Report Number: WUCS-88-04

1988-01-01

Probable Performance of Steiner Tree Algorithms

Bernard M. Waxman

In this paper we consider the probable performance of three polynomial time approximation algorithms for the Steiner tree problem with respect to a specific random graph model. The Steiner problem asks us to find a minimum cost spanning subgraph (tree) for a subset D of modes in a graph.

PROBABLE PERFORMANCE OF STEINER TREE ALGORITHMS

Bernard M. Waxman

WUCS-88-04

Department of Computer Science
Washington University
Campus Box 1045
One Brookings Drive
Saint Louis, MO 63130-4899

Abstract

In this paper we consider the probable performance of three polynomial time approximation algorithms for the Steiner tree problem with respect to a specific random graph model. The Steiner problem asks us to find a minimum cost spanning subgraph (tree) for a subset D of nodes in a graph. In our model, graphs have an edge probability that is given by $p(n)=\frac{(Cnlogn)^{1/d}}{n}$. Graphs with edge probability $\frac{(Cnlogn)^{1/d}}{n}$, where $d \in Z^+ - \{1\}$ and $C>2$, have diameter equal to d almost always (a.a.) according to the results of work done by Bela Bollobas. We show for $k(n)<n^{(1-\varepsilon)/d(d+1)}$ for any $\varepsilon > 0$ that all three algorithms yield solutions with identical cost a.a., where $k(n)=|D|$. In addition if $\varepsilon n^{1/d(d+1)} \leq k(n) \leq (n \log n)^{(1-\varepsilon)/2d}$ for any $\varepsilon > 0$ we show that one algorithm a.a. produces a solution with cost less than that of the other two.

This work supported by Bell Communications Research, Itatel SIT, NEC and National Science Foundation grant DCI-8600947.

PROBABLE PERFORMANCE OF STEINER TREE ALGORITHMS

Bernard M. Waxman

1. Introduction

In this paper we consider a version of the Steiner tree problem referred to as the Steiner tree problem in graphs. In this problem we are given an undirected graph $G(V, E)$, a cost function $C: E \rightarrow Z^+$, and $D \subset V$. We are asked to find a spanning subgraph for D with minimum cost, called the minimum Steiner tree, where the cost of a graph is defined to be the sum of its edge costs. This problem was shown to be NP-complete by Karp in 1972 [5] and remains NP-complete even if C is a constant function. Thus, there is interest in polynomial time approximation algorithms for the Steiner problem.

There are a number of approximation algorithms for deriving solutions to the Steiner tree problem in graphs which run in polynomial time. None of these is known to give a solution that is better than twice the optimal solution in the worst case. In this paper we consider three algorithms: the *minimum distance tree* heuristic (MDT), the *minimum spanning tree* heuristic (MST), and Rayward-Smith's algorithm (RS).

In the first algorithm, MDT, we designate one node in D to be the root r. We then construct a shortest path tree from r to the other nodes in D. MDT can be implemented using any of the shortest path algorithms, and thus, has the virtue of being relatively simple. Unfortunately it has worst case performance which can be nearly as bad as k times the cost of an optimal solution, where $k = |D|$.

The second algorithm, MST, begins by constructing a complete graph on the set D, where the distances between nodes correspond to the distances in the original graph G. A minimum spanning tree for this derived graph is then determined using one of the standard minimum spanning tree algorithms. Finally each edge of this tree is translated into a path in G to produce a solution. MST will produce solutions whose cost is no worse than twice the cost of an optimal solution. In addition two is the best worst case bound since we can construct examples for which MST yields a solution which has cost $(2 - \epsilon)$ times optimal for any $\epsilon > 0$. For more details on MST see [1,4].

RS is an algorithm developed by Rayward-Smith [6,7]. RS works by defining a function $f: V \rightarrow \Re^+$. Initially we create a collection of single node trees $\mathcal{T}$ consisting of the nodes in D. At each stage of RS we choose a node v, for which $f(v)$ is minimum, and a pair of trees in $\mathcal{T}$, which are two trees closest to v. The pair of trees is joined by a shortest path through v. In other words RS is analogous to Kruskal's algorithm, where the role of a shortest edge is replaced by a path through vertex v for which $f(v)$ is minimum. Informally f is the average cost of making r joins to $r+1$ trees through a node v, where $1 \leq r < |\mathcal{T}|$. More formally we define

$$f_i(v) \;=\; \min_{S \subseteq \mathcal{T}, |S|>1} \left\{ \frac{1}{|S|-1} \sum_{T \in S} \operatorname{dist}(v, T) \right\}.$$

for each stage $i \in [0..|D|-1]$ of RS.

Our graph model will consist of the sample space of all labeled graphs of order n with edge probability $0 \leq p(n) \leq 1$. In addition we choose a cost function that maps each edge to 1. We indicate that a graph G is drawn from this sample space by saying $G \in \Omega_n(p)$, i.e. G is a standard random graph with edge probability p and edge weight of 1. We use the expression almost always (a.a.) to indicate that a graph drawn from $\Omega_n(p)$ has a given property with probability approaching 1 as $n \to \infty$. For many of our results we choose a specific class of functions for $p(n)$. This work was inspired by some interesting results on the diameter of random graphs [2] due to Bela Bollobas. We use the following proposition.

Proposition 1.1 *If we let* $p(n) = \frac{(Cn \log n)^{1/d}}{n}$ *for fixed* $d \in Z^+ - \{1\}$ *and* $C > 2$, *we have that graphs in* $\Omega_n(p(n))$ *have diameter* d, *a.a.*

This follows directly from corollary 7(i) for $d = 2$ and from corollary 8(i) for $d > 2$ in the paper by Bollobas.

For convenience we shall refer to the model $\Omega_n(\frac{(Cn \log n)^{1/d}}{n})$ for fixed $d \in Z^+ - \{1\}$ and fixed $C > 0$ as $\Lambda_n(C, d)$.

2. Preliminaries

Throughout this paper we make use of the following notation. We let $P(\text{expr})$ stand for the probability that *expr* is true, $f(n) \to v$ stand for the $\lim_{n\to\infty} f(n)$ and $f(n) \sim g(n)$ indicate that $\frac{f(n)}{g(n)} \to 1$. For a graph $G(V, E)$ we will always use n to stand for $|V|$ and we define $V_i(u)$ to be all nodes at a distance of i from u for $0 \leq i \leq \operatorname{dia}(G)$ and $\emptyset$ otherwise. We then define $n_i(u) = |V_i(u)|$, $S_i(u) = \bigcup_{k=0}^{i} V_k(u)$ and $s_i(u) = |S_i(u)|$. Often we drop the the u where no confusion results. Finally we indicate by $g \in B(n, p)$ that g is drawn from a binomial distribution of order n with probability p.

We state a number of facts which we use in several proofs that follow. Note that p stands for a probability so $0 \leq p \leq 1$.

1. For the geometric series $R_n = \sum_{i=0}^{n} r^i$ if $r > 1$ it is clear that $r^n < R_n < \frac{r}{r-1} r^n$.

2. Given $i < \frac{\log n}{\log pn}$ it follows that $(\exists \epsilon > 0)([(1+\epsilon)(pn)]^i < n)$.
3. Hence, given $i < \frac{\log n}{\log pn}$ we have $(\exists \epsilon > 0)([(1+\epsilon)(pn)]^{i-1}p < 1)$.
4. $1-(1-p)^x \leq px$ if $x \geq 1$.

Finally we state a second proposition (Bela Bollobas [2]) which gives concrete limits for the tails of a binomial distribution, derived from the De Moivre-Laplace *limit theorem.*

Proposition 2.1 *Let* $x \in B(n,p)$, $0 < p < \frac{1}{3}$, $0 < \epsilon < \frac{1}{21}$, *and* $\epsilon np > 40$. *Then*

$$P\Big(|x - pn| \geq \epsilon np\Big) \ < \ \frac{\exp(-\frac{\epsilon^2 np}{3})}{\epsilon(np)^{\frac{1}{2}}} \ < \ \exp(-\frac{\epsilon^2 np}{3})\,.$$

Although not stated explicitly we assume, throughout this paper, that $\epsilon < \frac{1}{21}$ whenever we make use of proposition 2.1. We also note that for $\Lambda_n(C,d)$ the other conditions required for this proposition hold, for large enough n.

3. Distance in $\Lambda_n(C,d)$

Lemma 3.1 *Given* $G(V,E) \in \Lambda(C,d))$, $u \in V$, *and* $i \in Z_0^+$, *there exists an* $N > 0$ *such that for all* $n > N$ *and all* i

$$P\Big(n_i(u) \ \leq \ [(1+\epsilon)pn]^i\Big) \ \geq \ \left(1 - \exp(-\frac{\epsilon^2 pn}{3})\right)^i \ \to \ 1 \tag{1}$$

for any $\epsilon > 0$.

Proof: Let $t(n) = \frac{\log n}{\log pn}$. Then for $i > t(n)$ we have $n_i \leq n$, and the lemma clearly holds. For $i \leq t(n)$ we prove $P\Big(n_i(u) \leq [(1+\epsilon)pn]^i\Big) \geq (1 - \exp(-\frac{\epsilon^2 pn}{3}))^i \to 1$ for large enough N by induction on i. Clearly for $i = 0$ this statement holds by definition of $V_0(u)$. Assume the result holds for $i-1 \geq 0$ where $i \leq t(n)$. Then $n_i \in B(n - s_{i-1}, 1-(1-p)^{n_{i-1}})$ so that

$$n_i \ \leq \ (1+\epsilon)(n - s_{i-1})(1-(1-p)^{n_{i-1}})$$

with probability $\geq 1 - \exp(-\frac{\epsilon^2 pn}{3})$ by proposition 2.1. Then

$$\begin{aligned} n_i &\leq (1+\epsilon)n(1-(1-p)^{n_{i-1}}) \\ &\leq (1+\epsilon)n(n_{i-1}p) \\ &\leq (1+\epsilon)n[(1+\epsilon)pn]^{i-1}p \\ &= [(1+\epsilon)pn]^i \end{aligned}$$

with probability $\geq (1 - \exp(-\frac{\epsilon^2 pn}{3}))^i$.

□

Lemma 3.2 *Given $G(V,E) \in \Lambda(C,d)$ and $u \in V$, there exists an $N > 0$ such that*

$$P\Big(\frac{s_i(u)}{n} \;\leq\; \frac{(1+\epsilon)^{i+1}}{(pn)^{1-\epsilon}}\Big) \;\geq\; (1-\exp(-\frac{\epsilon^2 np}{3}))^i$$

for any $0 < \epsilon < \frac{1}{21}$ if $n > N$. Thus $\frac{s_i(u)}{n} \to 0$ for $0 \leq i < d$ with probability approaching 1.

Proof: By lemma 3.1 and fact 1 on geometric series we have for $\epsilon > 0$, with probability $\geq (1-\exp(-\frac{\epsilon^2 pn}{3}))^i$, that

$$\begin{aligned}\frac{s_i(u)}{n} &\leq \frac{\frac{pn}{pn-1}[(1+\epsilon)pn]^i}{n}\\ &\leq \frac{\frac{pn}{pn-1}(1+\epsilon)^i}{(pn)^{d-\epsilon'-i}}\\ &\leq \frac{\frac{pn}{pn-1}(1+\epsilon)^i}{(pn)^{1-\epsilon'}} \qquad (2)\end{aligned}$$

where $\epsilon' \geq d - \frac{\log n}{\log pn}$. Note that

$$\frac{\log n}{\log pn} = \frac{\log n}{\frac{1}{d}\log(Cn\log n)} = \frac{d\log n}{[\log C + \log n + \log\log n]}\,.$$

Thus, $\lim\limits_{n\to\infty} d - \frac{\log n}{\log pn} = 0$ and hence, we can choose $\epsilon' < \epsilon$ for large enough N_1 where $n > N_1$. If we choose N_2 large enough then for all $n > N_2$ we have that $\frac{pn}{pn-1} < 1+\epsilon$. Then if we let $N = \max\{N_1, N_2\}$ the lemma follows.
□

Theorem 3.1 *Given $G(V,E) \in \Lambda_n(C,d)$, we have for any $u, v \in V$ the following:*

1. $\mathrm{dist}(u,v) \geq d$
2. *if $C > 2$ then* $\mathrm{dist}(u,v) = d$

a.a.

Proof: These result follows directly from lemma 3.2 and proposition 1.1.
□

4. Performance of MDT, MST, RS When $|D|$ is Fixed

Using the results from the previous section we can say something about the cost of Steiner trees in graphs $G(V,E) \in \Lambda_n(C,d)$ where $C > 2$. Proposition 1.1, due to Bollobas, clearly indicates that for any $D \subset V$ there will a.a. be a Steiner tree of cost $d(k-1)$ where $k = |D|$. For fixed $|D|$ it follows from theorem 3.1 that MST and MDT will yield trees of cost $d(k-1)$ a.a. In the next section we make this result stronger by by showing that MST and MDT still yield solutions with cost $d(k-1)$ even if $k = k(n)$ grows as a function of n as long as $k(n) \leq (n \log n)^{(1-\epsilon)/2d}$ for some $\epsilon > 0$. Of course we assume that $k(n) \in Z^+ - \{1\}$ for all n.

We prove that the solutions produced by RS also have cost $d(k-1)$ a.a. for fixed $k \in Z^+$. Thus, as long as the size of the set D is fixed all three algorithms have the same probable performance. We note that a solution produced by RS will be better than that given by MST iff there exists a node x such that $f(x) < d$, where f is the function defined in RS. We proceed by proving that no such node x is expected to exist if k is fixed. In the next section we consider what happens when k grows as a function of n.

In the next lemma we evaluate the limit of an expression which we use in theorem 4.1 to determine the probability that a node x is connected to every node in a set $S \subset V$ with a cost $\leq (\ell+1)|S|$, for a graph $G(V,E) \in \Lambda_n(C,d)$, where $\ell + 1 < \text{dia}(G)$. We note that a node x is connected to a node u by a path of length $\leq \ell+1$ with probability approaching $1-(1-p)^{s_\ell(u)}$ since $s_\ell(u)/n \to 0$. Then clearly the probability that x is connected to each node in S with a path of length $\leq \ell+1$ is given by $[1-(1-p)^{s_\ell}]^b$ where $b = |S|$. Finally, noting that there are at most n possible choices for the node x, we see that the expression in the lemma below is related to the desired probability.

Lemma 4.1 *Let $b, d \geq 2$ let $0 \leq \ell < d-1$ and let $\delta, C > 0$. If $(\ell+1)(b) \geq d(b-1)$ then*

$$\lim_{n\to\infty}\left\{1-\left[1-\left(1-\frac{(Cn\log n)^{1/d}}{n}\right)^{\delta^\ell(Cn\log n)^{\ell/d}}\right]^b\right\}^n = 0 \tag{3}$$

else if $(\ell+1)(b) < d(b-1)$

$$\lim_{n\to\infty}\left\{1-\left[1-\left(1-\frac{(Cn\log n)^{1/d}}{n}\right)^{\delta^\ell(Cn\log n)^{\ell/d}}\right]^b\right\}^n = 1\,. \tag{4}$$

Proof: Using the binomial expansion we note that

$$\left(1-\frac{a}{n}\right)^{(\delta a)^\ell} = 1-\delta^\ell\frac{(a)^{\ell+1}}{n}\dots$$

If we set $a = (Cn\log n)^{1/d}$ we have $\lim_{n\to\infty} \delta^\ell \frac{a^{\ell+1}}{n} = 0$ for $\ell < d-1$ therefore,

$$1-\left(1-\frac{a}{n}\right)^{(\delta a)^\ell} \sim \delta^\ell\frac{a^{\ell+1}}{n}\,.$$

Substituting a for $(Cn\log n)^{1/d}$ and $g = g(n)$ for $\delta^{b\ell}a^{b(\ell+1)}n^{-b}$ it follows that

$$\begin{aligned}\lim_{n\to\infty}\left\{1-\left[1-\left(1-\frac{(Cn\log n)^{1/d}}{n}\right)^{\delta^\ell(Cn\log n)^{\ell/d}}\right]^b\right\}^n &= \lim_{n\to\infty}\left\{1-\left[1-\left(1-\frac{a}{n}\right)^{(\delta a)^{\ell/d}}\right]^b\right\}^n \\ &= \lim_{n\to\infty}\left(1-\delta^{b\ell}\frac{a^{b(\ell+1)}}{n^b}\right)^n \\ &= \lim_{n\to\infty}\left[(1-g)^{1/g}\right]^{gn}\end{aligned}$$

The second equality follows from the fact that $\lim_{n\to\infty}(1-f)^n = \lim_{n\to\infty}(1-h)^n$ if $f \sim h$ and $\lim_{n\to\infty} h = 0$. Note that g is a function of n that is strictly decreasing and with $\lim_{n\to\infty} g = 0$. Therefore, we have that

$$\lim_{n\to\infty}\left[(1-g)^{1/g}\right]^{gn} = \lim_{n\to\infty}\exp(-gn)$$

Then expanding g we have that

$$\begin{aligned} gn &= \delta^{b\ell}(Cn\log n)^{\frac{b(\ell+1)}{d}} n^{1-b} \\ &= \delta^{b\ell}(C\log n)^{\frac{b(\ell+1)}{d}} n^{1-b(1-\frac{\ell+1}{d})} \end{aligned}$$

Since $\lim_{n\to\infty}\delta^{b\ell}(C\log n)^{\frac{b(\ell+1)}{d}} \to \infty$ we have that

$$\lim_{n\to\infty}\exp(-gn) = 0$$

if

$$\lim_{n\to\infty} n^{1-b(1-\frac{\ell+1}{d})} = 1 \quad \text{or} \quad \lim_{n\to\infty} n^{1-b(1-\frac{\ell+1}{d})} \longrightarrow \infty$$

and since n dominates $\log n$

$$\lim_{n\to\infty}\exp(-gn) = 1$$

if

$$\lim_{n\to\infty} n^{1-b(1-\frac{\ell+1}{d})} = 0.$$

Thus if

$$1 - b\left(1 - \frac{\ell+1}{d}\right) \geq 0.$$

we have that (3) holds and if

$$1 - b(1 - \frac{\ell+1}{d}) < 0.$$

we have that (4) holds and the lemma follows.
□

Theorem 4.1 *Suppose we are given a graph $G(V,E) \in \Lambda_n(C,d)$ with $C > 2$ and a set of point S, with $|S| = b$. Then*

$$P\Big(\big(\exists x \in G\big)\Big(\sum_{v\in S}\operatorname{dist}(x,v) < d(b-1)\Big)\Big) \rightarrow 0.$$

Proof: We note that $s_i(u) < ((2Cn\log n)^{1/d})^i$ a.a. using the results of lemma 3.1. We also note that for $G(V,E) \in \Lambda_n(C,d)$ that a vertex x is connected to u by a path of length $\leq j+1$ with probability less than

$$1 - \left(1 - \frac{(Cn\log n)^{1/d}}{n}\right)^{2^j(Cn\log n)^{j/d}} \tag{5}$$

for large enough n. Then x is connected to each node $u_i \in S$ by a path of length $\leq j_i + 1$ with probability less than

$$\begin{aligned}\prod_{i=1}^{b} 1 - \left(1 - \frac{(Cn\log n)^{1/d}}{n}\right)^{2^{j_i}(Cn\log n)^{j_i/d}} &\sim \prod_{i=1}^{b} \frac{2^{j_i}(Cn\log n)^{\frac{j_i+1}{d}}}{n} \\ &= \frac{2^{b\ell}(Cn\log n)^{\frac{b(\ell+1)}{d}}}{n^b} \\ &\sim \left[1 - \left(1 - \frac{(Cn\log n)^{1/d}}{n}\right)^{2^{\ell}(Cn\log n)^{\ell/d}}\right]^b \qquad (6)\end{aligned}$$

where $\ell = 1/b\sum_i j_i$. We use the argument from lemma 4.1 that

$$1 - \left(1 - \frac{a}{n}\right)^{2a^{j_i}} \sim \frac{2^{j_i}a^{j_i+1}}{n}$$

to get the result derived in (6) above.

Therefore, lemma 4.1 can be used to give us the probability that

$$\left(\neg\exists x \in V(G)\right)\left(\sum_{i=1}^{b} \operatorname{dist}(x, v_i) < d(b-1)\right). \qquad (7)$$

If $(\ell + 1)(b) < d(b-1)$ then equation (4) of lemma 4.1 holds and we have that the probability of (7) goes to 1.
□

Corollary 4.1 *Let $D \subset V$, where $G(V,E) \in \Lambda_n(C,d)$, $C > 2$, and let $k = |D|$ be fixed. $(\forall u \in V)(\forall i \in [0..k-1])(f_i(u) \geq d)$, where f_i is the function defined in RS.*

Proof: If k and d are fixed then the number of choices for the set S in theorem 4.1 is fixed for each step of RS. Thus, this corollary is a direct consequence of theorem 4.1.
□

Theorem 4.2 *Let $D \subset V$, where $G(V,E) \in \Lambda_n(C,d)$, $C > 2$. If $k = |D|$ is fixed MDT, MST, and RS will produce solutions with cost $d(k-1)$ a.a.*

Proof: This result follows directly from corollary 4.1 and theorem 3.1.
□

We conjecture that when k is fixed the solutions found by each of the algorithms is optimal but do not attempt to prove it here.

5. When MST is Better than MDT

At this point we have shown that for fixed diameter graphs and for fixed size of the set D the cost of a solution produced by either MST, MDT, or RS is almost always $d(|D|-1)$. We now consider the case in which the size of D grows as a function of n, that is we let $|D| = k(n)$, where, as in the previous section, $k(n) \in Z^+ - \{1\}$ for all n.

Lemma 5.1 *Given $G(V,E) \in \Lambda_n(C,d)$, with $C > 2$, if $k(n) \leq (n \log n)^{\frac{1-\epsilon}{2d}}$, for any $\epsilon > 0$, a solution produced by MST for an instance (G, D) of the Steiner tree problem has cost $d(k-1)$ a.a.*

Proof: Given two nodes u and v, from lemma 3.2 we have

$$P(\text{dist}(u,v) \geq d) \geq \left(1 - \exp(-\frac{\epsilon^2 np}{3})\right)^{d-1}\left(1 - \frac{(1+\epsilon)^d}{(pn)^{1-\epsilon}}\right) \tag{8}$$

for any $\epsilon > 0$ i.e. as long we choose $n > N$ for N large enough. Given a set of nodes D such that $|D| = k(n)$ it follows that

$$\begin{aligned} P((\forall u, v \in D)(\text{dist}(u,v) \geq d)) &\geq \left[\left(1 - \exp(-\frac{\epsilon^2 np}{3})\right)^{d-1}\left(1 - \frac{\alpha}{(pn)^{1-\epsilon}}\right)\right]^{\binom{k(n)}{2}} \\ &\geq \left[\left(1 - \exp(-\frac{\epsilon^2 np}{3})\right)^{d-1}\left(1 - \frac{\alpha}{(pn)^{1-\epsilon}}\right)\right]^{k^2(n)} \\ &= \left[1 - \exp(-\frac{\epsilon^2 np}{3})\right]^{(d-1)k^2(n)}\left[1 - \frac{\alpha}{(pn)^{1-\epsilon}}\right]^{k^2(n)} \\ &\geq \left[1 - (d-1)k^2(n)\exp(-\frac{\epsilon^2 np}{3})\right]\left[1 - \frac{\alpha k^2(n)}{(pn)^{1-\epsilon}}\right] \end{aligned} \tag{9}$$

where $\alpha = (1+\epsilon)^d$. Hence, for any function $k(n)$ such that $k(n) \leq (n \log n)^{\frac{1-\epsilon'}{2d}} = [(1/C)^{1/d} pn]^{\frac{1-\epsilon'}{2}}$, for some $\epsilon' > 0$, the expression in (9) above will go to 1 as $n \to \infty$ as long as we choose ϵ, $0 < \epsilon < \epsilon'$. Thus, using proposition 1.1 we have that the distance between all pairs $u, v \in D$ will be d a.a. and the lemma follows.
□

We can prove a similar result for the MDT algorithm which we state in the next lemma.

Lemma 5.2 *Given $G(V,E) \in \Lambda_n(C,d)$, with $C > 2$, if $k(n) \leq (n \log n)^{\frac{1-\epsilon}{d}}$, for any $\epsilon > 0$, a solution produced by MDT for an instance (G, D) of the Steiner tree problem has cost $d(k-1)$ a.a.*

Proof: The proof is nearly identical to the proof of lemma 5.1 except that we only consider pairs of nodes r, v where r is the root chosen by MDT. Thus we need consider only $k(n)$ pairs instead of $\binom{k(n)}{2}$.
□

Lemma 5.3 *Let $G(V,E) \in \Lambda_n(C,d)$, $u \in V$, and $i \in Z_0^+$. If $i < d$*

$$P(n_i(u) \geq ((1-\epsilon)pn)^i) \;\; \rightarrow \;\; 1 \tag{10}$$

for any $\epsilon > 0$.

Proof: We prove $P(n_i(u) \geq ((1-\epsilon)pn)^i) \rightarrow 1$ by induction on $i < d$. Clearly the result holds for $i = 0$. Assume it holds for $i-1$, where $i < d$, then $n_i \in B(n - s_{i-1}, 1-(1-p)^{n_{i-1}})$ and it follows that

$$n_i \;\; \geq \;\; \left(1 - \frac{\epsilon}{2}\right)(n - s_{i-1})(1-(1-p)^{n_{i-1}}) \tag{11}$$

with probability approaching 1. By lemma 3.2, $n - s_{i-1} > (1-\frac{\epsilon}{2})n$ for large enough n so

$$\begin{aligned} n_i \;\; &\geq \;\; \left(1-\frac{\epsilon}{2}\right)\left(1-\frac{\epsilon}{2}\right) n(1-(1-p)^{n_{i-1}}) \\ &\geq \;\; \left(1-\frac{\epsilon}{2}\right)^2 npn_{i-1} \\ &\geq \;\; (1-\epsilon)np((1-\epsilon)pn)^{i-1} \\ &\geq \;\; ((1-\epsilon)pn)^i. \end{aligned}$$

□

Lemma 5.4 *Let $G(V,E) \in \Lambda_n(C,d)$, with $C > 2$. If $k(n) \geq (n \log n)^{\frac{1+\epsilon}{2d}}$, for some $\epsilon > 0$, a solution produced by MST for an instance (G,D) of the Steiner tree problem has cost $< d(k-1)$ a.a.*

Proof: Using the result of the previous lemma and proposition 2.1 we have that there exist an $N > 0$ such for all $n > N$

$$\begin{aligned} \frac{s_{d-1}(u)}{n} \;\; &\geq \;\; \frac{[(1-\epsilon)pn]^{d-1}}{n} \\ &\geq \;\; \frac{(1-\epsilon)^{d-1}}{(pn)} \end{aligned}$$

with probability $\geq (1-\exp(-\frac{\epsilon^2 pn}{3}))^{d-1}$. Therefore, if we let $\alpha = (1-\exp(-\frac{\epsilon^2 pn}{3}))^{d-1}$ then the probability of finding a pair of nodes $u, v \in D$ such that $\text{dist}(u,v) < d$ for $n > N$ is given by

$$P((\exists u, v \in D)(\text{dist}(u,v) < d)) \;\; \geq \;\; 1 - \left[1 - \alpha\frac{(1-\epsilon)^{d-1}}{pn}\right]^{\binom{k(n)}{2}}.$$

So if $k(n) \geq (n \log n)^{(1+\epsilon)/2d}$ then for some ϵ', $0 < \epsilon' < \epsilon$ it follows that $\binom{k(n)}{2} \geq ((1-\epsilon)^{1-d} Cn \log n)^{(1+\epsilon')/d}$ assuming that $n > N$ for large enough N. Therefore, since $pn = (Cn \log n)^{1/d}$, it follows that

$$\begin{aligned} P((\exists u, v \in D)(\text{dist}(u,v) < d)) \;\; &\geq \;\; 1 - \exp\left(-\alpha(Cn \log n)^{\frac{\epsilon'}{d}}\right) \\ &\rightarrow \;\; 1 \end{aligned}$$

since $\alpha \to 1$ and $(Cn\log n)^{\epsilon'/d} \to \infty$.
□

Theorem 5.1 *Let $G(V,E) \in \Lambda_n(C,d)$, with $C > 2$. The cost of a solution produced by MST for an instance (G,D) of the Steiner tree problem is a.a. less than the cost of a solution produced by MDT if*

$$(n\log n)^{\frac{1+\epsilon}{2d}} \leq k(n) \leq (n\log n)^{\frac{1-\epsilon}{d}},$$

for any $\epsilon > 0$.

Proof: This theorem follows directly from lemmas 5.2 and 5.4.
□

6. When RS is Better than MST and MDT

From lemma 5.3 we can conclude that $n_i > (\frac{pn}{a})^i$ for any $a > 1$ so, for example, we can choose $a = 2$. Then using an argument similar to that used in the proof of lemma 4.1 we have the following.

Lemma 6.1 *Let $G(V,E) \in \Lambda_n(C,d)$ where $C > 2$, let $S \subset D$, let $b = |S|$, and let $0 \leq \ell < d$, where $b((l+1) \in Z^+$. Then the probability that $(\exists x \in V)(\exists S \subset D)$ $(\sum_{v_i \in S} \mathrm{dist}(x,v_i) < b(\ell+1))$ goes to 1 if*

$$\lim_{n\to\infty} \binom{k(n)}{b} n^{1-b(1-\frac{\ell+1}{d})} > 0 \tag{12}$$

and goes to 0 if

$$\lim_{n\to\infty} \binom{k(n)}{b} n^{1-b(1-\frac{\ell+1}{d})} = 0 . \tag{13}$$

Proof: This lemma follows from an argument similar to that given in the proof of lemma 4.1. When $0 \leq i < d$ we note that $n - s_i > \frac{n}{2}$ from lemma 3.2 and that $n_i > (\frac{(pn)}{2})^i$ from lemma 5.3. We note that the whether δ in lemma 4.1 equals $\frac{1}{2}$ or 2 has no effect on the result of the lemma. Thus expression (5) actually is asymptotic to the probability that a node u is connected to a node v by a path of length $\leq j+1$. Therefore, from theorem 4.1 we have the existence of a node x with probability approaching

$$\lim_{n\to\infty} 1 - \left\{\left[1 - \left(1 - \left(1 - \frac{(Cn\log n)^{1/d}}{n}\right)^{(Cn\log n)^{\ell/d}}\right)^b\right]^{\frac{n}{\alpha}}\right\}^{\binom{k(n)}{b}} \tag{14}$$

for some α, $1 \leq \alpha < 2$. Using an argument identical to that of lemma 4.1 we have that (14)

$$\begin{aligned}
&= \lim_{n\to\infty} 1 - \left[\exp - \left(\frac{1}{\alpha}(C\log n)^{\frac{b(\ell+1)}{d}} n^{1-b(1-\frac{\ell+1}{d})}\right)\right]^{\binom{k(n)}{b}} \\
&= \lim_{n\to\infty} 1 - \left(\exp - \binom{k(n)}{b} \frac{1}{\alpha}(C\log n)^{\frac{b(\ell+1)}{d}} n^{1-b(1-\frac{\ell+1}{d})}\right) .
\end{aligned}$$

Since $\frac{1}{\alpha}(C \log n)^{\frac{b(\ell+1)}{d}} \to \infty$ and since n dominates $\log n$ the result follows.
□

We note, for fixed d, that if $k(n) \to \infty$ then $\binom{k(n)}{d+1} \sim (k(n))^{d+1}$ and we have the following lemma.

Lemma 6.2 *Let $G(V,E) \in \Lambda_n(C,d)$, $C > 2$, $D \subset V$, and $b = d+1$. If $|D| = k(n) \geq \epsilon n^{1/d(d+1)}$ for any $\epsilon > 0$ then $\exists x \in V$ and $\exists S \subset D$ with $|S| = b$ such that x is connected to every $u \in S$ with a path of length less that d, e.g. $\ell + 1 \leq d - 1$, with probability approaching 1.*

Proof: If $\ell + 1 = d - 1$ and $k(n) \geq \epsilon n^{1/d(d+1)}$ for an $\epsilon > 0$ then

$$\begin{aligned}
\lim_{n\to\infty} \binom{k(n)}{b} n^{1-b(1-\frac{\ell+1}{d})} &= \lim_{n\to\infty} \binom{k(n)}{d+1} n^{1-(d+1)(1-\frac{d-1}{d})} \\
&\sim \lim_{n\to\infty} (k(n))^{d+1} n^{1-(d+1)(1-\frac{d-1}{d})} \\
&\geq \lim_{n\to\infty} \epsilon^{d+1} n^{\frac{1}{d}} n^{1-(d+1)(1-\frac{d-1}{d})} \\
&= \lim_{n\to\infty} \epsilon^{d+1} n^{\frac{-1}{d}+\frac{1}{d}} \\
&= \epsilon^{d+1}.
\end{aligned}$$

Since $\epsilon > 0$ we have that $\epsilon^{d+1} > 0$. Thus lemma 6.1 yields the correct result. □

Theorem 6.1 *Let $G(V,E) \in \Lambda_n(C,d)$, with $C > 2$. If $\epsilon n^{1/d(d+1)} \leq k(n) \leq (n \log n)^{(1-\epsilon)/2d}$, for some $\epsilon > O$, then the cost of a solution produced by RS, for an instance (G,D) of the Steiner tree problem, is almost always less than the cost of a solution produced by MST.*

Proof: By lemma 5.1 the cost of a solution produced by MST is a.a. $d(|D|-1)$ under the conditions specified. From lemma 6.2 for some $S \subset D$ with $|S| = d+1$ there exists an $x \in V$ such that $\text{dist}(x,s) \leq d-1$ for all $s \in S$. Therefore, the cost to connect the nodes in S will be less that or equal to $(d-1)(d+1) = d(|S|-1) - 1$. RS will find the node x or one that yields a solution at least as good. Hence the cost of a solution produced by RS will be less than or equal to $d(|D|-1)-1$. We therefore, conclude that RS yields a solution which is better than the solution given by MST or MDT provided that $k(n)$ grows at the rate specified.
□

Lemma 6.3 *Let $G(V,E) \in \Lambda_n(C,d)$, with $C > 2$, and let $D \subset V$. If $k(n) \leq n^{(1-\epsilon)/d(d+1)}$, for any $\epsilon > 0$, then $\forall x \in V$ and $\forall S \subset D$ with $b = |S|$ it follows that $\sum_{u \in S} \text{dist}(x,u) \geq d(b-1)$ with probability approaching 1.*

Proof: Let us assume that we have a set $S \subset D$ and a node x such that $\sum_{u \in S} \text{dist}(x,u) < d(b-1)$ where $b = |S|$. Then clearly $(\exists S' \subset S)(\sum_{u \in S'} \text{dist}(x,u) < d(b-1))$ and

$(\forall u \in S')(\text{dist}(x,u) < d)$. Thus without loss of generality we need only consider pairs (x, S) where $(\forall u \in S)(\text{dist}(x,u) \leq d-1)$. We now consider two cases.

case i $\{b > d\}$ In this case we need consider only those sets S where $|S| = d+1$. If $|S| > d+1$ and there is a node x such that $\sum_{u\in S} \text{dist}(x,u) < d(b-1)$ then $\forall S' \subset S$ such that $|S'| = d+1$ it follows that $\sum_{u\in S'} \text{dist}(x,u) < d(b-1)$ since $(\forall u \in S)(\text{dist}(x,u) \leq d-1)$. Thus, we can substitute S' for S. Using a derivation similar to that in lemma 6.2 and noting that $\ell+1 \leq d-1$ we have, for $k(n) < n^{(1-\epsilon)/d(d+1)}$, that

$$\begin{aligned}
\lim_{n\to\infty} \binom{k(n)}{b} n^{1-b(1-\frac{\ell+1}{d})} &\leq \lim_{n\to\infty} n^{\frac{1-\epsilon}{d}} n^{\frac{-1}{d}} \\
&= \lim_{n\to\infty} n^{\frac{-\epsilon}{d}} \\
&= 0
\end{aligned}$$

for any $\epsilon > 0$. Thus this theorem hold for case (i).

case ii $\{b \leq d\}$ In this $b = d-j$, for an integer j, where $0 \leq j \leq d-2$. Then, for any pair (x, S) such that $\sum_{u\in S} \text{dist}(x,u) < d(b-1)$, it follows that

$$(\ell + l)(d-j) \;\leq\; d(d-j-1)-1\,,$$

where $b(\ell+1) = \sum_{u\in S} \text{dist}(x,u)$, since the sum must be an integer. Thus for a fixed j using the derivation in lemma 6.2 we have for $k(n) < n^{(1-\epsilon)/d(d+1)}$ that

$$\begin{aligned}
\lim_{n\to\infty} \binom{k(n)}{b} n^{1-b(1-\frac{\ell+1}{d})} &= \lim_{n\to\infty} \binom{k(n)}{b} n^{1-(d-j)(1-\frac{d(d-j-1)-1}{d(d-j)})} \qquad (15)\\
&= \lim_{n\to\infty} \binom{k(n)}{b} n^{\frac{-1}{d}} \\
&\leq \lim_{n\to\infty} n^{\frac{1-\epsilon}{d}} n^{\frac{-1}{d}} \\
&= \lim_{n\to\infty} n^{\frac{-\epsilon}{d}} \\
&= 0
\end{aligned}$$

for any $\epsilon > 0$. Thus by lemma 6.1 we have the required result.
□

This last lemma allows us to strengthen the result that tells us that the three algorithms do equally well as long as the size of D is fixed. The next theorem states this stronger result.

Theorem 6.2 *Let $G(V,E) \in \Lambda_n(C,d)$, with $C > 2$. If $k(n) < n^{\frac{1-\epsilon}{d(d+1)}}$ then MDT, MST, and RS yield solutions of cost $d(k(n)-1)$ a.a. for an instance (G,D) of the Steiner tree problem.*

Proof: This theorem follows immediately from lemmas 5.1, 5.2 and 6.3.
□

7. Conclusion

We have compared the performance of three Steiner tree approximation algorithms on a random graph model for which the diameter is almost always constant. We have shown that all three perform equally well almost always when the size of the set D, for which we are to find a Steiner tree, grows more slowly than $n^{(1-\epsilon)/d(d+1)}$ for any $0 < \epsilon < 1$. We also have shown that MST almost always does better than MDT when $k(n) = |D|$ satisfies $(n \log n)^{(1+\epsilon)/d^2} \leq k(n) \leq (n \log n)^{(1-\epsilon)/d}$. Finally we have shown that RS does better than both MST and MDT when $\epsilon n^{1/d(d+1)} \leq k(n) \leq (n \log n)^{(1-\epsilon)/d^2}$.

If $k(n) > (n \log n)^{(1-\epsilon)/d^2}$ we conjecture that RS still does better than the other two algorithms with some nonzero probability, even though the results presented here tell us nothing about the relative performance in this case. In addition our results do not give us any quantitative comparison of the performance of the three algorithms nor do they tell us anything about how the performance of these algorithms compares to that of an optimal algorithm.

Each of the points in the paragraph above warrant further investigation. The consideration of other random graph models to further compare the three algorithms discussed in this paper, as well as others, should also prove to be useful. Most helpful would be a random graph model for which the cost of an optimal solution, or at least an upper bound, is known. Unfortunately at this time it is not clear to us how to construct such a model so that it will yield useful information.

References

[1] K. Bharath-Kumar and J.M. Jaffe. "Routing to multiple destinations in computer networks." *IEEE Transactions on Communications*, COM-31(3):343–51, March 1983.

[2] B. Bollobas. "The diameter of random graphs." *Transactions of the American Mathematical Society*, 267:41–52, September 1981.

[3] E.N. Gilbert. "Random graphs." *Annals of Mathematical Statistics*, 30:1141–44, 1959.

[4] L. Kou, G. Markowsky, and L. Berman. "A fast algorithm for Steiner trees." *Acta Informatica*, 15:141–5, 1981.

[5] R.E. Miller and J.W. Thatcher. *Complexity of Computer Computations*, chapter Reducibility among combinatorial problems - by R.M. Karp, pages 85–103. Plenum Press, 1972.

[6] V.J. Rayward-Smith. "The computation of nearly minimal Steiner trees in graphs." *International Journal of Math. Ed. Sci. Tech.*, 14(1):15–23, 1983.

[7] V.J. Rayward-Smith and A. Clare. "On finding Steiner vertices." *Networks*, 16:283–94, 1986.