A note on random minimum length spanning trees

Alan Frieze* Miklós Ruszinkó†
Lubos Thoma‡
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA15213, USA
alan@random.math.cmu.edu,
ruszinko@lutra.sztaki.hu,
thoma@qwes.math.cmu.edu

Submitted June 28, 2000, Accepted August 11, 2000

Abstract

Consider a connected r-regular n-vertex graph G with random independent edge lengths, each uniformly distributed on $[0, 1]$. Let $mst(G)$ be the expected length of a minimum spanning tree. We show in this paper that if G is sufficiently highly edge connected then the expected length of a minimum spanning tree is $\sim \frac{n}{r}\zeta(3)$. If we omit the edge connectivity condition, then it is at most $\sim \frac{n}{r}(\zeta(3) + 1)$.

1 Introduction

Given a connected simple graph $G = (V, E)$ with edge lengths $\mathbf{x} = (x_e : e \in E)$, let $mst(G, \mathbf{x})$ denote the minimum length of a spanning tree. When $\mathbf{X} = (X_e : e \in E)$ is a family of independent random variables, each uniformly distributed on the interval $[0, 1]$, denote the expected value $\mathbf{E}(mst(G, \mathbf{X}))$ by $mst(G)$. Consider the complete graph K_n. It is known (see [2]) that, as $n \to \infty$, $mst(K_n) \to \zeta(3)$. Here $\zeta(3) = \sum_{j=1}^{\infty} j^{-3} \sim 1.202$. Beveridge, Frieze and McDiarmid [1] proved two theorems that together generalise the previous results of [2], [3], [5].

*Supported in part by NSF Grant CCR9818411 email: alan@random.math.cmu.edu
†Permanent Address Computer and Automation Research Institute of the Hungarian Academy of Sciences, Budapest, P.O.Box 63, Hungary-1518. Supported in part by OTKA Grants T 030059 and T 29074 FKFP 0607/1999. email: ruszinko@lutra.sztaki.hu
‡Supported in part by NSF grant DMS-9970622. email: thoma@qwes.math.cmu.edu

Theorem 1 *For any n-vertex connected graph G,*

$$mst(G) \geq \frac{n}{\Delta}(\zeta(3) - \epsilon_1)$$

where $\Delta = \Delta(G)$ denotes the maximum degree in G and $\epsilon_1 = \epsilon_1(\Delta) \to 0$ as $\Delta \to \infty$.

For an upper bound we need expansion properties of G.

Theorem 2 *Let $\alpha = \alpha(r) = O(r^{-1/3})$ and let $\rho = \rho(r)$ and $\omega = \omega(r)$ tend to infinity with r. Suppose that the graph $G = (V, E)$ is connected and satisfies*

$$r \leq \delta \leq \Delta \leq (1 + \alpha)r, \tag{1}$$

where $\delta = \delta(G)$ denotes the minimum degree in G. Suppose also that

$$|(S : \bar{S})|/|S| \geq \omega r^{2/3} \log r \;\; \textit{for all } S \subseteq V \textit{ with } \; r/2 < |S| \leq \min\{\rho r, |V|/2\}, \tag{2}$$

where $(S : \bar{S}) = \{(x, y) \in E : \; x \in S, y \in \bar{S} = E \setminus S\}$. Then

$$\left| mst(G) - \frac{n}{r}\zeta(3) \right| \leq \epsilon_2 \frac{n}{r}$$

where the $\epsilon_2 = \epsilon_2(r) \to 0$ as $r \to \infty$.

For regular graphs we of course take $\alpha = 0$.

The expansion condition in the above theorem is probably not the "right one" for obtaining $mst(G) \sim \frac{n}{r}\zeta(3)$. We conjecture that high edge connectivity is sufficient: Let $\lambda = \lambda(G)$ denote the *edge connectivity* of G.

Conjecture 1
Suppose that (1) holds. Then,

$$\left| mst(G) - \frac{n}{r}\zeta(3) \right| \leq \epsilon_3 \frac{n}{r}$$

where $\epsilon_3 = \epsilon_3(\lambda) \to 0$ as $\lambda \to \infty$.

Note that $\lambda \to \infty$ implies $r \to \infty$.

Along these lines, we prove the following theorem.

Theorem 3 *Assume $\alpha = \alpha(r) = O(r^{-1/3})$ and (1) is satisfied. Suppose that $r \geq \lambda(G) \geq \omega r^{2/3} \log n$ where $\omega = \omega(r)$ tends to infinity with r. Then*

$$\left| mst(G) - \frac{n}{r}\zeta(3) \right| \leq \epsilon_4 \frac{n}{r}$$

where the $\epsilon_4 = \epsilon_4(r) \to 0$ as $r \to \infty$.

Remark: It is worth pointing out that it is not enough to have $r \to \infty$ in order to have the result of Theorem 2, that is, we need some extra condition such as high edge connectivity. For consider the graph $\Gamma(n, r)$ obtained from n/r r-cliques $C_1, C_2, \ldots, C_{n/r}$ by deleting an edge (x_i, y_i) from C_i, $1 \leq i \leq n/r$ then joining the cliques into a cycle of cliques by adding edges (y_i, x_{i+1}) for $1 \leq i \leq n/r$. It is not hard to see that

$$mst(\Gamma(n,r)) \sim \frac{n}{r}\left(\zeta(3) + \frac{1}{2}\right)$$

if $r \to \infty$ with $r = o(n)$. We repeat the conjecture from [1] that this is the worst-case, i.e.

Conjecture 2 *Assuming only the conditions of Theorem 1,*

$$mst(G) \leq \frac{n}{\delta}\left(\zeta(3) + \frac{1}{2} + \epsilon_5\right)$$

where $\epsilon_5 = \epsilon_5(\delta) \to 0$ *as* $\delta \to \infty$.

We prove instead

Theorem 4 *If G is a connected graph then*

$$mst(G) \leq \frac{n}{\delta}(\zeta(3) + 1 + \epsilon_6)$$

where the $\epsilon_6 = \epsilon_6(\delta) \to 0$ *as* $\delta \to \infty$.

We finally note that high connectivity is not necessary to obtain the result of Theorem 2. Since if $r = o(n)$ then one can tolerate a few small cuts. For example, let G be a graph which satisfies the conditions of Theorem 2 and suppose $r = o(n)$. Then taking 2 disjoint copies of G and adding a single edge joining them we obtain a graph G' for which $mst(G') \sim \frac{1}{2} + \frac{n'}{r}\zeta(3) \sim \frac{n'}{r}\zeta(3)$ where $n' = 2n$ is the number of vertices of G'.

2 Proof of Theorem 3

Given a connected graph $G = (V, E)$ with $|V| = n$ and $0 \leq p \leq 1$, let G_p be the random subgraph of G with the same vertex set which contains those edges e with $X_e \leq p$. Let $\kappa(G)$ denote the number of components of G. We shall first give a rather precise description of $mst(G)$.

Lemma 1 *[1]*
For any connected graph G,

$$mst(G) = \int_{p=0}^{1} \mathbf{E}(\kappa(G_p))dp - 1. \tag{3}$$

□

We substitute $p = x/r$ in (3) to obtain

$$mst(G) = \frac{1}{r}\int_{x=0}^{r} \mathbf{E}(\kappa(G_{x/r}))dx - 1.$$

Now let $C_{k,x}$ denote the total number of components in $G_{x/r}$ with k vertices. Thus

$$mst(G) = \frac{1}{r}\int_{x=0}^{r} \sum_{k=1}^{n} \mathbf{E}(C_{k,x})dx - 1. \tag{4}$$

Proof of Theorem 3

In order to use (4) we need to consider three separate ranges for x and k, two of which are satisfactorily dealt with in [1]. Let $A = (r/\omega)^{1/3}$, $B = \lfloor (Ar)^{1/4} \rfloor$ so that each of $B\alpha$, AB^2/r and $A/B \to 0$ as $r \to \infty$. These latter conditions are needed for the analysis of the first two ranges.

Range 1: $0 \le x \le A$ and $1 \le k \le B$ – see [1].

$$\frac{1}{r}\int_{x=0}^{A} \sum_{k=1}^{B} \mathbf{E}(C_{k,x})dx \le (1 + o(1))\frac{n}{r}\zeta(3).$$

Range 2: $0 \le x \le A$ and $k > B$ – see [1].

$$\frac{1}{r}\int_{x=0}^{A} \sum_{k=B}^{n} \mathbf{E}(C_{k,x})dx = o(n/r).$$

Range 3: $x \ge A$.

We use a result of Karger [4]. A cut $(S : \bar{S}) = \{(u, v) \in E : \ u \in S, v \notin S\}$ of G is γ-minimal if $|(S : \bar{S})| \le \gamma\lambda$. Karger proved that the number of γ-minimal cuts is $O(n^{2\gamma})$. We can associate each component of G_p with a cut of G. Thus

$$\sum_{k=1}^{n} \mathbf{E}(C_{k,x}) \le O\left(\sum_{s=\lambda}^{\infty} n^{2s/\lambda}\left(1 - \frac{x}{r}\right)^s\right) = O\left(\sum_{s=\lambda}^{\infty} (n^{2r/\lambda}e^{-x})^{s/r}\right)$$
$$= O\left(\int_{s=\lambda}^{\infty} (n^{2r/\lambda}e^{-x})^{s/r} ds\right) = O\left(\frac{rn^2 e^{-x\lambda/r}}{x - \frac{2r}{\lambda}\log n}\right),$$

and using $A\lambda \ge \omega^{2/3} r \log n$ we obtain

$$\frac{1}{r}\int_{x=A}^{r} \sum_{k=1}^{n} \mathbf{E}(C_{k,x})dx = O\left(\int_{x=A}^{r} \frac{n^2 e^{-x\lambda/r}}{x - \frac{2r}{\lambda}\log n} dx\right)$$
$$= O\left(A^{-1}\int_{x=A}^{r} n^2 e^{-x\lambda/r} dx\right) = O\left(\frac{rn^2}{A\lambda} e^{-A\lambda/r}\right) = o(n/r).$$

We complete the proof by applying Lemma 1. □

3 Proof of Theorem 4

We keep the definitions of A, B and Ranges 1,2, but we split Range 3 and let $\delta = r$.
Range 3a: $x \geq A$ and $k \leq (1-\epsilon)r$, $0 < \epsilon < 1$, arbitrary – see [1] (here $\epsilon = 1/2$ but the argument works for arbitrary ϵ).

$$\frac{1}{r}\int_{x=A}^{r}\sum_{k=1}^{(1-\epsilon)r}\mathbf{E}(C_{k,x})dx = o(n/r).$$

Range 3b: $x \geq A$ and $k > (1-\epsilon)r$.
Clearly

$$\sum_{k=(1-\epsilon)r}^{n} C_{k,x} \leq \frac{n}{(1-\epsilon)r}$$

and hence

$$\frac{1}{r}\int_{x=A}^{r}\sum_{k=(1-\epsilon)r}^{n}\mathbf{E}(C_{k,x})dx \leq \frac{n}{(1-\epsilon)r}.$$

We again complete the proof by applying Lemma 1. □

References

[1] A. Beveridge, A. M. Frieze and C. J. H. McDiarmid, *Minimum length spanning trees in regular graphs*, Combinatorica 18 (1998) 311-333.

[2] A. M. Frieze, *On the value of a random minimum spanning tree problem*, *Discrete Applied Mathematics* **10** (1985) 47 - 56.

[3] A. M. Frieze and C. J. H. McDiarmid, *On random minimum length spanning trees*, Combinatorica 9 (1989) 363 - 374.

[4] D. R. Karger, *A Randomized Fully Polynomial Time Approximation Scheme for the All Terminal Network Reliability Problem*, Proceedings of the twenty-seventh annual ACM Symposium on Theory of Computing (1995) 11-17.

[5] M. Penrose, *Random minimum spanning tree and percolation on the n-cube*, Random Structures and Algorithms **12** (1998) 63 - 82.