

On Bounded Diameter MST Problem on Random Instances

Edward Kh. Gimadi^{1,2}, Alexey M. Istomin¹, and Ekaterina Yu. Shin²

¹ Sobolev Institute of Mathematics,

4 Acad. Koptuyug avenue, 630090, Novosibirsk, Russia

² Novosibirsk State University,

2 Pirogova Str. 630090, Novosibirsk, Russia

gimadi@math.nsc.ru, alexeyistomin@gmail.com, katherine15963@gmail.com

Abstract. We give an approximation deterministic algorithm for solving the Random bounded diameter minimum spanning tree (BDMST) problem on an undirected graph. The algorithm has a quadratic time complexity. A probabilistic analysis was performed under conditions that edge weights of given graph are identically independent uniformly distributed random variables on an interval $(a_n; b_n)$. Conditions of asymptotic optimality are presented.

Keywords: Graph · Bounded diameter minimum spanning tree · Minimum spanning tree · Asymptotically optimal algorithm · Probabilistic analysis · Performance guarantees · Random inputs · Uniform

1 Introduction

The Minimum Spanning Tree (MST) problem is a one of the classic discrete optimization problems. Given undirected weighted graph $G = (V, E)$, MST is to find a spanning tree of a minimal weight. MST is polynomially solvable, there are classic algorithms by Boruvka (1926), Kruskal (1956) and Prim (1957). These algorithms have complexity $\mathcal{O}(n^2)$ and $\mathcal{O}(M \log n)$ where $M = |E|$ and $n = |V|$.

In current paper a modification of the classical MST is studied. We study a bounded diameter minimum spanning tree problem (BDMST). The goal is to find in the graph G_n a spanning tree T_n of minimal total weight having its diameter limited by given number d . The diameter of a tree is the number of edges on the longest path between two leaves in the tree. This problem is *NP*-hard in the common case [10].

The Bounded Diameter Minimum Spanning Tree Problem has many practical applications in various fields such as telecommunication networks and linear

light wave network design [5], bit compression for information retrieval [7] and distributed mutual exclusion [28, 30, 31, 34].

A good example of usage is the Distributed Mutual Exclusion algorithms. Here we have a computer network of k computers and the internal communication is done by sending messages between computers along a tree. Only One Computer is Allowed to Enter a Critical Section. If Some Computer Wants the Right to Enter a Critical Section It must request it by sending a message to the computer which currently has this right. The time of this request depends on the number of edges in the path to the computer with the right. The goal is to build a communication tree with the minimal cost and bounded time of communication. And the solution is exactly the Bounded Diameter Minimum Spanning Tree.

Techniques for solving the BDMST problem may be classified into tree categories: exact methods, heuristic methods with experimentally measured performance ratio and algorithms with guaranteed performance ratio.

There are exact approaches for solving the BDMST problem based on mixed linear integer programming [2], [17] and 0-1 integer linear programming based branch and cut approaches [18]. But, these approaches could only be used to solve small problem instances, like complete graphs with less than 100 nodes.

As for the heuristic methods with experimentally measured performance ratio, there was presented [1] a greedy heuristic algorithm - the One Time Tree Construction (OTTC) for solving the BDMST problem followed by its modification [27], called Randomized Greedy Heuristics (RGH). Later it was also studied and extended in [32] and [6]. Genetic algorithms for solving BDMST problems were considered as well [26], [21], [22], [23]. Local search approaches were considered in [19], [20].

There were not really much attempts to solve BDMST by algorithms with guaranteed performance ratio. A study was done in [3], however the proof is not really easy to follow. In this paper we give the first approximation deterministic polynomial time algorithm for solving the Random DBMST on an undirected graph.

In the papers [15, 14] this problem was studied with a graph diameter bounded from below. In the current paper we consider the problem with a graph diameter bounded from above. We introduce a polynomial-time algorithm to solve this problem and provide conditions for this algorithm to be asymptotically optimal. A probabilistic analysis was performed under conditions that edges weights of given graph are identically independent distributed random variables.

By $F_A(I)$ and $OPT(I)$ we denote respectively the approximate (obtained by some approximation algorithm A) and the optimum value of the objective function of the problem on the input I . An algorithm A is said to have *performance guarantees* $(\varepsilon_A(n), \delta_A(n))$ on the set of random inputs of the problem of the size n , if

$$\Pr\{F_A(I) > (1 + \varepsilon_A(n))OPT(I)\} \leq \delta_A(n), \quad (1)$$

where $\varepsilon_A(n)$ is an estimation of *the relative error* of the solution obtained by algorithm A, $\delta_A(n)$ is an estimation of *the failure probability* of the algorithm,

which is equal to the proportion of cases when the algorithm does not hold the relative error $\varepsilon_A(n)$ or does not produce any answer at all.

Following by [13], we say that an algorithm A is called *asymptotically optimal* on the class of instances of the problem, if there are exist such performance guarantees that $\varepsilon_A(n) \rightarrow 0$ and $\delta_A(n) \rightarrow 0$ as $n \rightarrow \infty$. Apparently, judging by the review article [33], the first examples of asymptotically optimal algorithms were presented in the works [11, 12] for the traveling salesman problem on random input data.

Let's denote $UNI(a_n; b_n)$ a class of complete graphs with n vertices where edge weights are independent identically distributed random variables with uniform distribution on an interval $(a_n; b_n)$.

Frieze shown that the mathematical expectation of weight of classic MST on a random graph can be unexpectedly small. So for example on a complete graph with weights of edges from class $UNI(0; 1)$, the weight of a MST w.h.p. (with high probability) is close to the constant 2.02 ... [9].

As it was said in the papers [14, 15] the MST was studied with a graph diameter bounded from below. In [14] presented an asymptotically optimal algorithm \tilde{A} with time-complexity $O(n^2)$ for graphs which belong to $UNI(a_n; b_n)$ -class. On the first stage algorithm \tilde{A} build a d -vertex path P , using the greedy strategy "Go to the nearest unexplored vertex", starting from an arbitrary vertex. On the second stage in a graph G with edge weights equal to a_n for all $e \in P$, by means of Prim's algorithm [25], a spanning tree of a minimal weight is built. This tree is taken as a solution.

On graphs which belong to $UNI(a_n; b_n)$ class algorithm \tilde{A} has the following performance guarantees:

$$\epsilon_n = \mathcal{O}\left(\frac{b_n/a_n}{n/\ln \frac{n-1}{n-d}}\right), \delta_n = e^{-0.25(n-d)}.$$

Thus, the sufficient conditions for the asymptotic optimality of the algorithm \tilde{A} are

$$b_n/a_n \leq \frac{n}{\ln n}, \quad d = o(n).$$

Next, let's proceed to the description of the algorithm for solving the BDMST problem.

2 An Algorithm \mathcal{A} for Finding a Bounded Diameter MST

Let d be a parameter exceeding the tree diameter.

Stage 1. Arbitrary select d vertices subset from V , let's denote selected subset V_1 . Using the Prim's algorithm [25] construct in the graph $G[d]$ induced by these d vertices a minimum spanning tree T_0 : edge by edge grow up tree by edges e_1, \dots, e_{d-1} . Obviously, its diameter is smaller than the parameter d . Put $V_2 = V \setminus V_1$.

Stage 2. Every vertex $u \in V_2$ is connected by the shortest possible edge with a vertex $v \in V_1$. As a result we obtain an n -vertex tree $T_{\mathcal{A}}$ which is an approximate solution of the problem.

A comment. If the constructed tree T_0 has the form of a chain, then the nearest vertices v are selected from the set $V_1 \setminus v'$, where v' is one of the two end vertices of the chain.

Finally we built the spanning tree $T_{\mathcal{A}}$ with a diameter smaller than the parameter d .

Further, we denote by $W(G')$ the weight of the subgraph G' of the given graph G , by $W_{\mathcal{A}}$ the weight of the solution built by algorithm \mathcal{A} . Also by $\mathbf{E}X$ we will denote an expectation of a random variable X and by $\mathbf{Var}X$ its variance.

3 Analysis of Algorithm \mathcal{A}

The algorithm has polynomial complexity $O(n^2)$, since the construction of the tree T_0 in Stage 1 is done by the Prim's algorithm [25] in time $O((n-d)^2)$, and in the Stage 2 it takes about $d(n-d)$ comparison operations.

A probabilistic analysis we perform under conditions that graph edges weights are identically independent distributed random variables with uniform distribution on a set (a_n, b_n) , $0 < a_n \leq b_n < \infty$. Further we suppose that the parameter d is defined on the set of values d in the range $\ln n \leq d < n$.

Statement 1. The spanning tree $T_{\mathcal{A}}$ is restricted by a diameter not exceeding d , since on the second Stage the diameter of the tree T_0 can increase by no more than 1.

Statement 2. $W(T_{\mathcal{A}}) = W(T_0) + S$, where $S = \sum_{u \in V_2} \phi_u^d$, ϕ_u^d is a random variable equal to minimum from d identically independent distributed random variables from with uniform distribution on an interval $[a_n, b_n]$, $a_n > 0$.

Also, according to step 1 of the algorithm \mathcal{A} , weight of selected edge e_i is a random variable equal to minimum from i identically independent distributed random variables from with uniform distribution on an interval $[a_n, b_n]$, $a_n > 0$.

So,

$$W(T_0) = \sum_{i=1}^{d-1} \phi_{e_i}^d.$$

Or using variables ξ_u^d , where $\xi_u^d = \frac{\phi_u^d - a_n}{b_n - a_n}$, distributed on $[0, 1]$, we get

$$W(T_0) = (d-1)a_n + (b_n - a_n) \sum_{i=1}^{d-1} \xi_{e_i}^d = (d-1)a_n + (b_n - a_n)W(T_0)'$$

We denote $W(T_0)' = \sum_{i=1}^{d-1} \xi_{e_i}^d$.

Denote $S' = \sum_{u \in V_2} \xi_u^d$. Obviously,

$$S = (n-d)a_n + S'(b_n - a_n).$$

ξ_u^d is a random variable equal to minimum from d identically independent distributed random variables with uniform distribution on an interval $[0, 1]$.

We have

$$W(T_{\mathcal{A}}) = (n - 1)a_n + (b_n - a_n)(W(T_0)' + S').$$

Statement 3. We can estimate the expectation of $W(T_0)'$ by the sum

$$\frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{d} \leq \ln d$$

for the expectation of the length $\mathbf{E}Ch \leq \ln d$, where Ch is a chain in $G[d]$, obtained by the greedy procedure "Go to the nearest city". So we have $W(T_0)' \leq h \ln d$ w.h.p., where the constant h is large 1.

Let λ_n be a positive constant. Denote

$$\varepsilon_n = \frac{b_n (h \ln d + (1 + \lambda_n)\mathbf{E}S')}{a_n (n - 1)}. \tag{2}$$

Statement 4.

$$\Pr\{W_{\mathcal{A}} \leq (1 + \varepsilon_n)OPT\} \geq 1 - \delta_n, \tag{3}$$

where

$$\delta_n = \Pr\{S' > (1 + \lambda_n)\mathbf{E}S'\}. \tag{4}$$

Proof.

$$\frac{W_{\mathcal{A}}}{OPT} \leq \frac{(n - 1)a_n + (b_n - a_n)(W(T_0)' + S')}{(n - 1)a_n} \leq 1 + \frac{b_n (h \ln d + S')}{a_n (n - 1)}.$$

By virtue of formulas (2) and (3) the inequality can be continued with the probability $1 - \delta_n$.

$$\frac{W_{\mathcal{A}}}{OPT} \leq 1 + \frac{b_n (h \ln d + S')}{a_n (n - 1)} \leq 1 + \frac{b_n (h \ln d + (1 + \lambda_n)\mathbf{E}S')}{a_n (n - 1)} = 1 + \varepsilon_n.$$

The Statement 4 is proved.

Statement 5.

$$\mathbf{E}S' = \frac{n - d}{d}.$$

Proof. S' is equal to the sum of $n - d$ random independent identically distributed variables each of them equal to minimum over $d - 1$ uniformly distributed on a segment $[0, 1]$ variables.

Using the statement 5 we have the following expression for the ε_n

$$\varepsilon_n = \frac{b_n (h \ln d + (1 + \lambda_n)(n - d)/d)}{a_n (n - 1)} \leq \frac{b_n}{a_n} \left(\frac{h \ln d}{n - 1} + \frac{1 + \lambda_n}{d} \right). \tag{5}$$

Next for the probabilistic analysis of Algorithm \mathcal{A} we need the following

Petrov's Theorem [24]. Consider independent random variables X_1, \dots, X_n . Let there be positive constants g_1, \dots, g_n and T such that for all $1 \leq k \leq n$ and $0 \leq t \leq T$

$$\mathbf{E}e^{tX_k} \leq \exp\left\{\frac{g_k t^2}{2}\right\}. \quad (6)$$

Put $S = \sum_{k=1}^n X_k$ and $G = \sum_{k=1}^n g_k$. Then

$$\mathbf{Pr}\{S > x\} \leq \begin{cases} \exp\left\{-\frac{x^2}{2G}\right\}, & \text{for } 0 \leq x \leq GT, \\ \exp\left\{-\frac{Tx}{2}\right\}, & \text{if } x \geq GT. \end{cases}$$

Theorem 1. Let the parameter d be defined so that

$$\ln n \leq d < n. \quad (7)$$

Then Algorithm \mathcal{A} solves the problem asymptotically optimal w.h.p.

Proof. We introduce a proof for two cases for a values of the parameter d : $\ln n \leq d < n/2$ and $n/2 \leq d < n - 1$.

Case 1: $\ln n \leq d < n/2$.

Put $\lambda_n = \sqrt{\frac{4 \ln n}{n}}$.

According to the formula (5)

$$\varepsilon_n \leq \frac{b_n}{a_n} \left(\frac{h \ln d}{n-1} + \frac{1 + \lambda_n}{d} \right).$$

We see that $\varepsilon_n \rightarrow 0$ under condition

$$\frac{b_n}{a_n} = o(d_n).$$

Now using Petrov's Theorem estimate the fault probability

$$\delta_n = \mathbf{Pr}\{S' > (1 + \lambda_n)\mathbf{E}S'\},$$

Put

$$T = \frac{d}{2};$$

$$G = \frac{n-d}{d^2};$$

$$x = \lambda_n \mathbf{E}S^0 = \lambda_n \frac{n-d}{d}.$$

The inequality $TG > x$ is satisfied. Indeed from

$$TG = \frac{1}{2} \frac{n-d}{d} > x = \lambda_n \mathbf{E}S' = \lambda_n \frac{n-d}{d}$$

it follows that

$$\frac{1}{2} > \lambda_n = \sqrt{\frac{4 \ln n}{n}}.$$

According to Petrov's Theorem, we have an estimate for the failure probability of the algorithm \mathcal{A} :

$$\delta_n = \Pr\{\hat{S}' > x\} \leq \exp\left\{-\frac{x^2}{2G}\right\}.$$

Now show that

$$\frac{x^2}{2G} \geq \ln n.$$

Indeed, since $n-d \geq \frac{n}{2}$, according to inequality (7), we get

$$\frac{x^2}{2G} = \frac{\left(\lambda_n \frac{(n-d)}{d}\right)^2}{2 \frac{(n-d)}{d^2}} = \frac{(n-d)}{2} \lambda_n^2 = \frac{n-d}{2} \left(\frac{4 \ln n}{n}\right) \geq \ln n.$$

From this it follows that

$$\delta_n = \Pr\{S' > x\} \leq \exp\left\{-\frac{x^2}{2G}\right\} \leq \exp(-\ln n) = \frac{1}{n} \rightarrow 0,$$

as $n \rightarrow \infty$. So in the Case 1 Algorithm \mathcal{A} solves the problem asymptotically optimal.

Case 2 : $n/2 \leq d < n-1$.

Put $\lambda_n = \ln n$.

According to the formula (5)

$$\varepsilon_n \leq \frac{b_n}{a_n} \left(\frac{h \ln d}{n-1} + \frac{1 + \lambda_n}{d} \right).$$

We see that within the values of the parameter d for the case 2, the expression in parentheses reaches a maximum at $d = n/2$. So $\varepsilon_n = O\left(\frac{b_n \ln n}{a_n n}\right)$ and $\varepsilon_n \rightarrow 0$ under condition

$$\frac{b_n}{a_n} = o\left(\frac{n}{\ln n}\right).$$

Now using Petrov's Theorem estimate the fault probability

$$\delta_n = \Pr\{S' > (1 + \lambda_n)\mathbf{E}S'\},$$

Put

$$\begin{aligned} T &= \frac{d}{2}; \\ G &= \frac{(n-d)}{d^2}; \\ x &= \lambda_n \mathbf{E}S^0 = \ln n \frac{n-d}{d}. \end{aligned}$$

The inequality $TG < x$ is satisfied.

$$TG = \frac{1}{2} \frac{(n-d)}{d} < x = \ln n \frac{(n-d)}{d}.$$

According to Petrov's Theorem, we have an estimate for the failure probability of the algorithm \mathcal{A} :

$$\delta_n = \Pr\{S' > x\} \leq \exp\left\{-\frac{Tx}{2}\right\}.$$

Now

$$\frac{Tx}{2} = \frac{d}{2} \ln n \frac{(n-d)}{d} \geq \ln n.$$

From this it follows that

$$\delta_n = \Pr\{\hat{S}' > x\} \leq \exp\left\{-\frac{Tx}{2}\right\} \leq \exp(-\ln n) = \frac{1}{n} \rightarrow 0,$$

as $n \rightarrow \infty$. So in the Case 2 Algorithm \mathcal{A} solves the problem asymptotically optimal as well.

Theorem 1 is completely proved.

4 Conclusion

It would be interesting to investigate (a) the Random BDMST problem on input data with infinite support like exponential or truncated-normal distribution, (b) the problem of finding several edge-disjointed spanning trees with a bounded diameter.

Acknowledgement. Authors are supported by the Russian Foundation for Basic Research (project 16-07-00168), and by the program of fundamental scientific researches of the SB RAS I.5.1.

References

1. Abdalla, A., Deo, N., Gupta P.: Random-tree diameter and the diameter constrained MST. In: Proceedings of Congress on Numerantium. pp. 161–182 (2000)

2. Achuthan, N.R., Caccetta, L., Caccetta, P., Geelen, A.: Computational Methods for the Diameter Restricted Minimum Weight Spanning Tree Problem. *Australian Journal of Combinatorics* 10, 51–71 (1994)
3. Angel, O., Flaxman, A. D., Wilson, D. B.: A sharp threshold for minimum bounded-depth and bounded-diameter spanning trees and Steiner trees in random networks. Preprint; arXiv:0810.4908v2 [math.PR] (2011)
4. Angluin, D., Valiant, L. G.: Fast probabilistic algorithms for Hamiltonian circuits and matchings. *Journal of Computer and System Sciences* 18(2), 155–193 (1979)
5. Bala, K., Petropoulos, K., Stern, T. E.: Multicasting in a linear lightwave network. In: *Proceedings of IEEE INFOCOM'93*. pp. 1350–1358 (1993)
6. Binh, H. T. T., Hoai, N. X., McKay, R. I. I.: A new hybrid genetic algorithm for solving the bounded diameter minimum spanning tree problem. In: *Proceedings of IEEE World Congress on Computational Intelligence, Hong Kong, LNCS* (2008)
7. Bookstein, A., Klein, S. T.: Compression of correlated bit. *Inf. Syst.* 16, 110–118 (1996)
8. Cooper, C., Frieze, A., Ince, N., Janson, S., Spencer, J.: On the length of a random minimum spanning tree. *Combinatorics, Probability and Computing* 25(1), 89–107 (2016)
9. Frieze, A.: On the value of a random MST problem. *Discrete Applied Mathematics* 10, 47–56 (1985)
10. Garey, M. R., Johnson, D. S.: *Computers and Intractability*. Freeman, San Francisco (1979)
11. Gimadi, E. Kh., Perepelitsa, V. A.: On a problem of finding minimal Hamiltonian circuit with weighted arcs. *Diskretni analiz, Novosibirsk*, 15, 57–65 (1969), (in Russian)
12. Gimadi, E. Kh., Perepelitsa, V. A.: An asymptotical approach to solving the traveling salesman problem. *Upravlyaemye sistemy, Novosibirsk*, 12, 35–45 (1974), (in Russian)
13. Gimadi, E. Kh., Glebov, N. I., Perepelitsa, V. A.: Algorithms with Estimates for Discrete Optimization Problems. *Problemy Kibernetiki* 31, 35–42 (1975), (in Russian)
14. Gimadi, E. Kh., Serdyukov, A. I.: A probabilistic analysis of approximation algorithm for tree spanning problem with a bounded from below diameter In: *Oper. Res. Proceed.* 99 (Inderfurth K. ed.). pp. 63–68. Springer, Berlin (2000)
15. Gimadi, E. Kh., Shin, E. Yu.: Probabilistic analysis of an algorithm for the minimum spanning tree problem with diameter bounded below. *Journal of Applied and Industrial Mathematics* 9(4), 480–488 (2015)
16. Gimadi, E. Kh., Istomin, A. M., Shin, E. Yu.: On algorithm for the minimum spanning tree problem bounded below. In: *Proc. DOOR 2016, Vladivostok, Russia, September 19-23, 2016. CEUR-WS. vol. 1623*, pp. 11–17 (2016)
17. Gouveia, L., Magnanti, T.L., Requejo C.: A 2-path approach for odd diameter constrained minimum spanning and steiner trees. *Network* 44 (4), 254–265 (2004)
18. Gruber, M., Raidl G.R.: A New 0-1 ILP approach for the bounded diameter minimum spanning tree problem. In: *Proceedings of the 2nd International Network Optimization Conference*. pp. 178-185 (2005)
19. Gruber, M., Raidl, G.R.: Variable neighbourhood search for the bounded diameter minimum spanning tree problem. In: *Proceedings of the 18th Mini Euro Conference on Variable Neighborhood Search, Spain* (2005)
20. Gruber, M., Hemert, J., and Raidl, G.R.: Neighbourhood searches for the bounded diameter minimum spanning tree problem embedded in a VNS, EA and ACO. In:

- Proceedings of Genetic and Evolutionary Computational Conference (GECCO-2006). pp. 1187–1194 (2006)
21. Julstrom, B.A., Raidl, G.R.: A permutation coded evolutionary for the bounded diameter minimum spanning tree problem. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2003). pp. 2-7 (2003)
 22. Julstrom, B.A.: Encoding bounded diameter minimum spanning trees with permutations and random keys. In: Proceedings of Genetic and Evolutionary Computational Conference (GECCO-2004). LNCS. vol. 3102, pp. 1272-1281 (2004)
 23. Nghia, N. D., Binh, H. T. T.: A new recombination operator for solving bounded diameter minimum spanning tree problem. In: Proceedings of RIVF-2007, LNCS (2007)
 24. Petrov, V. V.: Limit Theorems of Probability Theory. Sequences of Independent Random Variables. Clarendon Press, Oxford, 304 p. (1995)
 25. Prim, R. C.: Shortest connection networks and some generalizations. Bell System Tech. J. 36, 1389–1401 (1957)
 26. Raidl, G.R., Julstrom, B.A.: Edge-sets: an effective evolutionary coding of spanning trees. IEEE Transactions on Evolutionary Computation 7, 225–239 (2003)
 27. Raidl, G.R., Julstrom, B.A.: Greedy heuristics and an evolutionary algorithm for the bounded-diameter minimum spanning tree problem. In: Proceeding of the ACM Symposium on Applied Computing. pp. 747–752 (2003)
 28. Raymond, K.: A tree-based algorithm for distributed mutual exclusion. ACM Trans Comput Syst 7, 61–77 (1989)
 29. Roskind, J., Tarjan, R. E.: Note on finding minimum-cost edge-disjoint spanning trees. Math. Oper. Res. 10(4), 701–708 (1985)
 30. Satyanarayanan, R., Muthukrishnan, D. R.: A note on Raymond’s tree-based algorithm for distributed mutual exclusion. Inf Process Letters 43, 249–255 (1992)
 31. Satyanarayanan, R., Muthukrishnan, D. R.: A static tree-based algorithm for the distributed readers and writers problem. Comput Sci Inform 24, 21–32 (1994)
 32. Singh, A., Gupta, A.K.: An improved heuristic for the bounded diameter minimum spanning tree problem. Journal of Soft Computing 11, 911–921 (2007)
 33. Slominski, L.: Probabilistic analysis of combinatorial algorithms: a bibliography with selected annotations. Computing 28, 257–267 (1982)
 34. Wang, S., Lang, S. D.: A Tree-based Distributed Algorithm for the K-entry Critical Section Problem. In: Proceedings of the 1994 International Conference on Parallel and Distributed Systems. pp. 592–597 (1994)