Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 42, S. 180-186, Ekim 2022

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue. 42, pp. 180-186, October 2022

Research Article

Optimization of Gas Distribution in Istanbul Using Minimum Spanning Tree with Probabilistic Approach

Mehmet Savsar[1], Abdulrahman A. Hetari[2]

[1]Uskudar University, Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, İstanbul, Turkey. msavsar@gmail.com
[2]Uskudar University, Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, İstanbul, Turkey. a.556688@hotmail.com

(2nd International Conference on Engineering and Applied Natural Sciences ICEANS 2022, October 15-18, 2022, Konya, Turkey)

(**DOI:** 10.31590/ejosat.1202232)

ATIF/REFERENCE: Savsar, M. & Hetari, A. A. (2022). Optimization of Gas Distribution in Istanbul Using Minimum Spanning Tree with Probabilistic Approach. *European Journal of Science and Technology*, (42), 180-186.

Abstract

The minimum spanning tree (MST) is a method that aims to find the minimum distance for connecting all vertices or nodes in a network of cities or any other centers. For finding the minimum distance that connects all nodes, two different types of algorithms have been commonly used: Kruskal and Prim algorithms. Our objective in this research is to find the most cost-efficient process for linking the natural gas pipeline's pathways in Istanbul using minimum spanning tree with randomness or uncertainty in the distances between different districts, which are represented by vertices. The goal is to minimize total pipeline distance which connects all districts in metropolitan area of Istanbul. The method proved to be very useful in solving this problem.

Keywords: Minimum Spanning Tree, Prim Algorithm, Kruskel Algorithm, Gas Distribution Network, Optimization.

İstanbul Gaz Dağıtımının Olasılıksal Yaklaşımla Minimum Yayılma Ağacı Kullanılarak Optimizasyonu

Öz

Minimum yayılan ağaç (MST), bir şehir ağındaki veya diğer merkezlerdeki tüm köşeleri veya düğümleri birbirine bağlamak için minimum mesafeyi bulmayı amaçlayan bir yöntemdir. Bütün merkezleri birbirine bağlayan minimum mesafeyi bulmak için yaygın olarak iki farklı algoritma türü kullanılmıştır: Kruskal ve Prim algoritmaları. Bu araştırmadaki amacımız, İstanbul'daki doğal gaz boru hattının güzergâhlarını, köşelerle temsil edilen farklı ilçeler arasındaki mesafeleri, rastgelelik veya belirsizlik altında minimum yayılma ağacı kullanarak en uygun maliyetli birbirine bağlamaktır. Hedef, İstanbul metropoliten bölgesindeki tüm ilçeleri birbirine bağlayan toplam boru hattı mesafesini en aza indirmektir. Burada açıklanan yöntem, bu sorunu çözmede çok yararlı olduğunu kanıtladı.

Anahtar Kelimeler: Minimum Yayilan Ağaç, Prim Algoritması, Kruskel Algoritması, Gaz Dağıtım Şebekesi, Optimizasyon

1. Introduction

The natural gas is essential for generating various useful resources in all countries. Generating electricity or providing ddistributed or central heating for homes and other facilities in cities essentially depends on the gas distribution network. The distribution of natural gas is defined as moving the gas from the distribution point to various consumer points in a network. There has been so many researches for design of an optimal natural gas network distribution through the use of minimum spanning tree method, all with varying results of efficiency depending in the measurement and novelty of the distribution points in the graph. References (1-5) provide some related research papers.

Our aim is to distribute the natural gas between 30 district points in the metropolitan city of Istanbul, using the minimum spanning tree (MST) technique, so that we can get the most optimized pathways in a very cost efficient manner. The consumers in those 30 districts are in total of about 4.6 million according to a report from the main natural gas company in Turkey (IGDAŞ).

There are three types of natural gas systems: Linear, Tree structured and Cyclic. According to our algorithms result we will find out which system suits the pipeline pathways. There are several algorithms that exist and can be used to get different degrees on effective results, namely Kruskal, Prim and Boruvka algorithms. We will be utilizing the prim algorithm, due to the fact that the tool we use is limited in regards to the available algorithms to be processed.

The prim algorithm was founded in 1957 by Robert C. Prim and rediscovered by Edsger Diijkstra in the same year. And thus, the algorithm is also known as DJP algorithm. In this paper we will try to cover the 30 districts in Istanbul and construct an MST that nets us a suitable algorithm result.

In the present day, our issues deal with the distribution of gas throughout a city, utilizing it as one of the main sources of energy and consumed for various other purposes. Using an MST procedure, we attempt to connect every district of Istanbul based on minimum total distance resulting in minimum total cost.

MST has been used for optimizing connection points between several distribution points, various path-line systems such as electrical pathway, water supply pipelines, transportation routes and Network lines.

Our goal when using the MST is to optimize the length of the route, or path lines of the pipes that are distributing from its nodes towards the consumers. Two of the methods of calculating distribution probabilities are utilized to find the most accurate and precise distance, namely normal distribution and uniform distribution. In this paper, we will discuss in detail the methodology used and how we optimized our results

2. Methodology

The Minimum spanning tree consists of nodes and lines, where each nodes represents a district in the city and the lines are the distances between pairs of specific districts. In a gas distribution network, connecting lines between the nodes end up forming a loop or a cycle in the network if they end up connecting the node to itself. We can denote the nodes by N and the lines by L in the network (N, L) that we are creating. In the pipeline network, the natural gas flows in an oriented direction, i.e. in one directions. If all the pipelines are connected, then such graph would be called a directed graph.

We have 30 districts and as such 30 nodes to work with, and we have 52 lines connecting these nodes. A spanning tree has the lines equal to the total number of the nodes subtracted by 1, i.e., $L=N-1$. Each directed graph has too many spanning trees, through the use of the algorithms you get the optimum network for distribution, resulting in an MST.

We obtained the fixed distances between district centers using google map's "measure distance" feature. This is a feature in Google maps to measure the straight line distance between two points that one chooses. It is not very precise, yet simple to use. You have to access it by right clicking on the point you would like to start measure form and choose "Measure Distance". Figure 1 illustrates the measuring process. We tried to be as accurate as possible, there was no maps that showed the distribution network of the current gas pipelines in Turkey, most likely for security reasons. Thus, we used some of the main centers of IGDAS service in all the relative districts as a reference for an estimation of their exact distribution location.

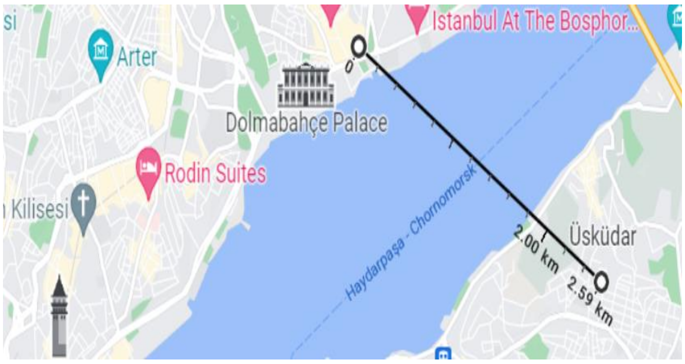

Figure 1. "Measure distance" command used in google maps

Since the distances were not accurate and precise, we assumed 20% less or more error value by utilizing some probabilistic concept to convert the actual constant distances into random values. This was unavoidable as we did not have any other way to get exact values. We applied to different probability distributions, which were normal and uniform distributions to represent the distances. For the uniform distribution related probability distribution function (pdf) and cumulative distribution function (cdf) formulas are as below. Figures 2 and 3 ilustrate related functions.

$$f(x) = 1 / (b - a) \qquad a \leq x \leq b$$

$$F(x) = \int f(x)dx=(x-a)/(b-a) \qquad x \geq a$$

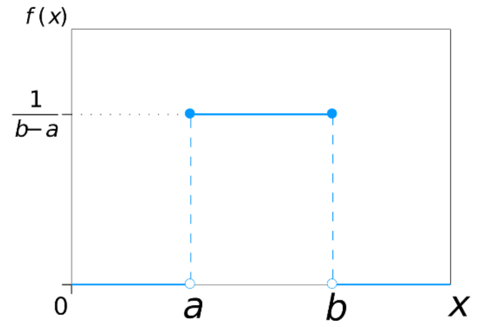

Fıgure 2. Uniform probability density function

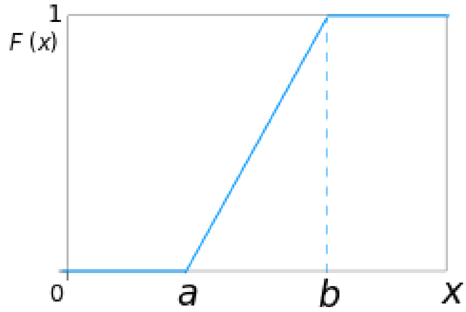

Figure 3. Uniform probability cumulative function

Next, we generated a uniform random variable X between a and b, which represents the distance between a and b, using inverse transformation. To to generate the random distance X between a and b, the cumulative distribution function which is between 0 and 1 ia set to a uniform random number between 0 and 1 (denoted as U) and by inverse transform, we determine X as a function of U.

F(x) = (x - a)/ (b - a) = U,

where U can be found using excel function RAND()

U = RAND ()

We can calculate the uniform distribution using this formula, for the 52 arcs we have. The random number generation has been repeated for 16 times, i.e., a sample of 16 random networks are generated and the MST slgorithm is applied to each to determine the average total distance that is best representative of the system. The random distance is obtained using the inverse transformation as follows:

x = a + U(b - a)

We find the optimum distance by trying 16 iterations and determining the MST for each followed by their average with the least error coefficient.

In case if the normal distribution is assumed for the random distances between districts, the fixed distance obtained by google is assumed to be the mean value and the standard deviation sigma is set as 20% of the mean.

σ = 0.2 * ($\bar{x}$)

Figure 4 shows the normal probability distribution function, for which the integral of the cumulative, and consequently its inverse is not straightforward. Fortunately, excel has a function to generate random variates from normal distribution as: NORM.INV(Probability, Mean, Std. Dev.), where probability is the U value. The function returns a random number from normal distribution with the specified mean and the standard deviation. The random distances are generated according this procedure for the complete gas distribution network.

We generate Distance=X_{ij} between all pairs of centers i and j. This procedure or sampling is repeated for 16 iterations. Finally, we substitute the values of X_{ij} into the subgraph, and apply the prim algorithm so that we find the MST for each sample of n=5 cases.

Next’ the average MST and the 95% confidence interval on the average is determined. Th upper confidence limit (UCL) and the lower confidence limit (LCL) values are calculated as follows:

UCL = $\bar{x}$ + 1.96(sigma/Sqrt(n))

LCL = $\bar{x}$ - 1.96(sigma/Sqrt(n))

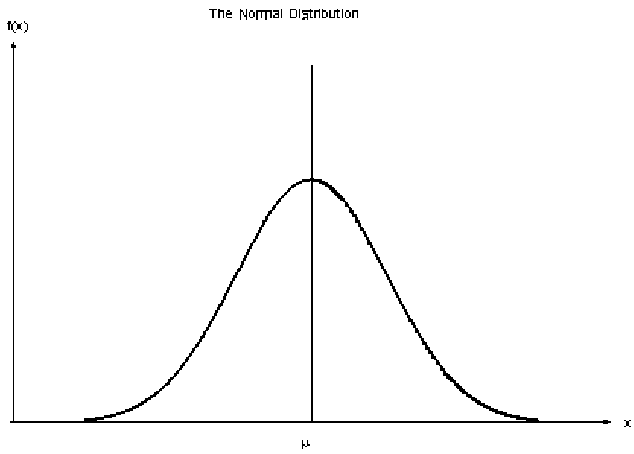

Figure 4. Normal probbility distribution function

After applying and processing our values we are confident by 95% of the accuracy of our distance measurements.. We decided to use Tuzla as the central point of distribution due to its proximity in relevance to the other nodes. And then we run the prim algorithm to get the MST. Figure 5 is a sample case example.

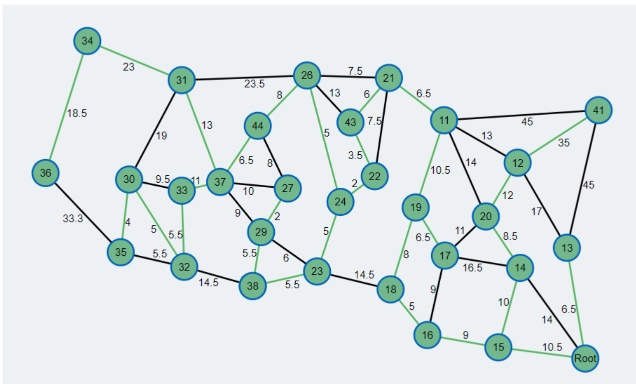

Figure 5. On network example and the MST (dark lines)

The procedure can be as outlined in steps as follows:

Step 1. Determine the X_{ij} distances between pairs of district centers for all centers using the Google Map measure feature.

Step 2. Generate random distance values using uniform or normal distribution by assuming the non-precise constant distances as base. In case of uniform distribution, assume the limits or parameters of the distribution a and b as ±20% of the constant distance, respectively and in case of normal distribution, assume the mean, μ=Constant Distance value and standard deviation σ=20% of the constant distance value.

Step 3. Determine the MST for the distribution network with random distance values.

Step 4. Repeat step 2 and 3 for enough iterations as sample size n. Find total minimum distance MST for each sample.

Step 5. Determine the average total minimum distance based on the samples of n minimum spanning threes (MST) and construct a confidence interval on the mean value.

3. Analysis of Results

As it was mentioned in the previus section, we generate a distribution network with distances between pairs of centers using the google map first. Next, we use the Prim algorithm to determine the MST for the generated network. The MST is obtained by using the algorithm available in the web through the link: 'https://algorithms.discrete.ma.tum.de/graph-algorithms/mst-prim/index_en.html'. Figure 6 illustrates a sub-weighted graph determined.

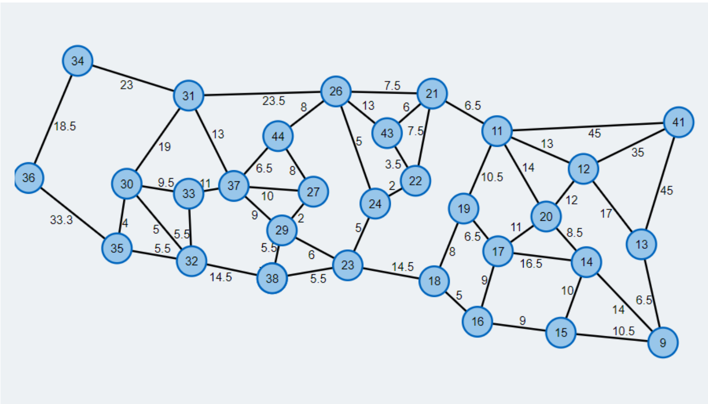

Figure 6. Districts-to-districts sub weighted graph of Istanbul. MST graph for the fixed distance measurements

These values between the nodes are the fixed distances between the districts which have been obtaned through the use of google maps. The total weight of the MST graph for the gas distribution network using the fixed values, is the sum of all the edges which equals 246 km. Since we are not quite sure of the precision of our values, we use the probability distribution to generate distances.

MST: Normal Distribution Case

As it was explained in previous section, distances are generated based on normal distribution. Table 1 illustrates random distance generation for three distance values (NORM1 to NORM3), five samples for each case.

Table 1. Random generation of distances (normal distribution)

1	min	5.2		11.2		8.4	
2	max	7.8		16.8		12.6	
3	mean	6.5		14		10.5	
4	sigma	1.3		2.8		2.1	
5		RAND 1	NORM 1	RAND 2	NORM 2	RAND 3	NORM 3
6							
7		0.8098	7.64	0.0458	9.28	0.7313	11.79
8		0.2893	5.78	0.1872	11.51	0.7807	12.13
9		0.8630	7.92	0.0606	9.66	0.3304	9.58
10		0.2374	5.57	0.3169	12.67	0.2342	8.98
11		0.5244	6.58	0.8577	17.00	0.0503	7.05

Figure 7 shows the related network for the first sample case. The total weight of the MST for the 1st iteration (sample 1), is the sum of all the edges which equals 250.49km. Figure 8 shows the MST for the second sample or the 2nd iteration of random distribution network. Figure 9 shows the distribution network and the MST for the 3rd sample; Figure 10 shows the MST for the 4th and Figure 11 shows the MST for 5th sample random network generated.

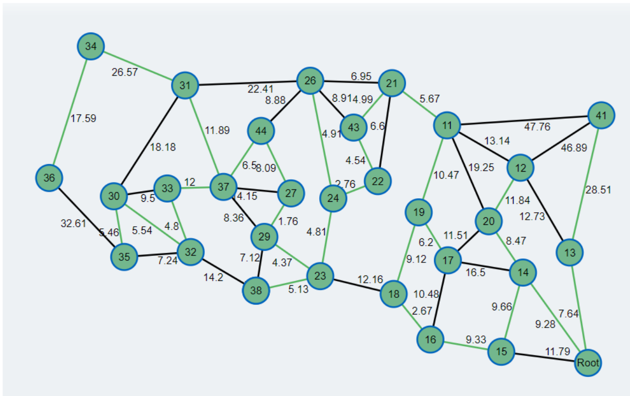

Figure 7. MST for the network based on 1st sample of random distances for normal distribution (MST=250.49 km)

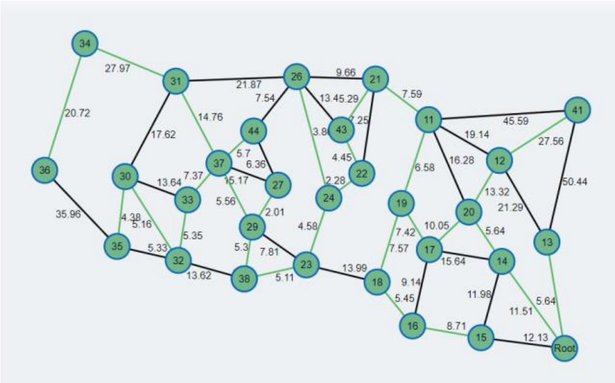

Figure 8. MST for the network based on 2nd sample of random distances for normal distribution (MST=246.61 km)

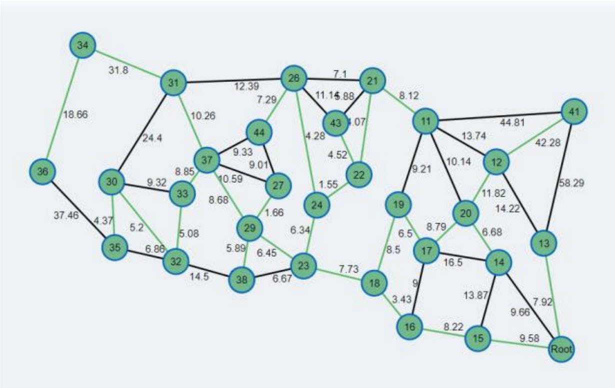

Figure 9. MST for the network based on 3rd sample of random distances for normal distribution (MST=264.52 km)

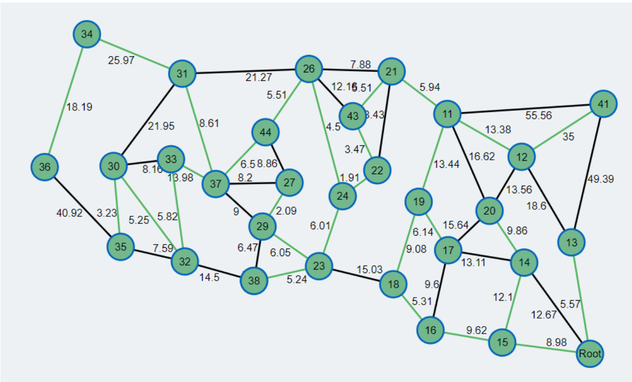

Figure 10. MST for the network based on 4th sample of random distances for normal distribution (MST=278.9km.

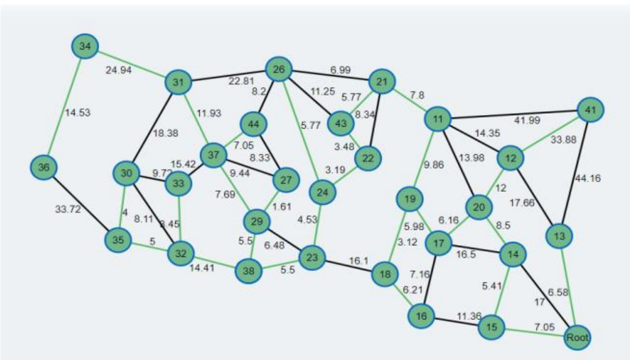

Figure 11. MST for the network based on 5th sample of random distances for normal distribution (MST=247.06 km)

MST: Uniform Distribution Case

As it was explained in previous section for uniform distribution, distances are generated based on cumulative uniform distribution function. Table 2 illustrates random distance generation for three district pair distance values, five samples for each case. A and B are the minimum and maximum values or the range of uniform distribution set by ±20% of the constant value obtained from Google map. Figure 12-16 show the related networks and MST for each sample of 5.

Table 3 summarizes all the results obtained by optimization for both the case of normal distribution and the uniform distribution. The table shows the minimum total distance, which connects all city centers, obtained by the MST Prim algorithm. Furthermore, the average total distance and the standard deviation values are also calculated. The number of samples (sample size) was kept small, because it was very much time consuming to construct the distribution network for each sample of distance generated. However, the average and the standard deviation indicate that the sample size was sufficient to represent the data and to determine a confidence interval on the mean. A 95% confidence interval was calculated on the mean for the each case using the formula below:

LCL and UCL = $\mu \pm Z_{\alpha/2} s\ S/n^{1/2}$

Values are given in Table 3. The confidence limits assure that we are 95% sure that the true mean is within these upper and lower limits.

Table 2. Random generation of distances (uniform distribution)

A	8.4			36			36	
B	12.6			54			54	
MEAN	10.5			45			45	
	RAND 3	X 3		RAND 4	X 4		RAND 5	X 5
	0.1067	8.85		0.9624	53.32		0.7349	49.23
	0.0553	8.63		0.8255	50.86		0.6186	47.13
	0.8872	12.13		0.7562	49.61		0.2464	40.43
	0.8477	11.96		0.9892	53.81		0.5739	46.33
	0.0444	8.59		0.3088	41.56		0.7413	49.34

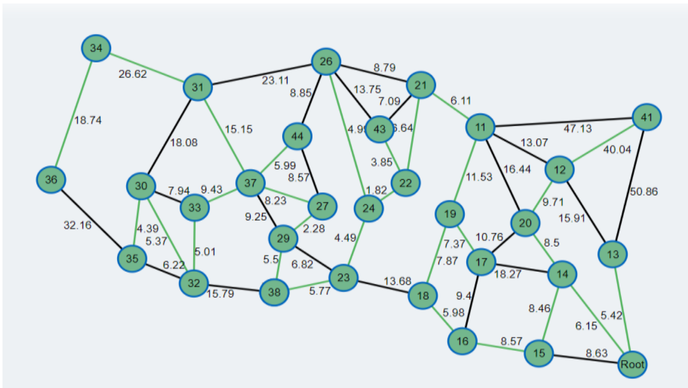

Figure 12. MST for the network based on 1st sample of random distances for uniform distribution (MST=273.46 km)

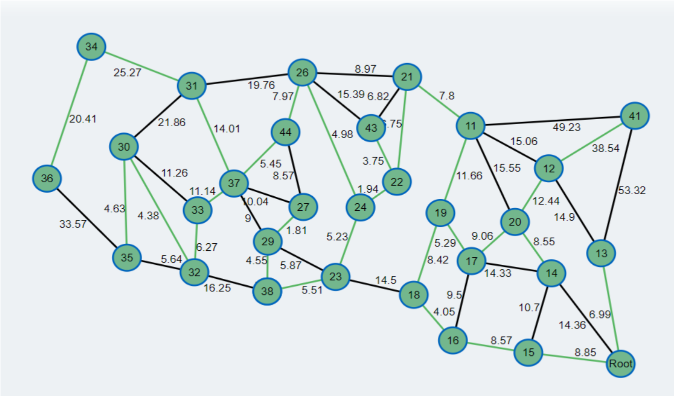

Figure 13. MST for the network based on 2nd sample of random distances for uniform distribution (MST=269.29 km)

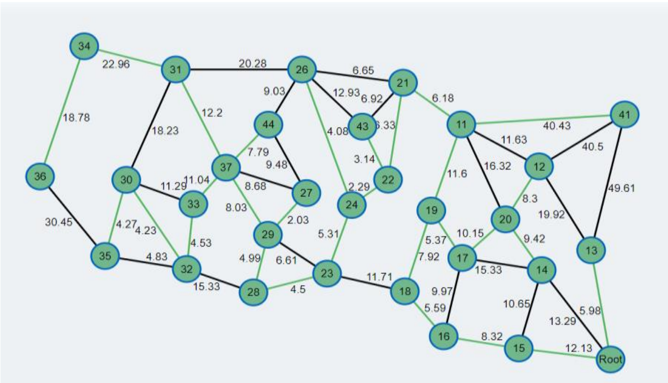

Figure 14. MST for the network based on 3rd sample of random distances for uniform distribution (MST=257.89 km)

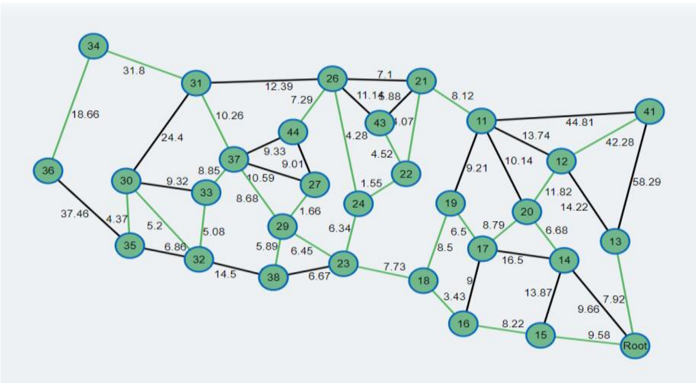

Figure 15. MST for the network based on 4th sample of random distances for uniform distribution (MST=264.70 km)

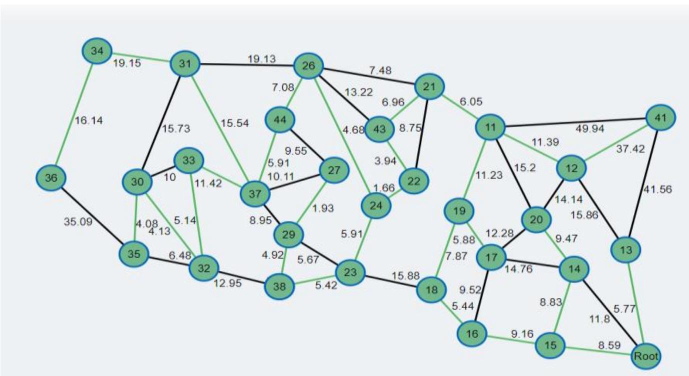

Figure 16. MST for the network based on 5th sample of random distances for uniform distribution (MST=263.39 km)

Table 3. Minimum total distance connecting all centers by MST

MSTs	Minimum total distance (MST) (normal)	Minimum total distance (MST) (uniform)
MST 1	250.49km	273.46km
MST 2	246.61km	269.29km
MST 3	264.52km	257.89km
MST 4	278.9km	264.70km
MST 5	247.06km	263.39km
Mean	257.52 km	265.75 km
Standard deviation	14.00 km	5.93 km
95% Lower Conf. Limit	245.24	260.55
95% Upper Conf. Limit	269.79	270.94

4. Conclusions

Through the use of random variates with normal and uniform distributions, we could get more reasonable distance measurements for the edges in the network to be processed by the Prim algorithm of the MST. Taking a sample of 5 for each case of normal and uniform distribution, gas distribution network is constructed and sub-weighted graphs were processed by Prim's algorithm. One of the principles of the minimum spanning tree is that the number of edges should be equal to Number of Nodes–1, or (n - 1). In our case, it was 30-1=29 edges. MST was obtained for each sample of random network constructed.

Introduction of probabilistic measures in estimating distances between districts proved to be extremely useful in dealing with uncertainties in measurements. The total distance of the MST with fixed value came out as 246 km. Taking 5 samples by generating different random networks, the average total minimum distance came out to be 257.52 km with a standard deviation of 14 km, and 95% lower and upper confidence limits of 245.24 km and 269.79 km respectively for normally distributed distances. In case of uniformly distributed distances, the average of five samples came out to be 265.75 km with a standard deviation of 5.93 km and 95% lower and upper confidence limits of 260.55 km and 270.94 km respectively. The confidence limits provide us a 95% assurance that the average total minimum distance would be between specified limits. This is a more meaningful information as compared to using constant values that we have no assurance about their accuracy.

In conclusion, the procedure outlined in this paper can be useful for the gas distribution companies to establish distrition networks with minimum total dstance, which direcly results in minimum total distibution costs.

References

1. I. (Mahdavi, N. Mahdavi-Amiri, A. Makui, A. Mohajeri, 2010), R. Tafazzoli, "Optimal gas distribution network using minimum spanning tree".
2. Minimum Spanning Tree of City to City Road Network in Nigeria Effanga,(E.O. & Edeke, Uwe. E,2013) Department Of Statistics, University Of Calabar, Nigeria
3. Nahla, M. A. I. (2011). Optimal Spanning Trees. African Institute for Mathematical Science.http://www/users. aims.ac.za/nahla/nahla.pdf
4. Mahdavi, I., ahdavi-Amiri, N., Makui, A., Mohajeri, A. (2022) "Optimal gas distribution network using minimum spanning tree", 17Th International Conference on IE&EM (IEEE).
5. Temnikov, D.O. and Suslov. D.Y. (2018) Algorithms of the Minimum Spanning Tree in Gas Supply Systems", J. Phys.: Conf. Ser. 1066-012001