c, f\}, \{f, g\}, \{f, e\}\}$

Since $V\backslash S = \{\ \}$, it means minimum spanning tree is complete.

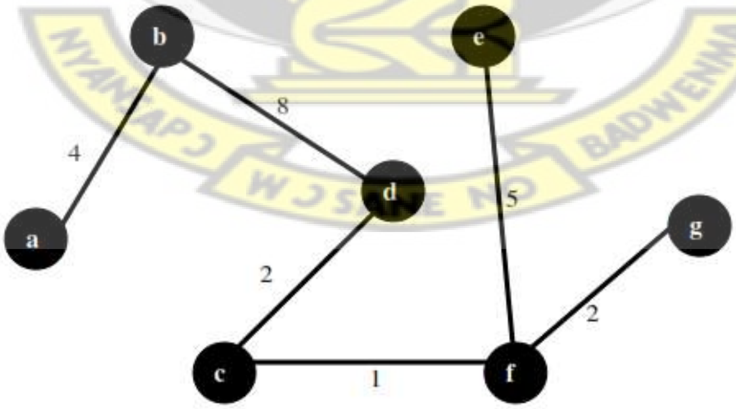


Figure 3.3h Minimum Spanning Tree

The minimum total weight of the tree is 4+8+2+1+2+5 = 22.

FINDING SOLUTION USING MATRIX METHOD

Step 0: With the matrix representing the network, choose a starting vertex. Delete the row corresponding to that vertex.

Step 1: Label with ‘1’ the column corresponding to the start vertex and ring the smallest undeleted entry in that column.

Step 2: Delete the row corresponding to the ringed entry.

Step 3: Label (with the next number) the column corresponding to the deleted row.

Step 4: Ring the lowest undeleted entry in all labeled columns.

Step 5: Repeat the last three steps until all rows are deleted. The ringed entries represent the edges in the minimum connector.

When there is a tie in the smallest values, it is broken arbitrarily.

Example 2

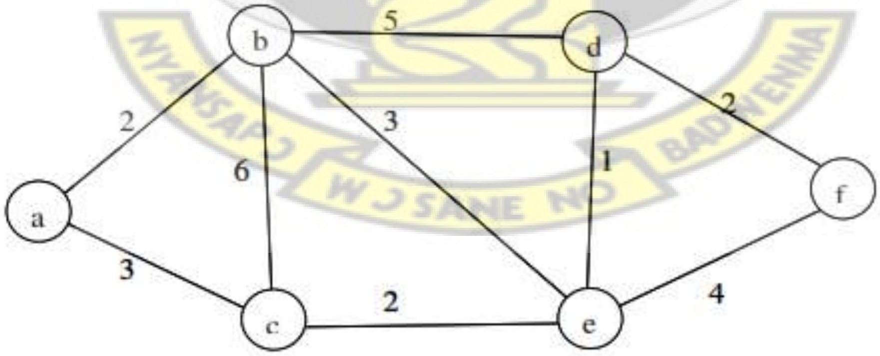


Figure 3.4: Hypothetical network

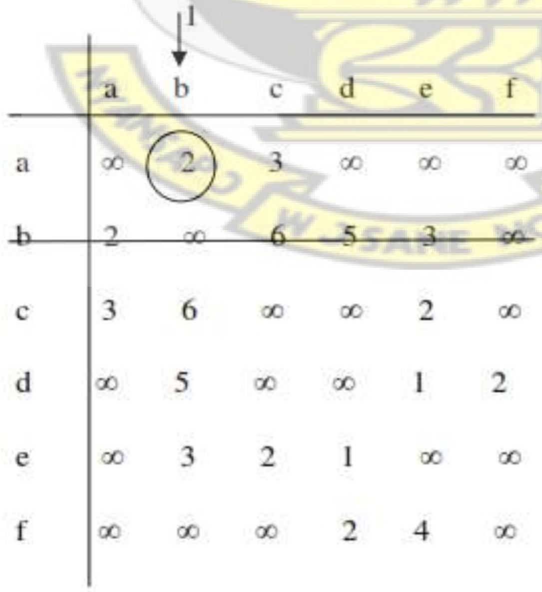
SOLUTION USING MATRIX METHOD

Table 3.1 Node distance matrix 1

	a	b	c	d	e	f
a	∞	2	3	∞	∞	∞
b	2	∞	6	5	3	∞
c	3	6	∞	∞	2	∞
d	∞	5	∞	∞	1	2
e	∞	3	2	1	∞	∞
f	∞	∞	∞	2	4	∞

Choose a starting vertex say **b,** delete row **b,** and look for the smallest entry in column **b.**

Table 3.2 solution matrix 1

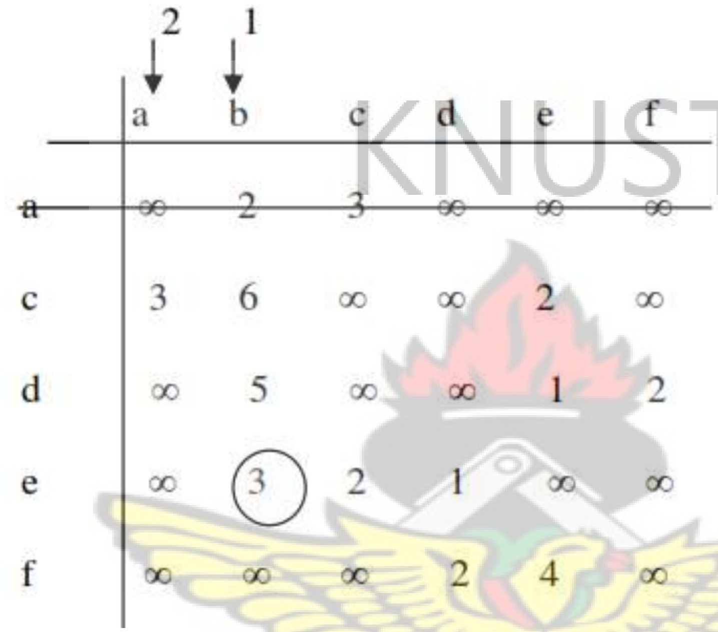


	a	b	c	d	e	f
a	∞	2	3	∞	∞	∞
~~b~~	~~2~~	~~∞~~	~~6~~	~~5~~	~~3~~	~~∞~~
c	3	6	∞	∞	2	∞
d	∞	5	∞	∞	1	2
e	∞	3	2	1	∞	∞
f	∞	∞	∞	2	4	∞

The edge **ba** is the smallest edge joining **b** to the other vertices. Put edge **ba** into the solution.

Delete row **a** and look for the smallest entry columns **b** and **a**.

Table 3.3 solution matrix 2



	2 ↓	1 ↓				
	a	b	c	d	e	f
~~a~~	~~∞~~	~~2~~	~~3~~	~~∞~~	~~∞~~	~~∞~~
c	3	6	∞	∞	2	∞
d	∞	5	∞	∞	1	2
e	∞	3	2	1	∞	∞
f	∞	∞	∞	2	4	∞

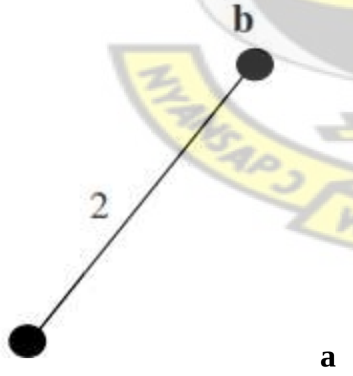


Figure 3.5 Partial pair connection

Table 3.4 solution matrix 3

	2	1			3	
	↓	↓			↓	
	a	b	c	d	e	f
c	3	6	∞	∞	2	∞
d	∞	5	∞	∞	①	2
e	∞	3	2	1	∞	∞
f	∞	∞	∞	2	4	∞

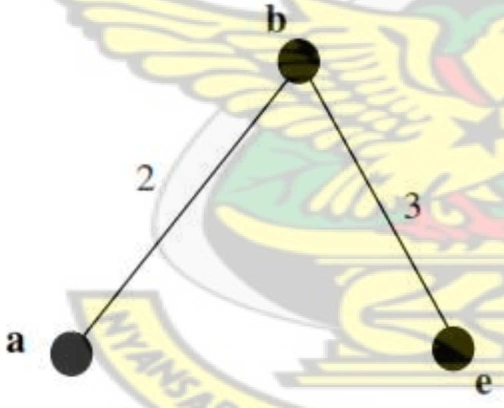


Figure 3.6: Partial Connection

ed is the smallest edge joining **b, a**, and **e**to the other vertices. Put the edge **ed** into the solution and delete row **d**. Look for the smallest entry in columns **b, a, e** and **d**

Table 3.5 solution matrix 4

	2 ↓	1 ↓		4 ↓	3 ↓	
	a	b	c	d	e	f
c	3	6	∞	∞	2	∞
~~d~~	~~∞~~	~~5~~	~~∞~~	~~∞~~	~~1~~	~~2~~
f	∞	∞	∞	2	4	∞

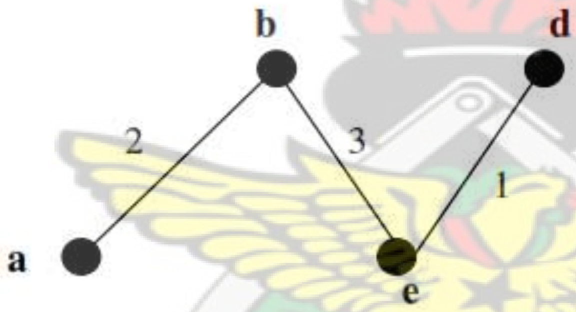


Figure 3.7 Partial Connections 1

df is the smallest edge joining **b, a, e**, and d to the other vertices. Put **df** into the solution and delete row **f**. Look for the smallest entry in columns **b, a, e**, **d** and **f.**

Table 3.6 solution matrix 5

	2	1		4	3	5
	a	b	c	d	e	f
c	3	6	∞	∞	②	∞
~~f~~	~~∞~~	~~∞~~	~~∞~~	~~2~~	~~4~~	~~∞~~

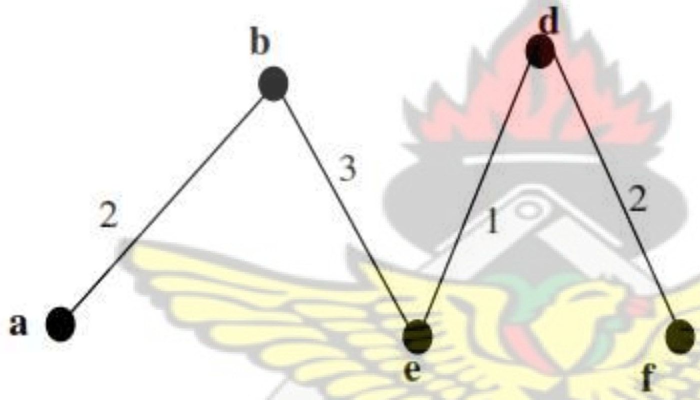


Figure 3.8 Partial connected network

Table 3.7 solution matrix 6

	2	1	6	4	3	5
	a	b	c	d	e	f
~~c~~	~~3~~	~~6~~	~~∞~~	~~∞~~	~~2~~	~~∞~~

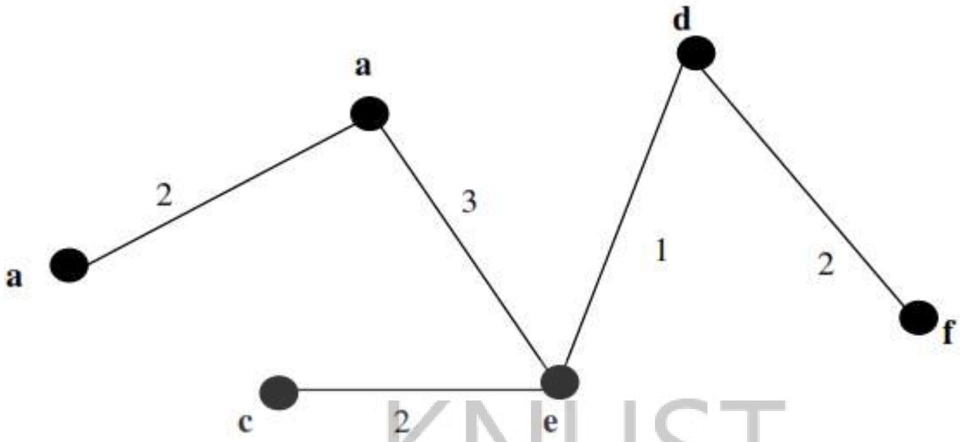


Figure 3.9 Minimum Spanning Tree (MST)

The minimum length (weight or cost) of the network (Figure 3.9) is 10 units, which is the total sum of the edge values (2+3+2+1+2=10).

3.2.3.2 KRUSKAL’S ALGORITHM

Kruskal's algorithm is a greedy algorithm in graph theory that finds a minimum spanning tree for a connected weighted graph. This means it finds a subset of the edges that forms a tree that includes every vertex, where the total weight of all the edges in the tree is minimized. If the graph is not connected, then it finds a minimum spanning forest (a minimum spanning tree for each connected component).

Let G = (V, E) be a connected, undirected, weighted graph with weight function w: E → R.

It starts with each vertex being its own component.

Repeatedly merges two components into one by choosing the light edge that connects them (i.e., the light edge crossing the cut between them).

Scans the set of edges in monotonically increasing order by weight.

Uses a disjoint-set data structure to determine whether an edge connects vertices in different components.

Pseudocode for Kruskal's Algorithm

```
Kruskal (G; w)
A = Ø
For each vertex v ∈ G.V
    Make – Set (v)
Put the edges of G.E into list sorted by non decreasing weight
For each (u, v) ∉ G.E taken from the sorted list
    If find-set (u) ≠ find-set(v)
        A = A ∪ {(u, v)}
        Union (u; v)
Return A
```

Here a disjoint set data structure is used:

Make-Set (v) creates a set containing a single element v

Find-Set (v) finds a set containing v.

Union (u; v) merges sets containing u and v into a single set

A is the minimum spanning tree.

Example

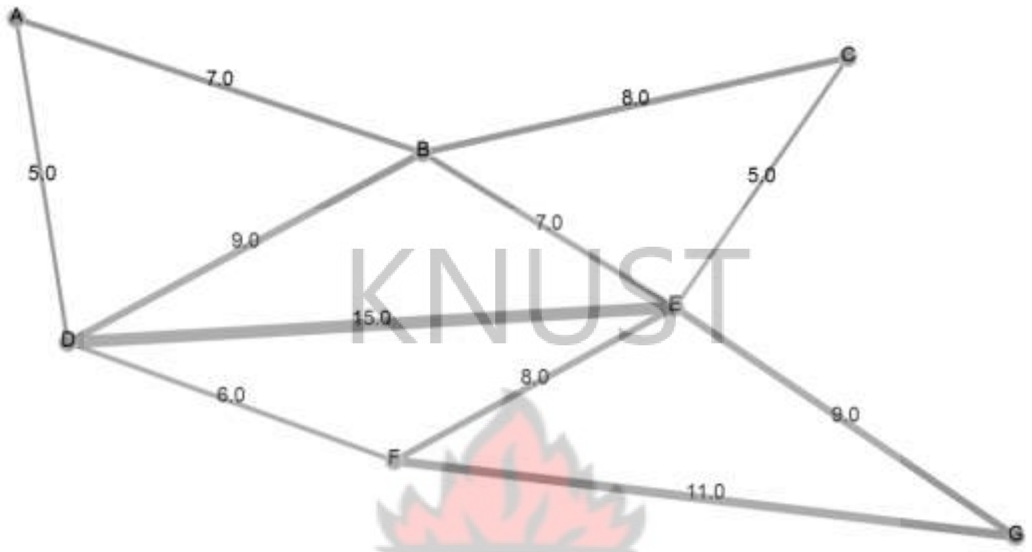


Figure 3.10 Hypothetical Network

1. **AD** and **CE** are the shortest edges, with length 5, and **AD** has been arbitrarily chosen.
2. **CE** is now the shortest edge that does not form a cycle, with length 5, so it is the second edge.
3. The next edge, **DF** with length 6, is chosen using much the same method.
4. The next-shortest edges are **AB** and **BE**, both with length 7. **AB** is chosen arbitrarily. The edge **BD** has been ignored because there already exists a path between **B** and **D**, so it would form a cycle (**ABD**) if it were chosen.
5. The process continues to choose the next-smallest edge, **BE** with length 7. Many more edges are chosen at this stage: **BC** because it would form the loop **BCE**, **DE** because it would form the loop **DEBA**, and **FE** because it would form **FEBAD**.

6. Finally, the process finishes with the edge **EG** of length 9, and the minimum spanning tree is found.

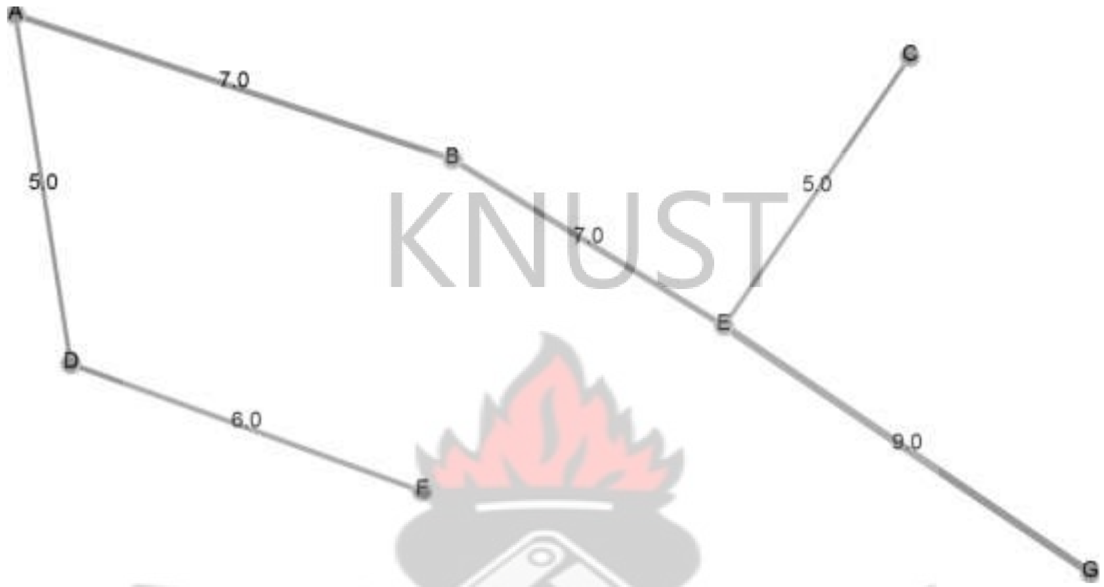


Figure 3.11 Minimum Spanning Tree

3.3 SUMMARY

In this chapter, we discussed the research methodology of the study.

The next chapter is devoted to the data collection and analysis of the study.

CHAPTER 4

DATA COLLECTION AND ANALYSIS

4.0 INTRODUCTION

This chapter shall be devoted to data collection and analysis. We shall determine the minimum spanning tree for the pipelines to be connected for the water supply system in the Kwahu South District of Ghana. We shall also analyze the data using the MATLAB (Matrix Laboratory) program (prims.m). The pseudocode for Prim's algorithm shall be implemented to establish the minimum spanning tree. In this study, the distance between any two towns (nodes) shall be taken to be equivalent to the length of the pipeline needed to connect a given edge.

4.1 DATA COLLECTION

The data used in the study were partly obtained from the office of the Town and Country Planning Department and the Department of Feeder Roads in the district. Part of the data was also obtained electronically (on the internet). The map of the district under consideration was used to obtain some salient data.

The towns to be covered shall include Abetifi (**AB**), Aduamoa (**AD**), Asakraka (**AS**), Atibie (**AT**), Bepong (**BE**), Kotoso (**KO**), Mpraeso (**MP**), Nkwatia (**NK**), Obo (**OB**), Obomeng (**OG**), Pepease (**PE**) and Tafo (**TA**). The distances are all in kilometres (km).

Table 4.1 The distances between all pairs of nodes.

	KO	TA	AS	BE	MP	AT	OG	OB	NK	AD	AB	PE
KO	∞	10	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
TA	10	∞	4	∞	∞	∞	∞	∞	∞	∞	∞	11
AS	∞	4	∞	7	∞	∞	∞	∞	∞	∞	∞	∞
BE	∞	∞	7	∞	3	∞	∞	∞	∞	∞	∞	∞
MP	∞	∞	∞	3	∞	1	3	∞	6	∞	∞	∞
AT	∞	∞	∞	∞	1	∞	4	∞	∞	∞	∞	∞
OG	∞	∞	∞	∞	3	4	∞	4	∞	∞	∞	∞
OB	∞	∞	∞	∞	∞	∞	4	∞	∞	4	∞	∞
NK	∞	∞	∞	∞	6	∞	∞	∞	∞	3	7	∞
AD	∞	∞	∞	∞	∞	∞	∞	4	3	∞	∞	∞
AB	∞	∞	∞	∞	∞	∞	∞	∞	7	∞	∞	5
PE	∞	11	∞	∞	∞	∞	∞	∞	∞	∞	5	∞

Table 4.1 was used to construct the network in Figure 4.1.

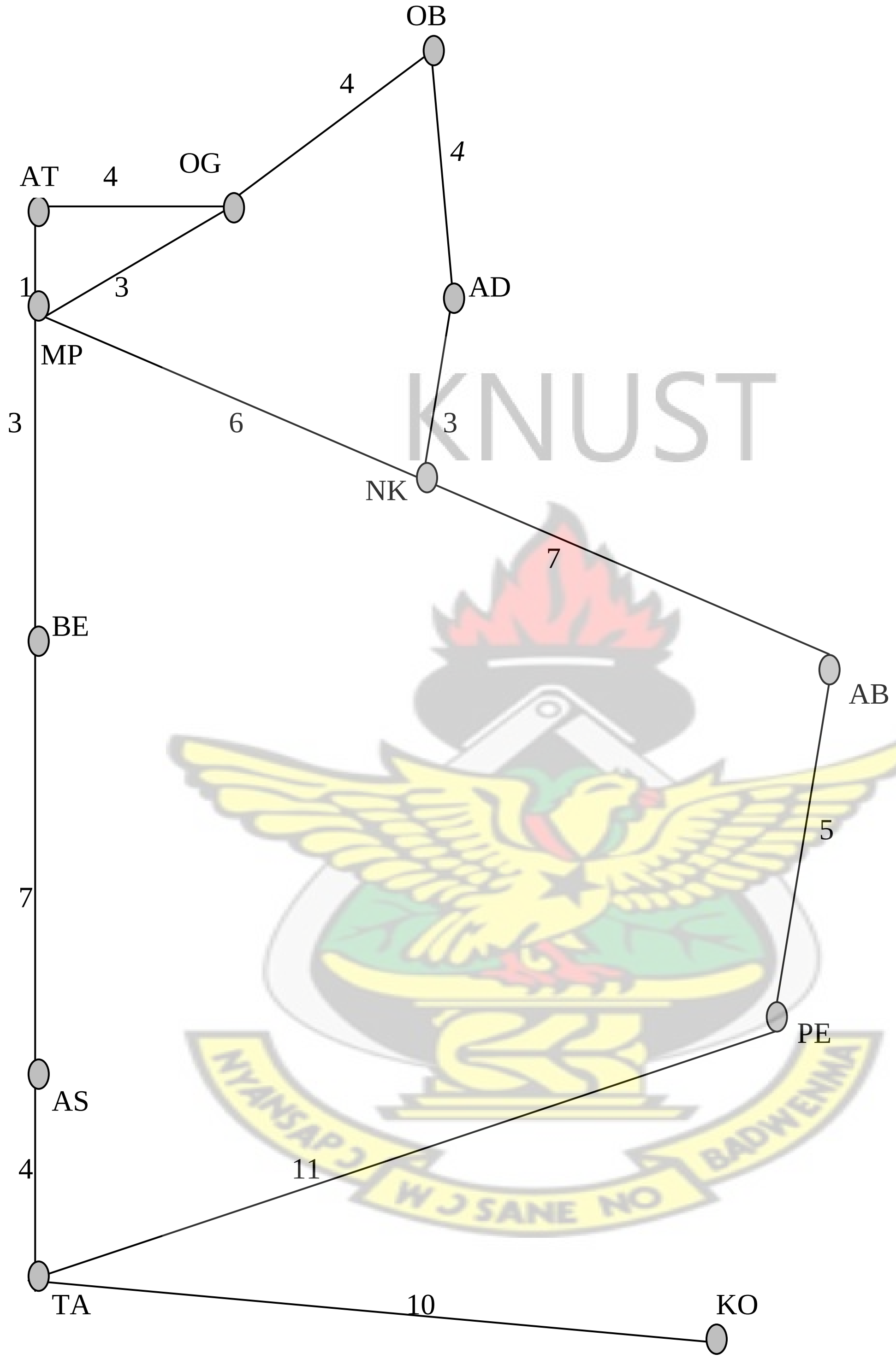


Figure 4.1 Network of the Lengths of Pipelines

4.2 DATA ANALYSIS

We used Prim's algorithm for the analysis of the data in table 4.1. The matrix method was employed in this situation. The output from the algorithm starting from **KO** is as follows:

Table 4.2 Output of the Prim's Algorithm

Iteration	Starting Node	End Node	Distance (km)
1	KO	TA	10
2	TA	AS	4
3	AS	BE	7
4	BE	MP	3
5	MP	AT	1
6	AT	OG	4
7	OG	OB	4
8	OB	AD	4
9	AD	NK	3
10	NK	AB	7
11	AB	PE	5

The output generated from the Prim's Algorithm was used to construct the minimum spanning tree in Figure 4.2.

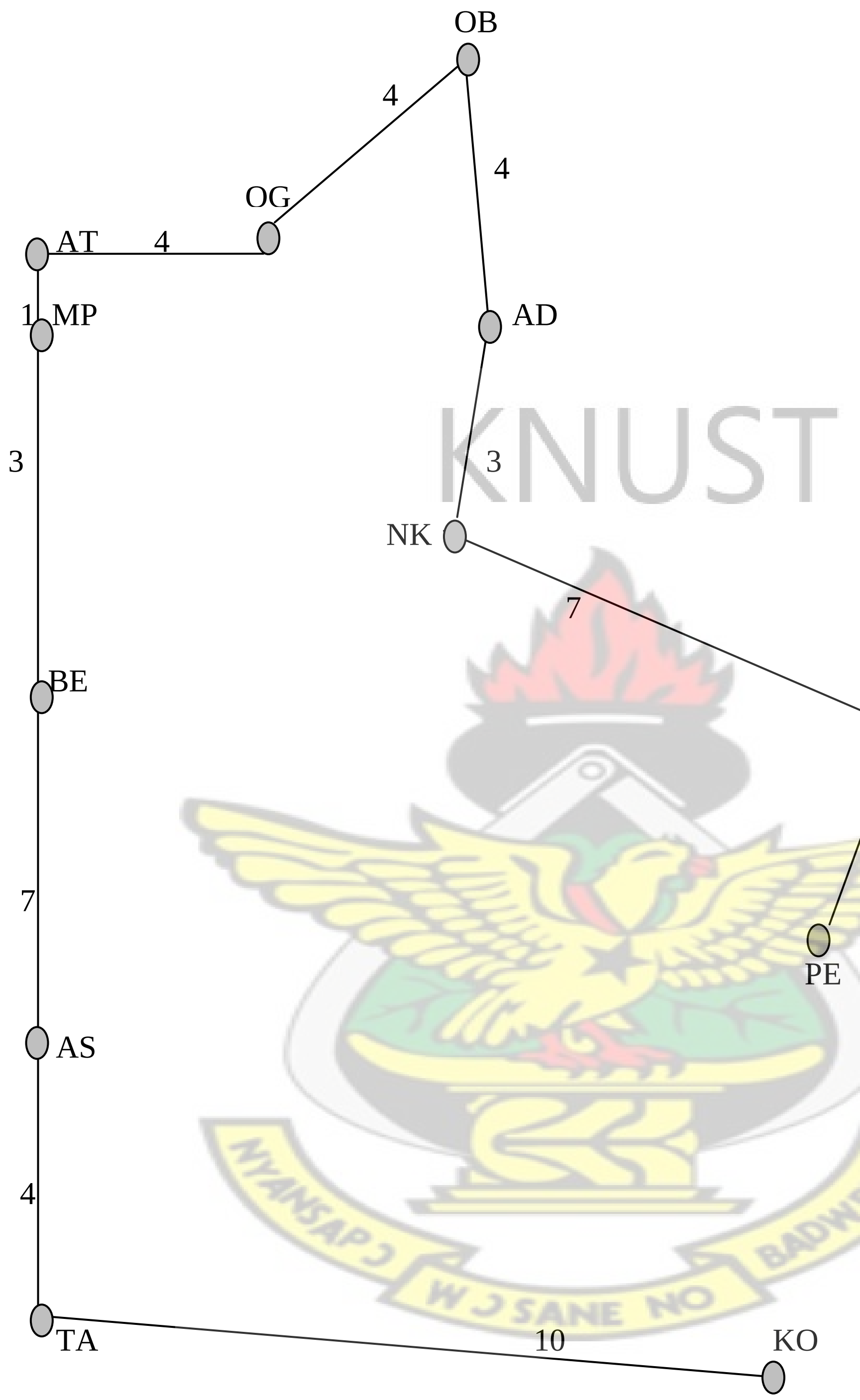


Figure 4.2 Minimum Spanning Tree for the Pipeline connections.

The total weight of the spanning tree is **52km** (= 10+4+7+3+1+4+4+4+3+7+5).

4.3 FINDINGS

The study has obtained the minimum spanning tree of pipelines needed for water supply in the major towns in the Kwahu South District of Ghana.

The pipelines in this spanning tree must thus be given the much needed priority in terms of development and maintenance if water supply in the district is to be at optimum levels.

From Figure 4.2, **KO** and **PE** are terminal nodes and are the remotest water supply points. None of the supply points could particularly be described as possible hubs of water supply but to a lesser degree **AT**, **OG** and **OB** since they are more of central supply points. Water supply at **AT, OG** or **OB** could be boosted since they can also serve as supply sources for **AD, NK, AB** and **PE**. Apart from the main source **KO**, **PE** could also have served as a perfect source for water supply. The other supply points could also serve as supply points with little problems as regards pipe lengths. The minimum spanning tree pipeline length if used in procuring the required pipes shall be a major cost cutting adventure. Laying pipes for water supply is a very capital intensive exercise and therefore minimizing the cost of such a project shall automatically make resources available for other equally important aspects of the water supply chain. The costs of pipelines are not fixed and even differ from brand to brand possibly due to quality issues.

Table 4.3 Shortest length of Pipelines from One Supply Point (Town) to another

FROM	TO	LENGTH (km)
KOTOSO	TAFO	10
TAFO	ASAKRAKA	4
ASAKRAKA	BEPONG	7
BEPONG	MPRAESO	3
MPRAESO	ATIBIE	1
ATIBIE	OBOMENG	4
OBOMENG	OBO	4
OBO	ADUAMOA	4
ADUAMOA	NKWATIA	3
NKWATIA	ABETIFI	7
ABETIFI	PEPEASE	5

4.4 SUMMARY

This chapter presented the data collection and analysis of the study.

The next, which is the final chapter of the study presents the conclusions and recommendations of the study.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.0 INTRODUCTION

This chapter presents conclusions and recommendations of the study. The main objective of the study was to analyze and determine the optimal pipeline connection for the water supply system in the Kwahu South District of Ghana. The objective stemmed from the fact that the cost of water production should be at the lowest levels possible without compromising the quality of the water so produced. The major towns were captured in a network and Prim's algorithm used to find the minimum spanning tree length of pipelines required to connect the nodal towns.

5.1 CONCLUSIONS

In this study, we exposed the major towns selected in the form of a network and used the Prim's algorithm to construct the minimum spanning tree, which is a sub graph of the network in Figure 4.1, containing all the nodal towns for which the total length is a minimum.

The study revealed that the minimum spanning tree length for the pipeline connections for the water supply system in the Kwahu South District of Ghana is approximately fifty – two (52) kilometres.

The study also showed that with this minimum spanning tree, the total cost of procuring pipelines would be significantly reduced. This consequently will make resources available for other productive ventures in the water treatment sector.

5.2 RECOMMENDATIONS

Since water supply has always being a critical issue for the government and the masses, it is strongly recommended that the minimum spanning tree of pipelines requirement be given a critical attention by the water supply agencies.

After all, the dangers associated with using unclean water are much more expensive to overcome than to provide treated water for domestic and industrial purposes.

Once water reaches these nodal towns in abundance, the other villages could easily tap into them.

It is also highly recommended that the water supply agencies incorporate this in their plans as regards water supply in order to minimize cost.

REFERENCES

Achuttan N. R and Caccetta L. (1992), Minimum Weight Spanning Tree With Bounded Diameter. Australasian Journal of Combinatorics 5 (1992), pp. 261-276

Ahuja R., Magnanti T. and Orlin J. (1993), Network Flows: theory, algorithms and applications. Prentice Hall

Akutsah F. (2011), Minimum Spanning Tree Route for Major Tourist Centers in the Brong Ahafo Region of Ghana. Kwame Nkrumah University of Science and Technology Institute of Distance Learning

Althaus E., Funke S., Har-Peled S, Könemann J, Edgar A. R. and Skutella M (2004), Approximating k-hop minimum-spanning trees In Operations Research Letters 33 (2005) 115 – 120.

Assad, A. and Xu, W. (1992), The Quadratic Minimum Spanning Tree Problem. Naval Research Logistics, 39: 399–417. doi: 10.1002/1520-6750(199204)39:3<399::AID-NAV3220390309>3.0.CO;2-0

Ben - Zwi O. (2006), Lower Bounds for Testing Euclidean Minimum Spanning Trees. Master's Thesis, University of Haifa, Department of Computer Science (July, 2006)

Berganti˜nos G. and Vidal - Puga J. (2009), The Folk Solution and Boruvka's Algorithm in Minimum Cost Spanning Tree Problems

Chartrand G. and Zhang P. (2009), Chromatic Graph Theory. CRC Press, USA

Cheong O. and Changryeol L. (2012), Single-Source Dilation-Bounded Minimum Spanning Trees. Computational Geometry (cs.CG). arXiv: 1206.6943v1

Chi - Chih A. Y. (1982), On Constructing Minimum Spanning Trees in k-Dimensional Spaces and Related Problems. SIAM, On Computing, 11(1982), 721-736...

Claus A. (2008), An Optimal Minimum Spanning Tree Algorithm.
A Master's Thesis Submitted to the Department of Computer Science, Aarhus University

Dae Y. S., Lee D.T. and Tien-Ching L. (2009), Geometric Minimum Diameter Minimum Cost Spanning Tree Problem. Springer-Verlag Berlin Heidelberg 2009

Donkoh E. K., Amponsah S. K. and Darkwah K. F. (2010), Optimal Pipeline Connection for the West African Gas Pipeline Project. Research Journal of Applied Sciences, Engineering and Technology 3(2): 67-73, 2011 ISSN: 2040-7467

Erciyes K. (2010), Design of Algorithms for Bioinformatics.
kayhan.erciyes@izmir.edu.tr.

Erickson J. (2009), Minimum Spanning Tree. http://creativecommons.org/liceses/by-nc-sa,

Frey H., Ingelrest F. and Simplot-Ryl D.(2007), Localized Minimum Spanning Tree Based Multicast Routing with Energy-Efficient Guaranteed Delivery in Ad Hoc and Sensor Networks. Theme COM Systemes communicants Projets POPS Rapport technique June 2007.

Golden B., Raghavan S. and Stanojević D. (2004), Heuristic Search for the Generalized Minimum Spanning Tree Problem. *Journal on Computing* Summer 2005 vol. 17 no. 3 290-304

Golden B., Raghavan S. and Stanojević D. (2006), The prize-collecting generalized minimum spanning tree problem. J Heuristics (2008) 14: 69–93 DOI 10.1007/s10732-007-9027 -1 Springer Science+Business Media, LLC 2007

Gorod A. and Sauser B. (2007), An Application of Prim's Algorithm in Defining a SoS Operational Boundary. CSER 2007, Stevens Institute of Technology , ISBN - 0-9787122-1-8 PROCEEDINGS CSER 2007 , March 14-16, Hoboken, NJ , USA

Holm J., Kristian de Lichtenberg and Thorup M. (1998), Poly-Logarithmic Deterministic Fully-Dynamic Algorithms for Connectivity, Minimum Spanning Tree, 2-Edge and Biconnectivity. Journal of the ACM, Vol. 48, No. 4, July 2001, pp. 723–760.

http://en.wikipedia.org/w/index.php?title=Kruskal%27s_algorithm&oldid=527562607

http://www.carlschroedl.com/blog/wp-content/uploads/2012/05/wiki_kruskal_example_mst.png.

Jose A. C. and Hero O. A. (2003), Applications of Entropic Graphs to Nonparametric Estimation. Dept. Of Electrical Engineering and Computer Science, University Michigan.

Joyner D., Ngugen M. V. and Cohen S. N. (2011), Algorithmic Graph Theory. http://code.google.com/p/graph-theory-algorithms-book/.

Katajainen J. and Nevalainen O. (1982), An alternative for the Implementation of Kruskal's Minimum Spanning Tree Algorithm. Science of Computer Programming 3 (1983) 205-216 North Holland

Khan M., Pandurangan G. and Anil V.S. K. (2006), Distributed Algorithms for Constructing Approximate Minimum Spanning Tree in Wireless Networks. IEEE Transactions on Parallel Distributed System, pp. 1-14.

Lundelius W. J., Lamport L. and Lynch N. (1988), A Lattice-Structured Proof Technique Applied to a Minimum Spanning Tree Algorithm (Extended Abstract). 0 1988 ACM O-89791-277-2/88/0007/0028

Maki, D. P. & Thompson, M. (1996), Finite Mathematics (4th Ed.), New York: McGraw-Hill.

Martin G. and Gunther R. R. (2009), Solving the Euclidean Bounded Diameter Minimum Spanning Tree Problem by Clustering-Based (Meta-) Heuristics. In proceeding of: Computer Aided Systems Theory-EUROCAST 2009, 12th International Conference, Las Palmas de Gran Canaria, Spain February 15-20, 2009

Mathias G. (2009), Combinatorial Optimization. MSM 3M02b,
www.web.mat.bham.ac.uk/m.gerdts

Mauricio G.C. Resende and Panos M. Pardalos (2003), Interior point algorithms for network -flow problems. AT&T Bell Laboratories, Murray Hill, NJ 07974-2070 USA
The University of Florida, Gainesville, FL 32611-6595 USA
Received: September 22, 2010 Accepted: January 20, 2011 Published: February 20, 2011.

Max A. (2010), Analysis of Algorithms. University of South Carolina, USA.

Mert R., Sabuncu and Peter R. (2008), Using Spanning Graphs for Efficient
Image Registration. IEEE Transaction on Image Processing. Vol. 17, No. 5, pp. 788-798.

Metaxas P., Gloor P. A., Johnson D. B., and Makedon F. (1993), A Visualization System for Correctness Proofs of Graph Algorithms. Computer Science Education, 3(3) pp 315-333, 1992

Micah A., Wolfgang D., Juurlink B., Miroslaw K.L. and Ingo R. (1999), Communication-Optimal Parallel Minimum Spanning Tree Algorithms.citeseerx.ist.psu.edu/viewdoc/summa…

Mohammad G., Mahini H., Mirjalali K., Shayan O. G., Amin S., Sayedi R. and Zadimoghaddam M. (2007), Spanning trees with minimum weighted degrees. Information Processing Letters 104 (2007) 113–116

Narsingh D. (1974), Graph Theory with Application to Engineering and Computer Science. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Narsingh D. and Abdalla A. (2000), Computing a Diameter-Constrained Minimum Spanning Tree in Parallel. CIAC 2000, LNCS 1767, pp. 17-31, 2000. Springer-Verlag Berlin Heidelberg 2000.

Pankaj K. Agarwal, Eppstein D., Leonidas J.,Guibas,Monika R. H. (1998), Parametric and Kinetic Minimum Spanning Trees. IEEE Proceedings of FOCS'98, 8-11

Papadimitriou et al., (1998), Combinatorial Optimization-Algorithms and Complexity. Dover Publications

Patterson R. and Pirkul H. (2000), Heuristic Procedure Neural Networks for the CMST Problem.

Perez L. F. C. (2009), Improving Resource Utilization in Carrier Ethernet Technologies. A Thesis Presented to the Graduate School, University de Girona, Catalonia, Spain.

Peter S. J. and Victor S. P. (2010), A Novel Algorithm for Minimum Spanning Clustering Tree. Journal of Theoretical and Applied Information Technology, 2005 - 2010 JATIT.

Pettie S. and Ramachandran V. (2002), Randomized Minimum Spanning Tree Algorithms Using Exponentially Fewer Random Bits. Proceedings 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 713-722, 2002.

Quoc P. T. (2012), A Heuristic Approach for Solving Minimum Routing Cost Spanning Tree Problem. (Email: phantanquoc@gmail.com).

Raj P., and Sood M. (2012), Ant Colony Optimization is the limited case of Prim's Algorithm. International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012, 4202-4204

Rama M. P. (2007), Operations Research (2^{nd} Ed.), New Delhi: New Age International.

Read N. and Jackson T. S. (2009), Theory of minimum spanning trees : Mean-field theory and strongly disordered spin-glass model, Phys. Rev. E 81, 021130 (2010) arXiv:0902.3651v3

Rico Z. (2011), Matroidal Degree-Bounded Minimum Spanning Trees.
NASA ADS arXiv: 1107.5329v1[cs.DS]. E-mail: ricoz@math.mit.edu.

Ruohonen K. (2008), Graph Theory

Sörensen K. and Janssens G. K. (2005), An algorithm to generate all spanning trees of a graph in order of increasing cost. Pesquisa Operacional, v.25, n.2, p.219-229, Maio a Agosto de 2005, versão impressa ISSN 0101-7438 / versão online ISSN 1678-5142

Subadra M., Bhagavan D. and RaoC. R. S.(2011), Directed Graph Algorithms for Tours – A Case Study. Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS) 2 (4): 615-618 (ISSN: 2141-7016)

Taha H. A. (2007), Operations Research; An Introduction (8thEdition), Pearson Prentice Hall.

Takahashi H. and Matsuyama A. (1979), An Approximate Solution For The Steiner Problem In Graphs. Math Japonica 24, No. 6(1980), 573-577.

Taylor L. (2010), Randomly Generated Edge-Disjoint Minimum Spanning Trees based Data Gathering Trees. http://www.answers.com/topic/cdma.

Turgay K. and Marwan K. (2001), Multi-Constrained Optimal Path Selection.
IEEE INFOCOM 2001.

Vikas C. S. (2010), Minimum Spanning Tree Algorithm. 2010 International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 8

Vincent F., Tijs S. and Shigeo M. (1994), Minimum Cost Spanning Extension
Proportional Rule and the Decentralized Rule

William B. M., Parikshit R. and Gray G. A (2010), Fast Euclidean Minimum Spanning Tree: Algorithm, Analysis, and Applications. KDD'10, July 25–28, 2010, Washington, DC, USA. Copyright 2010 ACM 978-1-4503-0055-110/07

www.cse.ust.hk/~dekai/271/notes/L07/L07.ps.

www.wikipedia.com

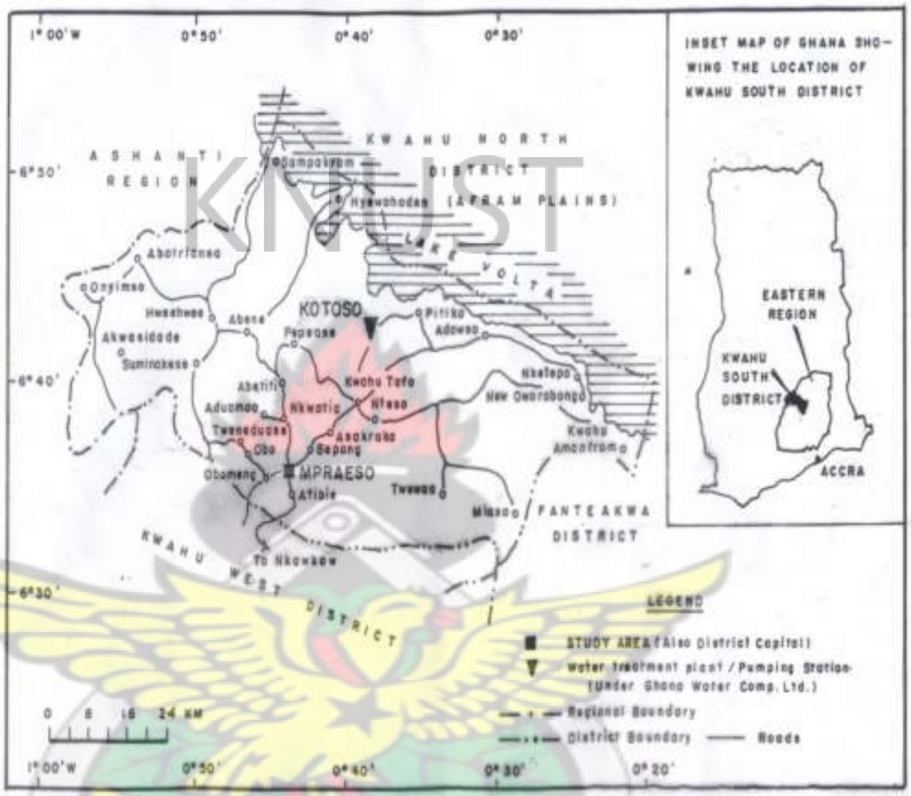
Yueyang L., Hong J. and Guangxin Y. (2006), An Energy-Efficient PEGASIS Based Enhanced Algorithm in Wireless Sensor Networks. China Communication 2006, Technology Forum, pp. 91-96.

Zhan F. B and Noon C. E. (1998), Shortest Path Algorithms: An Evaluation using Real Road Networks. Transportation Science Vol. 32, No. 1, February 1998. 1998 Institute for Operations Research and the Management Sciences

Zhou H. (2003), Efficient Steiner Tree Construction Based on Spanning Graphs. ISPD’03, April 6–9, 2003, Monterey, California, USA. Copyright 2003 ACM 1-58113-650-1/03/0004

APPENDIX

FIG. MAP OF KWAHU SOUTH DISTRICT



SOURCE : KWAHU SOUTH DISTRICT ASSEMBLY — MPRAESO

Pseudocode for Prim’s Algorithm

Given a connected weighted graph G = (V, E) with a weight function w and a minimum spanning tree T can be derived from the code below.

```
for any $v \in V$
    cost [v] ← ∞
    parent [v] ← NULL
r ← arbitrary vertex of V
cost [r] ← 0
Q ← V
while $Q \neq \{ \}$
    u ←  extract Min (Q)
    for each $v \in$ adja (u) do
        if $v \in Q$ and w (u, v) < cost [v] then
            parent [v] ← u
            cost[v] ← w (u, v)
T ← {(v, parent [v]| $v \in V – \{r\}$
Return T
```

Pseudocode of Kruskal's Algorithm

```
Kruskal (G; w)
A = Ø
For each vertex $v \in G.V$
    Make – Set (v)
Put the edges of G.E into list sorted by non decreasing weight
For each $(u, v) \notin G.E$ taken from the sorted list
    If find-set (u) $\neq$ find-set (v)
        $A = A \cup \{(u, v)\}$
        Union (u; v)
Return A
```