



1988

# Approximations for the Random Minimal Spanning Tree With Application to Network Provisioning

Anjani Jain  
*University of Pennsylvania*

John W. Mamer

Follow this and additional works at: [http://repository.upenn.edu/oid\\_papers](http://repository.upenn.edu/oid_papers)

 Part of the [Other Mathematics Commons](#)

## Recommended Citation

Jain, A., & Mamer, J. W. (1988). Approximations for the Random Minimal Spanning Tree With Application to Network Provisioning. *Operation Research*, 36 (4), 575-584. <http://dx.doi.org/10.1287/opre.36.4.575>

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/oid\\_papers/99](http://repository.upenn.edu/oid_papers/99)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Approximations for the Random Minimal Spanning Tree With Application to Network Provisioning

## **Abstract**

This paper considers the problem of determining the mean and distribution of the length of a minimal spanning tree (MST) on an undirected graph whose arc lengths are independently distributed random variables. We obtain bounds and approximations for the MST length and show that our upper bound is much tighter than the naive bound obtained by computing the MST length of the deterministic graph with the respective means as arc lengths. We analyze the asymptotic properties of our approximations and establish conditions under which our bounds are asymptotically optimal. We apply these results to a network provisioning problem and show that the relative error induced by using our approximations tends to zero as the graph grows large.

## **Keywords**

facilities/equipment planning: network planning, networks/graphs, stochastic: random minimal spanning trees, tree algorithms: approximations to MSTs

## **Disciplines**

Other Mathematics

DTIC FILE COPY

1

Working Paper No. 352

AD-A204 656

Contract N00014-85-K-0143

APPROXIMATIONS FOR THE RANDOM MINIMAL SPANNING TREE WITH APPLICATION TO NETWORK PROVISIONING

by

Anjani Jain

and

John W. Mamer

February 1986  
Revised November 1986

DTIC  
SELECTED  
S FEB 27 1989 D  
D C6

DISTRIBUTION STATEMENT A  
Approved for public release  
Distribution Unlimited

WESTERN MANAGEMENT SCIENCE INSTITUTE  
University of California, Los Angeles

89 2 21 062

# Approximations for the Random Minimal Spanning Tree With Application to Network Provisioning

Anjani Jain\*      John W. Mamer†

February 1986

Revised November 1986



|                    |                                     |
|--------------------|-------------------------------------|
| Accession For      |                                     |
| NTIS CRA&I         | <input checked="" type="checkbox"/> |
| DTIC TAB           | <input type="checkbox"/>            |
| Unannounced        | <input type="checkbox"/>            |
| Justification      |                                     |
| By <i>per HP</i>   |                                     |
| Distribution       |                                     |
| Availability Codes |                                     |
| Dist               | Available for Special               |
| <i>A-1</i>         |                                     |

\*The Wharton School, University of Pennsylvania

†Graduate School of Management, University of California, Los Angeles

(1)

## Abstract

→ This paper considers the problem of determining the mean and distribution of the length of a minimal spanning tree (MST) on an undirected graph whose arc lengths are independently distributed random variables. We obtain bounds and approximations for the MST length and show that our upper bound is much tighter than the naïve bound obtained by computing the MST length of the deterministic graph with the respective means as arc lengths. We analyze the asymptotic properties of our approximations and establish conditions under which our bounds are asymptotically optimal. We apply these results to a network provisioning problem and show that the relative error induced by using our approximations tends to zero as the graph grows large. (KR) ←

A spanning tree is a connected, acyclic subgraph that spans (*i.e.*, includes) all the nodes of a given graph (see Harary [1969]; Lawler [1976] for definitions and properties). A minimal (maximal) spanning tree is a spanning tree that has the minimum (maximum) total edge weight, the minimum (maximum) being taken over the set of all spanning trees of the given graph.

We consider the problem of determining the probability distribution function (DF) and expected value of the length of a minimal spanning tree (MST) for an undirected graph with random arc lengths. An obvious approximation to the expected length of a MST in random graphs is obtained by replacing the arc-length random variables by their respective expectations and computing the MST length for the resulting deterministic graph. This approximation is an upper bound for the true expected length of MST. Our approximation provides a tighter bound. Our approach also enables us to approximate the DF of the MST length, and we prove that our approximation is asymptotically precise with probability one.

A classical application for the minimal spanning tree is the design of communication networks. As an application of our results we consider a probabilistic version of the network design problem. An order is to be placed for material to construct a communication network (a minimal spanning tree). At the time that the order is placed, the arc lengths are not known with certainty. Over- or under-supply of material results in increased costs. We seek an order level which minimizes the expected total cost. Our approach will provide asymptotically optimal solutions to this problem.

The organization of this paper is as follows. Section 1 introduces the stochastic spanning tree problem and our approach to its approximation. In section 2 we present some Monte Carlo results as well as an analytical characterization of the accuracy of our approximation. Section 3 proceeds to establish some asymptotic properties of our bounds, and section 4 presents an application of these bounds to the network provisioning problem and results pertaining to their asymptotic optimality. Finally, section 5 concludes the paper with some remarks on open questions.

# 1 The Stochastic Spanning Tree Problem

The stochastic MST problem can be formulated in several ways. Gilbert [1965] considered the problem of constructing minimal spanning trees to connect  $n$  points placed at random in the unit circle  $\|P\| < 1$  according to a Poisson process, where  $\|\cdot\|$  represents the "distance" of a point from the origin. Gilbert considered three different norms (Euclidean, Manhattan, and maximum) on Cartesian surfaces and obtained asymptotic bounds on the expected length of the MST. For all choices of the norm, he estimated this length to be asymptotic to  $c(\pi n)^{\frac{1}{2}}$  and showed that the constant  $c$  is less than  $2^{-\frac{1}{2}}$ . Steele [1981] considered growth rates of MST's on  $n$  points distributed in a Euclidean space, and Lueker [1981] obtained asymptotic results on expected lengths of maximum spanning trees when arc lengths drawn independently from the unit normal distribution. In a remarkable paper, Frieze [1985] recently established that when arc lengths on a complete graph are non-negative i.i.d random variables having a common DF  $F$ , which is differentiable at zero with  $F'(0) = D > 0$ , and has finite mean and variance, then the length of the MST tends to  $\zeta(3)/D$  in expectation as well as in probability, where  $\zeta(3) = \sum_{k=1}^{\infty} \frac{1}{k^3} = 1.202\dots$

The stochastic framework we posit for our problem is close in spirit to the classical MST problem in the deterministic case: the graph structure, *i.e.* the configuration of the nodes and arcs, is considered given; only the arc-lengths are random variables, each with its own probability distribution.

Let  $G = (N, A)$  denote a graph with  $N$  being the set of nodes, and  $A \subseteq N \times N$  being the set of undirected pairs of nodes called arcs. We shall always label the nodes of  $G$  from 1 through  $n$ , where  $n = |N|$ .

Let  $X_{ij}$  denote the random variable representing the length of the arc between node  $i$  and node  $j$  (since the arcs are undirected, we make no distinction between  $X_{ij}$  and  $X_{ji}$ ), and let  $F_{ij}$  denote the distribution function of  $X_{ij}$ . The  $X_{ij}$ 's are assumed independent.

Let  $\mathcal{T}$  denote the class of all spanning trees of  $G$ . By Cayley's formula (see Riordan [1978]), if  $G$  is complete (*i.e.* if all links are permitted), then

$$|\mathcal{T}| = n^{n-2}.$$

In general, let  $|\mathcal{T}| = M \leq n^{n-2}$ . Then,  $\mathcal{T} = \{T_i\}_{i=1}^M$  where each  $T_i$  in  $\mathcal{T}$  is a spanning tree of  $G$ . We shall assume that  $G$  is connected.

Let

$$\xi = \min_{T_i \in \mathcal{T}} \{Y_i\}, \quad (1)$$

where

$$Y_i = \sum_{(j,k) \in T_i} X_{jk} \quad (2)$$

is the length of the  $i$ th spanning tree. It is clear that  $Y_i$  and  $\xi$  are well defined random variables. The characterization of the random variable  $\xi$  is, in general, quite difficult. Our attention will be focused on obtaining approximations to  $\xi$  and its distribution function  $F_\xi$ .

The obvious approximation to the expected length of the MST is obtained by replacing the random variables by their expectations and then constructing the MST. Let

$$g = \min_{T_i \in \mathcal{T}} \{EY_i\} = \min_{T_i \in \mathcal{T}} \left\{ \sum_{(j,k) \in T_i} EX_{jk} \right\} \quad (3)$$

represent this approximation. It may be trivially established that  $g$  does provide an upper bound, *i.e.*,

$$E\xi \leq g. \quad (4)$$

The value of  $g$  can be easily computed by any of the several "greedy" algorithms (Kruskal [1956]; Prim [1957], Loberman and Weinberg [1957]).

We shall use the following version of the Prim algorithm (Aho *et al.* [1983], p.235) to construct an analytic device which we shall use to obtain a tighter bound for  $E\xi$ . The algorithm is described in pseudo-Pascal notation.



```

Procedure PRIM ( $G(N, A)$ :graph; var  $S$ :set of arcs;
                var  $q$ :length;  $w$ :arc weights);
{PRIM constructs an MST for  $G$ , the nodes in  $N$  are numbered from 1 to  $n$ }
  var
     $U$  : set of nodes;
     $i, j$  : vertex;
    PRIMNUM : array[1 ..  $n$ ] of integer;
  begin
     $q := 0$ ;
     $S := \phi$ ;
     $U := \{1\}$ ;
    PRIMNUM(1) := 1;
     $k := 2$ ;
    while  $U \neq A$  do
      begin
        find  $(i, j)$  such that  $w_{ij} = \min_{\substack{u \in U \\ v \in A \setminus U}} \{w_{uv}\}$ ;

        PRIMNUM( $k$ ) :=  $j$ ;
         $S := S \cup \{(i, j)\}$ ;
         $U := U \cup \{j\}$ ;
         $q := q + w_{ij}$ ;
         $k := k + 1$ 
      end
    end; {PRIM}

```

The algorithm PRIM constructs a MST with respect to the set  $\{w_{ij}\}_{(i,j) \in A}$  of arc weights and stores the MST arcs in the set  $S$ . In variable  $q$  it computes the "length" of the MST so obtained. It also records in the array PRIMNUM, the node numbers of the nodes of  $G$  in the order they were selected for inclusion in the set  $U$ .

We shall make use of PRIM to generate a new numbering of the nodes of  $G$  w.r.t a given set  $\{w_{ij}\}_{(i,j) \in A}$ , according to the order in which they were included in the minimal spanning tree. Let this new numbering of nodes be called the *Prim-numbering of  $G$  w.r.t  $\{w_{ij}\}_{(i,j) \in A}$* . It then follows from this definition that if  $\{1', 2', \dots, n'\}$  represents a Prim-numbering of  $G$  with respect to arc weights  $EX_{ij}$ , then

$$g = \sum_{i'=2}^{n'} \min_{1 \leq j' < i'} \{EX_{i'j'}\}. \quad (5)$$

We shall henceforth assume that  $G$  is Prim-numbered w.r.t  $\{EX_{ij}\}_{(i,j) \in A}$ .

The following analytic device, motivated by Gilbert's paper [1965], enables us to obtain a better bound for  $E\xi$ . Consider a spanning tree (not necessarily minimal)

constructed for a given  $\omega$  (i.e., for a given realization of the graph), as follows:

For  $i = 2, 3, \dots, n$  choose the arc  $(i, j)$  such that

$$X_{ij}(\omega) = \min_{1 \leq k < i} \{X_{ik}(\omega)\}.$$

It is easily shown by induction on  $i$  that for each  $\omega$ , the  $n - 1$  arcs chosen as above form a spanning tree (Gilbert 1965). We shall refer to this tree as an *exodic tree* (the term is coined by Gilbert, who explains its interpretation on p.378 of his paper).

Define  $W_i = \min\{X_{i1}, \dots, X_{i,i-1}\}$  for  $i = 2, \dots, n$ , and

$$Z = \sum_{i=2}^n W_i = \sum_{i=2}^n \min_{1 \leq j < i} \{X_{ij}\}. \quad (6)$$

Then  $Z$  is a random variable that represents the length of the exodic tree.  $Z$  is uniquely defined for a given numbering of the nodes of  $G$  and is simpler to analyze than  $\xi$ . Indeed it is easy to show that  $EZ$  improves on  $g$  as an approximation of  $E\xi$ .

**Theorem 1.** *If  $G$  is Prim-numbered w.r.t  $\{EX_{ij}\}_{(i,j) \in A}$ , then*

$$E\xi \leq EZ \leq g.$$

*Proof:* The first inequality is trivial because by construction,  $\xi \leq Z$  almost surely.

Now, from (6)

$$\begin{aligned} EZ &= E \sum_{i=2}^n \min_{1 \leq j < i} \{X_{ij}\} \\ &= \sum_{i=2}^n E \min_{1 \leq j < i} \{X_{ij}\} \\ &\leq \sum_{i=2}^n \min_{1 \leq j < i} \{EX_{ij}\} \\ &= g \end{aligned}$$

where the last equation follows from (5). ■

It is noteworthy that our choice of the first node in the PRIM algorithm is quite arbitrary. There are, therefore, at least  $n$  different Prim-numberings of  $G$  which satisfy Theorem 1, and each yields a (possibly) different upper bound  $Z$ . (There would be more than  $n$  Prim-numberings if ties are encountered during execution of PRIM).

Theorem 1 holds even when there is lack of independence among  $X_{ij}$ 's. To obtain the distribution of  $Z$ , however, we invoke this independence. Let the random

variables  $W_i$  have distribution functions  $F_{W_i}$  for  $i = 2, \dots, n$ . Since,  $Z = \sum_{i=2}^n W_i$ , the DF of  $Z$  is given by the  $(n - 1)$ -fold convolution

$$F_Z(x) = (F_{W_2} * \dots * F_{W_n})(x) \quad (7)$$

where

$$F_{W_i}(x) = 1 - \prod_{k=1}^{i-1} (1 - F_{ik}(x)). \quad (8)$$

To illustrate the use of these formulas, consider the exponential case

$$F_{ij}(x) = 1 - e^{-\lambda_{ij}x} \quad x \geq 0$$

and let  $X_{ij}$  be independent. Then,

$$F_{W_i}(x) = 1 - \prod_{k=1}^{i-1} e^{-\lambda_{ik}x} = 1 - \exp\left(-\left(\sum_{k=1}^{i-1} \lambda_{ik}\right)x\right).$$

Let  $\lambda_i = \sum_{k=1}^{i-1} \lambda_{ik}$ . Then the Laplace-Stieltjes transform of  $F_{W_i}$  is given by

$$\tilde{F}_{W_i}(s) = \int_0^{\infty} \lambda_i e^{-sx} e^{-\lambda_i x} dx = \frac{\lambda_i}{s + \lambda_i},$$

and the Laplace-Stieltjes transform of  $F_Z$  is

$$\tilde{F}_Z(s) = \prod_{i=2}^n \frac{\lambda_i}{s + \lambda_i}$$

To invert the transform, we expand  $\tilde{F}_Z(s)$  as a partial fraction:

$$\tilde{F}_Z(s) = \frac{A_2}{s + \lambda_2} + \frac{A_3}{s + \lambda_3} + \dots + \frac{A_n}{s + \lambda_n}. \quad (9)$$

By successively substituting  $-\lambda_2, -\lambda_3, \text{ etc.}$  for  $s$  and solving for  $A_2, A_3, \text{ etc.}$ , we obtain

$$A_i = \frac{\prod_{k=2}^n \lambda_k}{\prod_{\substack{k=2 \\ k \neq i}}^n (\lambda_k - \lambda_i)} \quad i = 2, \dots, n \quad (10)$$

And now by the linearity of the Laplace transform, the individual terms in (9) can be inverted to yield the distribution of  $Z$ :

$$F_Z(x) = \sum_{i=2}^n \int_0^x A_i e^{-\lambda_i t} dt = \sum_{i=2}^n \frac{A_i}{\lambda_i} (1 - e^{-\lambda_i x})$$

which gives

$$EZ = \sum_{i=2}^n \frac{A_i}{\lambda_i^2} \quad (11)$$

When  $X_{ij}$  are i.i.d exponential with parameter  $\lambda$ , (11) yields

$$EZ = \sum_{i=2}^n \frac{1}{(i-1)\lambda}. \quad (12)$$

If the Laplace transform of  $F_Z$  is not directly invertible by inspection as above, then an explicit integration using the Inversion Formula may have to be used.

## 2 Accuracy of the Approximation

Some insight into the relationship between  $\xi$  and  $Z$  can be obtained from a Monte-Carlo simulation of the random minimal spanning tree. For our experiment, we generated graphs with random arc lengths for complete and sparse graphs with i.i.d and non-i.i.d arc lengths. Tables 1 through 4 compare  $EZ$ ,  $E\xi$ , and  $g$  for graphs of various sizes.  $EZ$  is computed for each value of  $n$  by using the formulas of section 3.  $E\xi$  is the average over 1000 Monte-Carlo realizations of the random graph whose structure (*i.e.*, the sets  $N$  and  $A$ , and  $\mu_{ij}$ 's, the mean arc lengths for arcs in  $A$ ) was chosen prior to the Monte-Carlo step, in accordance with the schemes indicated in the four tables. For sparse graphs (Tables 1 and 3), our code first checked for connectedness of graphs (not guaranteed by our method of generating random structures), though for a sparsity factor (*i.e.*, probability than an arc exists) of 0.1, we always got connected graphs.

Table 1:  $X_{ij}$  exponentially distributed with  $\mu_{ij}$  chosen uniformly between 1.0 and 100.0. Sparsity factor = 0.1 (i.e.  $\Pr\{\text{arc}(i,j) \text{ exists}\} = 0.1$ )

| $n$ | $E\xi$  | $EZ$    | $g$      |
|-----|---------|---------|----------|
| 40  | 626.046 | 791.389 | 1120.296 |
| 50  | 622.346 | 865.687 | 1228.896 |
| 60  | 381.298 | 566.956 | 962.151  |
| 70  | 422.141 | 601.960 | 1107.655 |
| 80  | 380.470 | 541.991 | 1111.642 |
| 90  | 354.753 | 532.965 | 1113.422 |
| 100 | 348.155 | 548.266 | 1143.779 |
| 110 | 348.654 | 552.688 | 1231.698 |
| 500 | 274.760 | 469.036 | 1590.662 |

Table 2:  $X_{ij}$  exponentially distributed with  $\mu_{ij}$  chosen uniformly between 1.0 and 100.0. Sparsity factor = 1.0 (i.e. Complete graph).

| $n$ | $E\xi$ | $EZ$   | $g$     |
|-----|--------|--------|---------|
| 10  | 34.276 | 46.627 | 105.640 |
| 20  | 31.642 | 49.353 | 137.619 |
| 30  | 28.040 | 43.484 | 135.129 |
| 40  | 27.191 | 42.128 | 136.703 |
| 50  | 26.367 | 47.483 | 157.473 |
| 60  | 27.955 | 51.071 | 173.729 |
| 70  | 27.157 | 50.821 | 185.730 |
| 80  | 26.064 | 51.209 | 187.194 |
| 90  | 27.156 | 50.809 | 203.407 |
| 100 | 27.432 | 50.940 | 222.503 |
| 110 | 26.963 | 54.678 | 230.275 |
| 500 | 26.140 | 74.147 | 518.133 |

Table 3:  $X_{ij}$  i.i.d exponential with  $\mu_{ij} = 1.0$ . Sparsity factor = 0.1  
(i.e.  $\Pr\{\text{arc}(i,j) \text{ exists}\} = 0.1$ ).

| $n$ | $E\xi$ | $EZ$   | $g$   |
|-----|--------|--------|-------|
| 40  | 16.562 | 27.261 | 39.0  |
| 50  | 16.172 | 34.778 | 49.0  |
| 60  | 13.466 | 32.552 | 59.0  |
| 70  | 13.646 | 34.047 | 69.0  |
| 80  | 13.453 | 34.834 | 79.0  |
| 90  | 12.476 | 36.364 | 89.0  |
| 100 | 12.196 | 37.180 | 99.0  |
| 110 | 11.908 | 38.747 | 109.0 |
| 500 | 12.346 | 55.674 | 499.0 |

Table 4:  $X_{ij}$  i.i.d exponential with  $\mu_{ij} = 1.0$ . Sparsity factor = 1.0  
(i.e. Complete graph).

| $n$ | $E\xi$ | $EZ$  | $g$   |
|-----|--------|-------|-------|
| 10  | 1.258  | 2.829 | 9.0   |
| 20  | 1.243  | 3.548 | 19.0  |
| 30  | 1.240  | 3.962 | 29.0  |
| 40  | 1.242  | 4.254 | 39.0  |
| 50  | 1.222  | 4.479 | 49.0  |
| 60  | 1.225  | 4.663 | 59.0  |
| 70  | 1.241  | 4.819 | 69.0  |
| 80  | 1.216  | 4.953 | 79.0  |
| 90  | 1.222  | 5.071 | 89.0  |
| 100 | 1.216  | 5.177 | 99.0  |
| 110 | 1.228  | 5.273 | 109.0 |
| 500 | 1.248  | 6.791 | 499.0 |

It is apparent in the cases described in Tables 1–4 that the exodic tree performs best in relatively sparse graphs with non-identical arcs. It also appears to be relatively better for small graphs. In each case,  $EZ$  provides a considerable improvement over  $g$  as an estimate of  $E\xi$ . Table 4 also exhibits convergence of  $E\xi$  to the Frieze [1985] limit  $\mu \cdot \zeta(3)$ .

In the remainder of this section we develop an analytical characterization of the

tightness of the exodic tree as an approximation to the random minimal spanning tree. We shall first establish the following lower bound.

**Proposition 2.** Let  $\Lambda_i^n = \min_{\substack{1 \leq j < n \\ j \neq i}} \{X_{ij}\}$  for  $i = 2, \dots, n$  and  $\Lambda = \sum_{i=2}^n \Lambda_i^n$ . Then for every realization of the graph  $G(N, A)$ ,

$$\xi \geq \Lambda.$$

*Proof:* We use induction and the following obvious facts about deterministic MST's:

1. Every MST (indeed, every tree) admits of at least two nodes of degree 1.
2. If a single-degree node and its (only) associated arc are deleted from a MST, the remaining arcs form a MST on the remaining nodes.

The Proposition trivially holds for  $n = 2$ . Assume it is true for every graph  $G(N, A)$  where  $|N| = n$ . Let  $G(N', A')$  be another graph with  $|N'| = n + 1$ . For a given  $\omega$  let  $i$  be one of the single-degree nodes on the MST for this realization of  $G(N', A')$ . Let  $j \in N'$  be the only node adjacent to  $i$  on this MST. Then

$$\xi_{G(N', A')}(\omega) = X_{ij}(\omega) + \xi_{G(N' \setminus \{i\}, A'_i)}$$

where subscripts on  $\xi$  indicate the graph w.r.t which the MST length is being computed, and  $A'_i = A' \setminus \{(i, k) | (i, k) \in A'\}$ . Since  $G(N' \setminus \{i\}, A'_i)$  is an  $n$ -node graph, we have by the induction hypothesis,

$$\begin{aligned} \xi_{G(N', A')}(\omega) &\geq X_{ij}(\omega) + \sum_{\substack{k=2 \\ k \neq i}}^{n+1} \min_{\substack{1 \leq j \leq n+1 \\ j \neq k, j \neq i}} \{X_{kj}(\omega)\} \\ &\geq \min_{\substack{1 \leq j \leq n+1 \\ j \neq i}} X_{ij}(\omega) + \sum_{\substack{k=2 \\ k \neq i}}^{n+1} \min_{\substack{1 \leq j \leq n+1 \\ j \neq k}} \{X_{kj}(\omega)\} \\ &= \Lambda_i^{n+1}(\omega) + \sum_{\substack{k=2 \\ k \neq i}}^{n+1} \Lambda_k^{n+1}(\omega) \\ &= \sum_{i=2}^{n+1} \Lambda_i^{n+1}(\omega) = \Lambda_{G(N', A')}(\omega) \quad \blacksquare \end{aligned}$$

The random variable  $\Lambda$  provides a useful lower bound for  $\xi$ . Also,  $Z - \Lambda$  provides an easily computed upper bound on the difference between  $Z$  and  $\xi$ . For instance,

it is readily seen that for the graphs summarized in Tables 3 and 4, the value of  $E\Lambda$  is 10 and 1 respectively. This eliminates the possibility that  $E\xi$  approaches zero as the number of nodes in the graph grows without limit.

For notational convenience, we denote  $\max\{x, 0\}$  as  $x^+$  for real  $x$ , and define

$$V_i = \begin{cases} \min\{X_{i,i+1}, \dots, X_{in}\} & \text{if } 2 \leq i \leq n-1 \\ \infty & \text{if } i = n \end{cases}$$

then, recalling that  $W_i = \min\{X_{i1}, \dots, X_{i,i-1}\}$ ,

$$\begin{aligned} \Lambda &= \sum_{i=2}^n \Delta_i = \sum_{i=2}^n \min\{W_i, V_i\} \\ &= \sum_{i=2}^n W_i - \sum_{i=2}^n (W_i - V_i)^+. \end{aligned}$$

Hence

$$Z - \Lambda = \sum_{i=2}^n (W_i - V_i)^+ \quad (13)$$

It is then clear from (13) that the necessary and sufficient condition for

$$E\Lambda = E\xi = EZ$$

is

$$V_i \geq W_i \quad \text{w.p. 1} \quad \forall i = 2, \dots, n.$$

Thus, a node numbering  $\pi$  of  $G$  is "optimal" (*i.e.*, the upper and lower bounds on  $E\xi$  with respect to  $\pi$  coalesce into  $E\xi$ ) if the above dominance condition is satisfied for  $\pi$ . This dominance condition, however, is not only difficult to verify, it may well not exist at all among any of the  $n!$  numberings of  $G$  (as is evident for any graph whose arc-lengths are distributed over the the same range). One simple situation in which the dominance condition does hold is when one node is closer to all other nodes than they are to each other. Suppose, for example, that for any  $i, j$ ,  $P\{X_{i1} \leq X_{ij}\} = 1$ . In this case the minimal spanning tree consists of the familiar "hub and spoke" pattern, *i.e.*, the arcs  $(1, j)$ ,  $j = 2, \dots, n$ . Clearly, however, the coalescence of  $\Lambda$ ,  $\xi$ , and  $Z$  involves very special circumstances.

### 3 Asymptotic Optimality

Two questions arise naturally from our discussion of approximations for the random MST. First, how well do the bounds perform for large graphs? Second, what are



the asymptotic properties of the random MST? Our next result shows that under appropriate conditions the relative error induced by using the exodic tree in place of the MST tends to zero as the number of nodes in  $G$  grows large.

In order to characterize the growth of graphs, we use the "incremental" model of Weide [1978]. Consider a sequence of graphs  $G_1, G_2, \dots$ , where  $G_n = (N_n, A_n)$ ,  $N_n, A_n$  are respectively the node and arc sets of  $G_n$ , and  $|N_n| = n$ . Let  $X_{ij}^n$  denote the arc length random variable for  $i, j \in N_n$  and  $(i, j) \in A_n$ . The incremental growth model assumes  $G_1 \subseteq G_2 \subseteq \dots$  in the sense that  $X_{ij}^n \equiv X_{ij}^{n+1} \equiv \dots$  for each  $(i, j) \in A_n$ . Thus, the graph  $G_{n+1}$  is identical to the graph  $G_n$  save for the addition of node  $n + 1$  and arcs incident to it.

To denote the functional dependence on the number of nodes, we shall henceforth annotate our symbols with  $n$ . Thus

$$\xi_n = \min_{T_i \in T_n} \left\{ \sum_{(j,k) \in T_i} X_{jk}^n \right\}$$

and

$$Z_n = \sum_{i=2}^n \min_{1 \leq j < i} \{X_{ij}^n\} = \sum_{i=2}^n W_i^n.$$

**Theorem 3.** *Let  $\{G_n\}$  be a sequence of complete graphs growing according to the incremental model. Let  $X_{ij}^n$  be independent non-negative random variables with distribution functions  $F_{ij}$  having a positive lower support for all  $(i, j) \in A_n$ , i.e.,  $\exists a > 0$  such that  $F_{ij}(a-) = 0 \quad \forall (i, j) \in A_n$ , and  $\exists$  a distribution function  $F_*$  such that*

1.  $F_*(x) \leq F_{ij}(x) \quad \forall x$  and  $\forall (i, j) \in A_n$ , and
2.  $F_*(a + \epsilon) > 0 \quad \forall \epsilon > 0$ .

Then as  $n \rightarrow \infty$ ,

$$(a) \quad \frac{Z_n}{\xi_n} \rightarrow 1 \quad \text{a.s.}$$

If, in addition,  $E(\sup_{n \geq 2} \min_{1 \leq j < n} \{X_{nj}^n\}) < \infty$ , then

$$(b) \quad E \left[ \frac{Z_n}{\xi_n} \right] \rightarrow 1, \quad \text{and}$$

$$(c) \quad \frac{EZ_n}{E\xi_n} \rightarrow 1.$$

Proof: Since  $F_{ij}(a-) = 0 \quad \forall (i, j) \in A_n$ , we have

$$X_{ij}^n \geq a \quad \text{w.p. 1} \quad \forall (i, j) \in A_n.$$

Hence  $\xi_n \geq (n-1)a$  w.p. 1, and

$$1 \leq \frac{Z_n}{\xi_n} \leq \frac{\sum_{i=2}^n W_i^n}{(n-1)a} \quad \text{w.p. 1} \quad (14)$$

Now,  $\forall \varepsilon > 0$ ,

$$P\{W_i^i > a + \varepsilon\} = \prod_{j=1}^{i-1} (1 - F_{ij}(a + \varepsilon)) \leq (1 - F_*(a + \varepsilon))^{i-1}.$$

Hence,

$$\sum_{i=2}^{\infty} P\{W_i^i > a + \varepsilon\} \leq \sum_{i=2}^{\infty} (1 - F_*(a + \varepsilon))^{i-1} < \infty \quad (15)$$

since  $F_*(a + \varepsilon) > 0$ ; and by the *Borel-Cantelli Lemma*,

$$P\{W_i^i > a + \varepsilon \text{ i.o.}\} = 0 \quad \text{i.e.,}$$

$$W_i^i \rightarrow a \quad \text{a.s.}$$

Therefore the Cesàro sum also converges, i.e., as  $n \rightarrow \infty$

$$\frac{\sum_{i=2}^n W_i^i}{n-1} \rightarrow a \quad \text{a.s.}$$

Now, the incremental growth of  $G_n$  ensures that for any  $n \geq i$ ,  $W_i^i \equiv W_i^n$ , so we have from (14)

$$\frac{Z_n}{\xi_n} \rightarrow 1 \quad \text{a.s.}$$

To prove (b), observe that

$$\begin{aligned} E \left[ \sup_{n \geq 2} \frac{Z_n}{\xi_n} \right] &\leq E \left[ \sup_{n \geq 2} \frac{Z_n}{(n-1)a} \right] \\ &\leq E \left[ \frac{(n-1) \sup_{i \geq 2} W_i^i}{(n-1)a} \right] \\ &= E \left[ \frac{1}{a} \sup_{i \geq 2} \min_{1 \leq j < i} \{X_{ij}\} \right] \\ &< \infty \end{aligned}$$

Hence by *Lebesgue Dominated Convergence Theorem*, part (a) implies

$$E \left[ \frac{Z_n}{\xi_n} \right] \rightarrow 1$$

Similarly, since  $E(\sup_{n \geq 2} W_n^n) = E(\sup_{n \geq 2} \min_{1 \leq j < n} \{X_{nj}\}) < \infty$ , we again have by *Lebesgue DCT* that

$$W_n^n \rightarrow a \text{ a.s.} \Rightarrow EW_n^n \rightarrow a.$$

Then (c) follows by the *Cesàro* argument since

$$1 \leq \frac{EZ_n}{E\xi_n} \leq \frac{EW_2^2 + \dots + EW_n^n}{(n-1)a}. \quad \blacksquare$$

The convergence results of *Theorem 3* can be established for sparse graphs as well, if we ensure that  $W_n^n \rightarrow a$  a.s. The following result extends *Theorem 3* to incomplete graphs. Define the *in-degree* of a node  $i$  in  $G_n$  to be the number of lower-numbered nodes connected to  $i$ . i.e., let  $d^n(i) = \sum_{j < i} 1_{\{(i,j) \in A_n\}}$ ,  $2 \leq i \leq n$ . Again, the incremental growth of  $G_n$  ensures that  $d^n(i) = d^i(i) \quad \forall n \geq i$ .

**Theorem 3'.** *Theorem 3 holds for incomplete graphs if there exist positive constants  $\delta, K_1, K_2$ , and  $N$  such that  $n > N$  implies*

$$d^n(n) > K_1 + K_2[\log n + (1 + \delta) \log \log n].$$

*Proof:* Let  $\beta_\epsilon = (1 - F_\epsilon(a + \epsilon))$ . Then the inequality in (15) reduces to

$$\sum_{n=2}^{\infty} P\{W_n^n > a + \epsilon\} \leq \sum_{n=2}^{\infty} \beta_\epsilon^{d^n(n)}.$$

Since  $\sum_n (1/n(\log n)^{1+\delta}) < \infty$  for all  $\delta > 0$ , we have by the comparison test that  $\sum_n \beta_\epsilon^{d^n(n)}$  converges if there exist positive constants  $C_1, N$  such that  $n > N$  implies

$$\beta_\epsilon^{d^n(n)} < C_1 \frac{1}{n(\log n)^{1+\delta}}.$$

Writing  $\log \beta_\epsilon = -C_2$ , with  $C_2 > 0$ , the condition for convergence is given as

$$d^n(n) > \frac{\log C_1}{-C_2} + \frac{1}{C_2}[\log n + (1 + \delta) \log \log n].$$

Writing  $K_1 = (\log C_1 / -C_2)$  and  $K_2 = (1/C_2)$  yields the desired condition. ■

It is of interest to note that under the growth conditions of Theorem 3,

$$\liminf_{n \rightarrow \infty} \frac{g_n}{E\xi_n} \geq \frac{\min_{ij} EX_{ij}}{a}.$$

Thus, for instance, when  $X_{ij}$  are i.i.d with a shifted exponential, i.e.,

$$P\{X_{ij} \leq x\} = \begin{cases} 0 & x < a \\ 1 - e^{-\lambda(x-a)} & x \geq a \end{cases}$$

then  $g_n/E\xi_n$  remains bounded below by  $1 + \frac{1}{a\lambda}$ , whereas

$$1 \leq \frac{EZ_n}{E\xi_n} \leq \frac{EZ_n}{E\Lambda_n} \leq \frac{(n-1)a + \frac{1}{\lambda} \log(n-1) + 1}{(n-1)a} = 1 + \frac{\log(n-1) + 1}{\lambda(n-1)a}$$

which indicates rapid convergence of  $EZ_n/E\xi_n$  to 1. For example if  $a = 1$ ,  $\lambda = \frac{1}{10}$ , and  $n = 500$ , then  $EZ_n$  overestimates  $E\xi_n$  by at most 15%.

## 4 Application to a Network Provisioning Model

Consider the task of constructing a communications network with the MST topology to interlink a given configuration of nodes. The cost of the connecting cable (e.g., a coaxial fiber optic cable) and its installation charge per unit length are deemed substantially high; and at the planning stage the length of cable needed to link a pair of nodes is not known with certainty (due to uncertainties about the exact path to be taken, wastages, etc.). The planner needs to place an order for the total length of cable required for the network based on probabilistic information about the inter-node distances, and the "true" distances (upon which the MST configuration is based) become known only later at the implementation stage. If the ordered length of cable falls short of the true requirement, then a supplemental order must be placed at a higher unit cost, possibly also accompanied by a fixed ordering cost. Contrariwise, the surplus length of cable can be disposed of at some salvage value. The decision problem is to determine the optimal quantity of cable to be ordered at the first stage so as to minimize the total expected cost of cable needed for the network. This problem is an example of the classical single-period "newsboy" problem. The difficulty lies in characterizing the "demand" distribution.

Let  $x$  be this order quantity and let  $c_1, c_2$  be the unit costs of cable in stage 1 and stage 2 respectively. Let  $s$  be the salvage value per unit length of cable, and  $K$  the fixed ordering cost should a supplemental order be needed in stage 2. Let  $0 \leq s < c_1 < c_2$ , and  $K \geq 0$ .

As before,  $G = (N, A)$  represents the graph with  $|N| = n$ , and the arc-lengths  $X_{ij}$ 's are independent random variables with DF's  $F_{ij}$ 's. Then the total cost of cable needed for the network is given by

$$C(\xi_n, x) = c_1x + c_2(\xi_n - x)^+ + K1_{\{\xi_n > x\}} - s(x - \xi_n)^+ \quad (16)$$

whence

$$EC(\xi_n, x) = c_1x + c_2 \int_x^\infty (t - x) dF_{\xi_n}(t) + K(1 - F_{\xi_n}(x)) - s \int_0^x (x - t) dF_{\xi_n}(t) \quad (17)$$

where  $F_{\xi_n}$  is the DF of the MST length. If  $F_{\xi_n}$  has density  $f_{\xi_n}$ , then the first-order condition requires that the optimal order satisfy the equation

$$(c_2 - s)F_{\xi_n}(x) - Kf_{\xi_n} = c_2 - c_1 \quad (18)$$

In case there is no fixed cost for the supplemental order, then, the optimal order quantity  $x_{\xi_n}^*$  is the  $\left(\frac{c_2 - c_1}{c_2 - s}\right)$ th fractile of  $F_{\xi_n}$ .

Since  $F_{\xi_n}$  is analytically intractable, we solve a surrogate problem of minimizing  $EC(Z_n, x)$ , the expected total cost for the exodic tree, and obtain the optimal order quantity  $x_{Z_n}^*$  by solving

$$(c_2 - s)F_{Z_n}(x) - Kf_{Z_n}(x) = c_2 - c_1$$

Again, if  $K = 0$ ,  $x_{Z_n}^*$  is obtained as the  $\left(\frac{c_2 - c_1}{c_2 - s}\right)$ th fractile of  $F_{Z_n}$  which can be computed as in Section 1. Notice that since  $\xi_n \leq Z_n$  w.p. 1, we have  $F_{\xi_n}(x) \geq F_{Z_n}(x) \quad \forall x$ , and hence  $x_{\xi_n}^* \leq x_{Z_n}^*$ .

The remainder of this section is devoted to characterizing the "goodness" of our approximation for the provisioning problem. Assume henceforth that  $G(N, A)$  satisfies the conditions of Theorem 3.

Our first result states that for any size  $x$  of the order, the surrogate problem provides an asymptotically correct approximation to the total cost.

**Proposition 4.**

$$\lim_{n \rightarrow \infty} \frac{C(Z_n, x)}{C(\xi_n, x)} = 1 \quad \text{a.s.} \quad \text{uniformly in } x$$

*Proof:* Since  $f^+ - g^+ \leq (f - g)^+$  for all functions  $f$  and  $g$ , we have from (16)

$$\left| \frac{C(Z_n, x)}{C(\xi_n, x)} - 1 \right| \leq \left| \frac{c_2(Z_n - \xi_n)^+ + s(Z_n - \xi_n)^+ + K(1_{\{Z_n > x\}} - 1_{\{\xi_n > x\}})}{c_1x + c_2(\xi_n - x)^+ + K1_{\{\xi_n > x\}} - s(x - \xi_n)^+} \right|$$

Now, since  $s < c_1 < c_2$ ,

$$c_1\xi_n \leq c_1x + c_2(\xi_n - x)^+ + K1_{\{\xi_n > x\}} - s(x - \xi_n)^+ \quad \text{a.s.} \quad \forall x$$

Therefore

$$\begin{aligned} \left| \frac{C(Z_n, x)}{C(\xi_n, x)} - 1 \right| &\leq \left| \frac{c_2}{c_1} \frac{(Z_n - \xi_n)^+}{\xi_n} + \frac{s}{c_1} \frac{(Z_n - \xi_n)^+}{\xi_n} + \frac{K}{c_1\xi_n} \right| \quad \text{a.s.} \\ &= \frac{c_2 + s}{c_1} \left( \frac{Z_n}{\xi_n} - 1 \right) + \frac{K}{c_1\xi_n} \quad \text{a.s. because } Z_n \geq \xi_n \text{ a.s.} \end{aligned}$$

Since  $\frac{Z_n}{\xi_n} \rightarrow 1$  a.s. by Theorem 3(a), and  $\xi_n \rightarrow \infty$  a.s., the Proposition follows by taking limits as  $n \rightarrow \infty$  and noting that the right-hand side is independent of  $x$ . ■

We require some technical lemmas in order to obtain our main result on the asymptotic precision of the exodic tree as a surrogate for the MST.

**Lemma 5.**

$$\lim_{n \rightarrow \infty} \frac{EC(Z_n, x)}{EC(\xi_n, x)} = 1 \quad \text{uniformly in } x$$

*Proof:* As in Proposition 4,  $s < c_1 < c_2 \Rightarrow c_1\xi_n \leq C(\xi_n, x)$  a.s.  $\forall x$ . Hence,

$$c_1E\xi_n \leq EC(\xi_n, x) \quad \forall x.$$

Now from (17),

$$\left| \frac{EC(Z_n, x)}{EC(\xi_n, x)} - 1 \right| \leq \left| \frac{c_2E(Z_n - \xi_n)^+ + sE(Z_n - \xi_n)^+ + K(P\{Z_n > x\} - P\{\xi_n > x\})}{c_1E\xi_n} \right|$$

Again, since  $Z_n \geq \xi_n$  a.s.,

$$\left| \frac{EC(Z_n, x)}{EC(\xi_n, x)} - 1 \right| \leq \frac{c_2 + s}{c_1} \left[ \frac{EZ_n}{E\xi_n} - 1 \right] + \frac{K}{c_1} \frac{1}{E\xi_n}$$

Application of Theorem 3(c) as  $n \rightarrow \infty$  then completes the proof. ■

**Lemma 6.** Let  $\langle f_n \rangle, \langle g_n \rangle$  be sequences of real-valued functions having global minima at points  $\langle x_n^* \rangle, \langle y_n^* \rangle$  respectively. If

$$\lim_{n \rightarrow \infty} \frac{f_n(u)}{g_n(u)} = 1 \quad \text{uniformly in } u,$$

then

$$\lim_{n \rightarrow \infty} \frac{f_n(x_n^*)}{g_n(y_n^*)} = 1.$$

*Proof:* By the optimality of  $x_n^*$  and  $y_n^*$

$$\frac{f_n(x_n^*)}{g_n(x_n^*)} \leq \frac{f_n(x_n^*)}{g_n(y_n^*)} \leq \frac{f_n(y_n^*)}{g_n(y_n^*)} \quad (19)$$

and by the assumed uniform convergence, the left and right sides of (19) both converge to 1. ■

The main result can now be established.

**Proposition 7.** Let there exist  $x_{\xi_n}^*$  and  $x_{Z_n}^*$  such that

$$EC(\xi_n, x_{\xi_n}^*) = \min_{x \geq 0} EC(\xi_n, x)$$

and

$$EC(Z_n, x_{Z_n}^*) = \min_{x \geq 0} EC(Z_n, x)$$

Then

$$(a) \lim_{n \rightarrow \infty} \frac{EC(Z_n, x_{Z_n}^*)}{EC(\xi_n, x_{\xi_n}^*)} = 1, \text{ and}$$

$$(b) \lim_{n \rightarrow \infty} \frac{EC(\xi_n, x_{Z_n}^*)}{EC(\xi_n, x_{\xi_n}^*)} = 1.$$

*Proof:* For definiteness, when  $x_{Z_n}^*$  and  $x_{\xi_n}^*$  are not unique, we take the smallest among such values. Then (a) is a direct application of Lemmas 5 and 6.

To prove (b), first observe that

$$C(Z_n, x) - C(\xi_n, x) = \begin{cases} c_2(Z_n - \xi_n) & \text{if } Z_n > \xi_n > x \\ c_2(Z_n - x) + K + s(x - \xi_n) & \text{if } Z_n > x > \xi_n \\ s(Z_n - \xi_n) & \text{if } x > Z_n > \xi_n \end{cases}$$

Therefore, since  $\xi_n \leq Z_n$  w.p. 1, we have by Theorem 1

$$C(\xi_n, x) \leq C(Z_n, x) \quad \forall x \quad \text{w.p. 1}$$

and hence

$$EC(\xi_n, x) \leq EC(Z_n, x) \quad \forall x \quad (20)$$

Notice that

$$1 \leq \frac{EC(\xi_n, x_{Z_n}^*)}{EC(\xi_n, x_{\xi_n}^*)} \leq \frac{EC(Z_n, x_{Z_n}^*)}{EC(\xi_n, x_{\xi_n}^*)}$$

where the first inequality follows from the optimality of  $x_{\xi_n}^*$  for  $EC(\xi_n, x)$ , and the second from (20). Now by taking limits as  $n \rightarrow \infty$  and using part (a), we get the desired convergence. ■

Proposition 7 may be interpreted in the following manner. For each instance of the problem, the MST network is constructed according to the “true” arc-lengths that become known at the implementation stage. Therefore, the expected total cost function that the planner really faces is  $EC(\xi_n, x)$ , which by definition is minimized at  $x_{\xi_n}^*$ . Then  $EC(\xi_n, x_{Z_n}^*)$  is the expected total cost the planner will incur by using  $x_{Z_n}^*$ , the optimal order quantity computed for the surrogate problem. Proposition 7(b) asserts that the proportional error in optimal expected cost, induced by using the surrogate problem, tends to zero. This approach to approximating the provisioning problem is similar in spirit to that of Dempster *et al.* [1983] in their analysis of hierarchical scheduling problems.

## 5 Concluding Remarks and Open Questions

It seems likely that the conclusions of Theorem 3 will hold under much weaker conditions. One generalization would be to establish these results under the more general “independent” growth model for random graphs (Weide [1978]). It may also be possible to relax the requirement of a *uniform* positive lower support  $a$  for all  $F_{ij}$ ’s, and prove the theorem for positive, but not necessarily equal, lower supports  $a_{ij}$ ’s for  $F_{ij}$ ’s respectively. However, we would not be able to do away entirely with positive supports since we know from Frieze’s [1985] result that in the absence of these, the MST length converges to a limit for complete graphs with



i.i.d arc lengths. In particular, for complete graphs with i.i.d. exponential( $\lambda$ ) arc lengths,  $E\xi_n \rightarrow \frac{1}{\lambda}\zeta(3)$ , and from (12),  $EZ_n \sim \frac{1}{\lambda} \log n$ .

One of the analytical difficulties in strengthening Theorem 3 lies in not having tight lower bounds that we can exploit. We use  $\Lambda$  in section 2 to examine how node numbering affects the “goodness” of  $EZ$ , and work with  $(n-1)a$  in Theorem 3, but none of these is very tight. A tighter lower bound to  $\xi$  is  $L = \sum_{k=1}^{n-1} X_{(k)}$ , the sum of the first  $n-1$  order statistics of  $X_{ij}$ ’s, and it remains an open question to see if this or any other lower bound might enable a generalization of our results.

**Acknowledgement:** We are grateful to an anonymous referee for simplifying an earlier proof of Proposition 2.

## References

- AHO, A.V., J.E. HOPCROFT AND J.D. ULLMAN. 1983. *Data Structures and Algorithms*. Addison-Wesley. Reading, Mass.
- DEMPSTER M.A.H., M.L. FISHER, L. JANSEN, B.J. LAGEWEG, J.K. LENSTRA AND A.H.G RINNOOY KAN. 1983. Analysis of Heuristics for Stochastic Programming: Results for Hierarchical Scheduling Problems. *Math. Opns. Res.* 8, 525-537.
- FRIEZE, A.M. 1985. On the Value of a Random Minimum Spanning Tree Problem. *Discrete Appl. Math.* 10, 47-56.
- GILBERT, E.N. 1965. Random Minimal Trees. *J. SIAM.* 13, 376-387.
- HARARY, E. 1969. *Graph Theory*. Addison-Wesley. Reading, Mass.
- KRUSKAL, J.B., JR. 1956. On the Shortest Spanning Subtree of a Graph and the Travelling Salesman Problem. *Proc. Amer. Math. Soc.* 7, 48-50.
- LAWLER, E. 1976. *Combinatorial Optimization: Networks and Matroids*. Holt, Rinehart and Winston. New York.
- LOBERMAN H. AND A. WEINBERGER. 1957. Formal Procedures for Connecting Terminals with a Minimum Total Wire Length. *J. ACM.* 4, 428-437.
- LUEKER, G.S. 1981. Optimisation Problems on Graphs with Independent Edge Weights. *SIAM J. Comput.* 10, 338-351.
- PRIM, R.C. 1957. Shortest Connection Networks and Some Generalizations. *Bell Sys. Tech. Jour.* 36, 1389-1401.
- RIORDAN, J. 1978. *An Introduction to Combinatorial Analysis*. Princeton University Press. Princeton, New Jersey.
- STEELE, J.M. 1981. Growth Rates of Minimal Spanning Trees of Multivariate Sample. Stanford Univ., Dept. of Statistics. Research Report.
- WEIDE, B. 1978. Statistical Methods in the Design of Algorithms. Ph.D Thesis, Carnegie-Mellon University.