The Minimum Spanning Tree Constant in Geometrical Probability and under the Independent Model; a Unified Approach

by

Florin Avram & Dimitris J. Bertsimas

OR 211-90 April 1990

The Minimum Spanning Tree Constant in Geometrical Probability and under the Independent Model; a Unified Approach

Florin Avram * Dimitris J. Bertsimas †

April 11, 1990

Abstract

Given n uniformly and independently points in the d dimensional cube of unit volume, it is well established that the length of the minimum spanning tree on these n points is asymptotic to $\beta_{MST}(d)n^{(d-1)/d}$, where the constant $\beta_{MST}(d)$ depends only on the dimension d. It has been a major open problem to determine the constant $\beta_{MST}(d)$. In this paper we obtain an exact expression of the constant $\beta_{MST}(d)$ as a series expansion. Truncating the expansion after a finite number of terms yields a sequence of lower bounds; the first 3 terms give a lower bound which is already very close to the empirically estimated value of the constant. Our proof technique unifies the derivation for the MST asymptotic behavior for the Euclidean and the independent model.

*Florin Avram, Department of Mathematics, Northeastern University.

†Dimitris Bertsimas, Sloan School of Management, MIT, Rm E53-359, Cambridge, Ma 02139.

1 Introduction

Research in the area of probabilistic analysis of combinatorial optimization problems in Euclidean spaces was initiated by the pioneering paper by Beardwood, Halton and Hammersley [3], where the authors prove the following remarkable result:

Theorem ([3]): If X_i are independent and uniformly distributed points in a region of R^d with volume a, then the length L_{TSP} of the traveling salesman tour (TSP) under the usual Euclidean metric through the points $X_1, \ldots, X_n$ almost surely satisfies

$$\lim_{n \to \infty} \frac{L_{TSP}}{n^{(d-1)/d}} = \beta_{TSP}(d) a^{1/d},$$

where $\beta_{TSP}(d)$ is a constant that depends only on the dimension d.

This result was generalized to other combinatorial problems defined on Euclidean spaces, including the minimum spanning tree (MST) ([13]), the minimum matching (M) ([10]), the Steiner tree (ST) ([12]), the Held and Karp (HK) lower bound for the TSP ([6]) and other problems. Indeed, Steele [12] generalized the previous theorem for a class of combinatorial problems called subadditive Euclidean functionals. These theorems assert that there exist constants that depend on the dimension d and on the functional F involved, such that $\lim_{n \to \infty} \frac{L_F}{n^{(d-1)/d}} = \beta_F(d)$ almost surely. Unfortunately the exact value of the constants $\beta_F(d)$ is not known for any interesting functional F. One of the important open problems in this area is the exact determination of these constants.

In a different direction researchers started the investigation of the values of combinatorial optimization problems under the independent model, in which the distances d_{ij} are independent and identically distributed random variables with a common cdf $F(x)$. Karp [8] introduced the model and analyzed the TSP and the assignment problem in [9]. Frieze [5] and Steele [11] analyzed the MST and proved that the MST converges in probability as $n \to \infty$ to $\zeta(3) = \sum_{k=1}^{\infty} \frac{1}{k^3}$ under the assumption that the d_{ij} are uniformly distributed. Until now the analysis under the independent and the Euclidean model use entirely different techniques. We believe that another important problem in the area is the unification of both models so that results for one model can be used for the other.

In this paper we make progress in both these directions for the MST. In particular we obtain an exact expression for the MST constant $\beta_{MST}(d)$ under the Euclidean toroidal model as a series expansion. The main reason we use the toroidal model is to avoid disturbing boundary effects. We believe, however, that the constant is the same with the usual Euclidean model. Moreover, our techniques generalize to the independent model. In this way we derive both these results in a very similar way, thus obtaining a certain degree of unification between the two models.

The paper is structured as follows. In the next section we introduce a set of conditions under which we can characterize the MST constant as a series expansion. In Section 3 we prove that the Euclidean toroidal model satisfies these conditions and therefore we find exactly $\beta_{MST}(d)$. In Section 4 we prove that the independent model also satisfies these conditions and thus we find the known results for the MST in the independent model in a simpler way. In Section 5 we use the series expansion from Section 3 to find better bounds for the MST in the plane. The last Section includes some concluding remarks.

2 The MST in a Unified Model

In this section we introduce the following model. We are given a set of distances d_{ij}, $1 \le i,j \le n$ of random distances with $d_{ii} = 0$ and $d_{ij} = d_{ji}$. We assume that the distances d_{ij} satisfy the following conditions:

1. (Isotropy of the points). The distribution of the random vectors $\{d_{1j}\}_{j=1}^{n}, \ldots, \{d_{nj}\}_{j=1}^{n}$ is exchangeable.

2. There exists a constant M so that $d_{ij} \le M$ almost surely.

3. If $F(x) = Pr\{d_{ij} \le x\}$ we assume that there exist constants d, c_d such that
$$\lim_{x \to 0} \frac{F(x)}{c_d x^d} = 1.$$

4. Let $G_n(z)$ denote the graph of all distances which are smaller than z and let $P_{k,n}(z) = Pr\{$a given point belongs to a component of $G_n(z)$ having exactly k elements$\}$. Fix k. We assume that the probabilities $P_{k,n}(z)$ satisfy:
$$\lim_{n \to \infty} P_{k,n}[(\frac{y}{nc_d})^{1/d}] = f_k(y).$$

5. For any $n > k$
$$P_{k,n}[(\frac{y}{nc_d})^{1/d}] \le l_k(y), \quad \text{where}$$
$$\int_0^{\infty} l_k(y) y^{1/d-1} dy < \infty.$$

6. For all $\epsilon > 0$ there exists a K (independent of n) such that
$$n^{1/d} \int_0^{\infty} [\sum_{k=K}^{n} \frac{1}{k} P_{k,n}(z) - \frac{1}{n}] dz < \epsilon.$$

Conditions 1, 2 are not crucial but convenient to work with. In condition 3 we ascribe to the independent model the same marginal distances as those of the d dimensional Euclidean model, thus creating a "d dimensional independent model". Informally, the dependence in the Euclidean model comes from the fact that neighboring spheres intersect, while in the d dimensional independent model their intersection is always void. Condition 4 will be seen below to be the natural scaling condition, which indeed leads to an expansion of $\beta_{MST}(d)$ in the parameters k and y. Conditions 5 and 6 stipulate that the contribution of large k and y becomes negligible, ensuring thereby the validity of the expansion.

To understand the scaling in condition 4 in the case of the Euclidean model, it is helpful to consider another model asymptotically equivalent but sometimes more convenient to work with, obtained by randomizing the number n of points in the torus, i.e. replacing it with a Poisson number of points with expectation n. Then, the points on the torus become a Poisson point process with intensity n. If we further rescale this model by a linear magnification factor of $n^{\frac{1}{d}}$ our point process becomes the restriction to the torus $[0, n^{\frac{1}{d}}]^d$ of a Poisson point process with intensity 1. For this model it is clear that $P_{k,n}[(\frac{y}{c_d})^{1/d}]$ converges to $f_k(y)$, where $f_k(y)$ represents now the probability that a given point belongs exactly to a k cluster in the graph $G((\frac{y}{c_d})^{1/d})$. In fact, the model as a whole converges to the Poisson point process with intensity 1. This approach, advocated by Aldous and Steele [1] reduces in effect the problem of computing the length of the MST constant to the problem of computing the "average" edge in the minimal tree build on a Poisson point process of intensity one.

We can now state and prove our main theorem.

Theorem 1 *Let T_n denote the length of the MST for a model that satisfies conditions 1-6 above. Then T_n satisfies:*

$$\beta_{MST}(d) = \lim_{n\to\infty} \frac{E[T_n]}{n^{(d-1)/d}} = \frac{1}{d(c_d)^{1/d}} \sum_{k=1}^{\infty} \frac{1}{k} \int_0^{\infty} f_k(y) y^{1/d-1} dy. \qquad (1)$$

Proof

Let $C_n(z)$ denote the number of components in the graph $G_n(z) = \{(i,j)|\ d_{ij} \leq z\}$. Then

$$T_n = \int_0^{\infty} [C_n(z) - 1] dz.$$

Indeed, let $0 = z_n < z_{n-1} < z_{n-2} \ldots < z_1$ be the distances, at which the graph $G_n(z)$ attains $n-1, n-2, \ldots, 1$ components. Then

$$\int_0^{\infty} [C_n(z)-1]dz = (n-1)z_{n-1} + (n-2)[z_{n-1} - z_{n-2}] + \ldots + 1[z_1 - z_2] = \sum_{j=1}^{n-1} z_j = T_n.$$

Note that in the last equation we used the fact that the greedy algorithm solves the MST so that indeed $\sum_{j=1}^{n-1} z_j = T_n$.

Since $C_n(z) - 1 \geq 0$ by Fubini-Tonelli's theorem, we have that

$$E[T_n] = \int_0^{\infty} E[C_n(z) - 1] dz. \qquad (2)$$

Introducing the indicator random variables: $X_{i,k}(z)$, which is 1 if point i belongs to a component of $G_n(z)$ with exactly k elements and 0 otherwise, we have that

$$C_n(z) = \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{X_{i,k}(z)}{k}.$$

Taking expectations and using condition 1 about the exchangability of the points we obtain that

$$E[C_n(z)] = n \sum_{k=1}^{n} \frac{P_{k,n}(z)}{k},$$

where $P_{k,n}(z) = Pr\{$a given point belongs to a component of $G_n(z)$ having exactly k elements$\}$.

Therefore, from (2)

$$\frac{E[T_n]}{n^{(d-1)/d}} = n^{1/d} \int_0^\infty [\sum_{k=1}^{n} \frac{P_{k,n}(z)}{k} - \frac{1}{n}] dz.$$

Taking limits we obtain

$$\beta_{MST}(d) = \lim_{n\to\infty} \frac{E[T_n]}{n^{(d-1)/d}} =$$

$$lim_{n\to\infty}[\int_0^\infty \sum_{k=1}^{K-1} \frac{1}{k} P_{k,n}(z) n^{1/d} dz + n^{1/d} \int_0^\infty (\sum_{k=K}^{n} \frac{1}{k} P_{k,n}(z) - \frac{1}{n}) dz].$$

Then from condition 6

$$\beta_{MST}(d) = \lim_{n\to\infty} [\sum_{k=1}^{K-1} \frac{1}{k} \int_0^\infty P_{k,n}(z) n^{1/d} dz] + \epsilon(K).$$

We introduce now the limit inside the finite summation. By making the change of variables $z = (\frac{y}{nc_d})^{1/d}$ we have

$$\beta_{MST}(d) = \frac{1}{d(c_d)^{1/d}} \sum_{k=1}^{K-1} \frac{1}{k} \lim_{n\to\infty} \int_0^\infty P_{k,n}[(\frac{y}{nc_d})^{1/d}] y^{1/d-1} dy + \epsilon(K).$$

Using condition 5 and the dominated convergence theorem we exchange the limit and the integration operation and use condition 4. Finally (1) follows from condition 6 by letting $K \to \infty$. □

Remark: Steele [13] considers the asymptotic behavior of the MST with power weighted edges, i.e. $T_n(a) = \sum_{i=1}^{n-1} z_i^a$, with $0 < a < d$. Using a straightforward modification of our method we can find that

Theorem 2 *Under assumptions 1, 2, 3, 4 and*

$5'$. *For any* $n > k$ $P_{k,n}[(\frac{y}{nc_d})^{1/d}] \le l_k(y)$, where $\int_0^\infty l_k(y) y^{a/d-1} dy < \infty$.

$6'$. *For all $\epsilon > 0$ there exists a K (independent of n) such that*

$$n^{a/d} \int_0^\infty [\sum_{k=K}^{n} \frac{1}{k} P_{k,n}(z) - \frac{1}{n}] dz^a < \epsilon,$$

the MST with power weights satisfies for $0 < a < d$

$$\beta_{MST}(d,a) = \lim_{n\to\infty} \frac{E[T_n(a)]}{n^{(d-a)/d}} = \frac{a}{d(c_d)^{a/d}} \sum_{k=1}^{\infty} \frac{1}{k} \int_0^\infty f_k(y) y^{a/d-1} dy. \square \quad (3)$$

For $a = d$ the subadditivity techniques of Steele [13] do not seem to work and it required the new techniques of Aldous and Steele [1] to prove that indeed $E[T_n(d,d)]$ converges (the result was first conjectured by Bland when $a = d = 2$ based on experimental evidence). Our expansion (3) is still valid for $a = d$.

We now prove in the next section that the Euclidean toroidal model satisfies the conditions 1-6.

3 The MST in the Euclidean toroidal model

We consider now the Euclidean toroidal model, i.e. the metric space $[-\frac{1}{2}, \frac{1}{2}]^k$, where boundary points are identified if their coordinates are equal *mod* 1, the distance between two points is the distance between one of them to the closest preimage of the other in R^d and the measure is the Lebesque measure. Conditions 1, 2, 3 hold obviously with c_d being the volume of the ball in dimension d with unit radius. In lemma 3 below we prove that the conditions 4, 5 are also satisfied.

Lemma 3 *In the Euclidean toroidal model, conditions 4, 5 hold with $f_k(y)$ and $l_k(y)$ defined in (7) and (8) respectively below.*

Proof

Let $x_0 = 0$. Let $B_k'(z)$ denote the set of all $\{x_0, x_1, x_2, \ldots, x_{k-1}\}$ such that the "torus" spheres $S'(x_j, \frac{z}{2})$ $j = 0, \ldots, k-1$ form a connected set. Another way to define $B_k'(z)$ is that it is the set of all points $\{x_0, x_1, x_2, \ldots, x_{k-1}\}$ such that there exists a tree on $\{x_0, x_1, x_2, \ldots, x_{k-1}\}$ with all distances less or equal to z. For $k = 1$ we define $B_1'(z)$ to be the entire torus. As an example, $B_2'(z) = S'(0, z)$.

Let $g_{k,z}'(x_0, x_1, x_2, \ldots, x_{k-1})$ denote the volume of $\cup_{j=0}^{k-1} S'(x_j, z)$, where $x_0 = 0$. With these definitions we obtain

$$P_{k,n}(z) = \binom{n-1}{k-1} \int_{B_k'(z)}' [1 - g_{k,z}'(x_0, x_1, x_2, \ldots, x_{k-1})]^{n-k} dx_1 \ldots dx_{k-1}, \quad (4)$$

where $\int'$ denotes integration over the $(k-1)$ times product of the d-dimensional torus with itself. Moreover, $P_{k,n}(z) = 0$ if $z \geq \sqrt{d}/2$ and $n > k$.

Symmetric sets which do not touch the torus boundary are identical on the torus and in R^{k-1} and thus for $z \leq \frac{1}{2k}$

$$P_{k,n}(z) = \binom{n-1}{k-1} \int_{B_k(z)} [1 - g_{k,z}(x_0, x_1, x_2, \ldots, x_{k-1})]^{n-k} dx_1 \ldots dx_{k-1}, \quad (5)$$

where $B_k(z)$ is the set of all $\{x_0, x_1, x_2, \ldots, x_{k-1}\}$ such that the spheres $S(x_j, \frac{z}{2})$ $j = 0, \ldots, k-1$ form a connected set, $g_{k,z}(x_0, x_1, x_2, \ldots, x_{k-1})$ is the volume of $\cup_{j=0}^{k-1} S(x_j, z)$, and the integral is a usual multiple integral in $R^{d(k-1)}$.

By changing variables in (5) to $u_{j,i} = \frac{x_{j,i}}{z}$, $i = 1, \ldots, d$, $j = 1, \ldots, k-1$ and noting that $g_{k,z}(zu_0, zu_1, zu_2, \ldots, zu_{k-1}) = z^d g_{k,1}(u_0, u_1, u_2, \ldots, u_{k-1})$ we get that for $z \leq \frac{1}{2k}$:

$$P_{k,n}(z) = \binom{n-1}{k-1} z^{d(k-1)} \int_{B_k(1)} [1 - z^d g_{k,1}(u_0, u_1, u_2, \ldots, u_{k-1})]^{n-k} du_1 \ldots du_{k-1}.$$

Rescaling to $z = (\frac{y}{nc_d})^{1/d}$ we obtain

$$P_{k,n}([\frac{y}{nc_d}]^{1/d}) = \frac{y^{k-1}}{c_d^{k-1}(k-1)!}\prod_{j=1}^{k-1}(1-\frac{j}{n})$$

$$\int_{B_k(1)}[1-\frac{y}{nc_d}g_{k,1}(u_0,u_1,u_2,\ldots,u_{k-1})]^{n-k}du_1\ldots du_{k-1}. \quad (6)$$

Since the integrand and the domain in the RHS of (6) are bounded, we get

$$f_k(y) = \lim_{n\to\infty} P_{k,n}([\frac{y}{nc_d}]^{1/d}) = \frac{y^{k-1}}{c_d^{k-1}(k-1)!}$$

$$\int_{B_k(1)} e^{-\frac{y}{c_d}g_{k,1}(u_0,u_1,u_2,\ldots,u_{k-1})}du_1\ldots du_{k-1}, \quad (7)$$

yielding condition 4. Note that since we are only interested in the limit as n goes to infinity, $(\frac{y}{nc_d})^{1/d} \leq \frac{1}{2k}$, and thus we can indeed apply (6).

We now turn to condition 5. Fix k and y. As we mentioned before, when $z \geq \sqrt{d}/2$, $P_{k,n}(z) = 0$. For $z < \sqrt{d}/2$ we will use:

$$g'_{k,z}(x_0,x_1,\ldots,x_{k-1}) \geq g'_{1,z/\sqrt{d}}(x_0,x_1) = c_d(\frac{z}{\sqrt{d}})^d$$

and since $1-x \leq e^{-x}$ we have:

$$P_{k,n}(z) \leq \binom{n-1}{k-1}\int'_{B'_k(z)} e^{-(n-k)g'_{k,z}(x_0,x_1,x_2,\ldots,x_{k-1})}dx_1\ldots dx_{k-1}$$

$$\leq \frac{n^{k-1}}{(k-1)!}e^{-(n-k)z^d c_d d^{-d/2}}\int_{B'_k(z)}dx_1\ldots dx_{k-1}.$$

Since $B'_k(z) \subseteq B_k(z)$ and $\frac{n-k}{n} \geq \frac{1}{k+1}$ we get that

$$P_{k,n}(z) \leq \frac{n^{k-1}}{(k-1)!}e^{-\frac{nc_d z^d}{(k+1)d^{d/2}}}k^{k-2}(c_d z^d)^{k-1}.$$

As a result,

$$P_{k,n}[(\frac{y}{nc_d})^{1/d}] \leq l_k(y) = \frac{y^{k-1}}{(k-1)!}k^{k-2}e^{-\frac{y}{(k+1)d^{d/2}}}, \quad (8)$$

and obviously $\int_0^\infty l_k(y) y^{1/d-1} dy < \infty$.□

We now turn our attention to condition 6.

Lemma 4 *In the Euclidean toroidal model, condition 6 is satisfied.*

Proof

Let $C_{K,n}(z)$ be the number of components having more than K points in the graph $G_n(z)$. As in the proof of theorem 1

$$C_{K,n}(z) = n \sum_{k=K}^{n} \frac{P_{k,n}(z)}{k}.$$

As a result,

$$n^{1/d} \int_0^\infty [\sum_{k=K}^{n} \frac{1}{k} P_{k,n}(z) - \frac{1}{n}] dz =$$
$$\frac{1}{n^{1-1/d}} \int_0^\infty (E[C_{K,n}(z)] - 1) dz.$$

As z increases it is clear that $C_{K,n}(z)$ will vary (both increase and decrease). Let $z_{i,K}^+, z_{i,K}^-$ be the lengths of the edges $l_{i,K}^+, l_{i,K}^-$ whose addition makes $C_{K,n}(z)$ to increase and decrease respectively. Let J the set of indices. Summation by parts yields that

$$\int_0^\infty (C_{K,n}(z) - 1) dz = \sum_{i \in J} [z_{i,K}^- - z_{i,K}^+] \leq \sum_{i \in J} z_{i,K}^-.$$

Our goal is to bound $\sum_{i \in J} z_{i,K}^-$ The edges $l_{i,K}^-$ connect components with at least K points. The edges $l_{i,K}^-$ do not form a tree, but rather a "pseudotree", in the sense that they connect clusters of points rather than individual points. This "pseudotree" has the smallest cost among all possible "pseudotrees". From each component in J choose an arbitrary point. Form the MST among the representatives. This tree has clearly largest cost than $\sum_{i \in J} z_{i,K}^-$, since it is also a "pseudotree" combining the clusters. But, in the Euclidean plane in dimension d the MST among any r points is less than $k_d r^{\frac{d-1}{d}}$ for some constant k_d. Therefore, $\sum_{i\in J} z_{i,K}^{-} \leq k_d |J|^{\frac{d-1}{d}}$. Since $|J| \leq \frac{n}{K}$, then

$$n^{1/d}\int_0^\infty [\sum_{k=K}^{n} \frac{1}{k} P_{k,n}(z) - \frac{1}{n}]dz$$

$$\leq \frac{1}{n^{1-1/d}} k_d (\frac{n}{K})^{\frac{d-1}{d}} = k_d \frac{1}{K^{1-1/d}} \epsilon \to 0,$$

as $K \to \infty \Box$.

Combining theorem 1 and lemmas 3, 4 we can now find a series expansion for the MST constant as follows:

Theorem 5 *In the Euclidean toroidal model*

$$\beta_{MST}(d) = \lim_{n\to\infty} \frac{E[T_n]}{n^{(d-1)/d}} = \frac{1}{d(c_d)^{1/d}} \sum_{k=1}^{\infty} \frac{1}{k} \int_0^\infty f_k(y) y^{1/d-1} dy, \qquad (9)$$

where $c_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$ *is the volume of the ball of unit radius in dimension* d,

$$f_k(y) = \frac{y^{k-1}}{c_d^{k-1}(k-1)!} \int_{B_k(1)} e^{-\frac{y}{c_d} g_{k,1}(u_0,u_1,u_2,\ldots,u_{k-1})} du_1 \ldots du_{k-1},$$

where the integration is performed on the set $B_k(1)$ *of all points* $\{u_0, u_1, u_2, \ldots, u_{k-1}\}$ ($u_0 = 0$) *such that the spheres* $S(u_j, \frac{1}{2})$ $j = 0, \ldots, k-1$ *form a connected set and* $g_{k,1}(u_0, u_1, u_2, \ldots, u_{k-1})$ *is the volume of* $\cup_{j=0}^{k-1} S(u_j, 1)$, *where* $u_0 = 0$.

As it is evident from the previous theorem the functions $f_k(y)$ are increasingly harder to obtain analytically as k increases. In section 5 we use the lower terms of this expansion to improve the best known lower bounds for the MST constant for $d = 2$.

Remark: If we define $h_d((\frac{y}{c_d})^{1/d}) = \sum_{k=1}^{\infty} \frac{f_k(y)}{k}$ to be the number of clusters per site (the free energy) in the continuous percolation model of spheres with radius $(\frac{y}{c_d})^{1/d}$ centered at points distributed according to a Poisson process with intensity 1 and we perform the change of variables $r = (\frac{y}{c_d})^{1/d}$ we see that the MST constant is the integral of the number of clusters per site then we can write

$$\beta_{MST}(d) = \int_0^\infty h_d(r)dr.$$

It is also natural to ask whether the above theorem which holds under the Euclidean toroidal model also holds under the usual Euclidean model. We only used the toroidal model to avoid the boundary effects. We conjecture that the boundary effects become negligible in the limit and thus theorem 5 also holds under the usual Euclidean model.

4 The MST in the independent model

In this section we consider the case when d_{ij} $1 \leq i, j, \leq n$ are i.i.d. random variables, whose distribution $F(x)$ satisfies:

$$\lim_{x \to 0} \frac{F(x)}{c_d x^d} = 1, \quad F(x) = 1 \quad \forall x \geq M.$$

Again conditions 1, 2, 3 hold trivially. In the following lemma we prove that conditions 4, 5 are also satisfied.

Lemma 6 *In the independent model, conditions 4, 5 hold with $f_k(y)$ and $l_k(y)$ defined in (11) and (12) respectively below.*

Proof

In this case

$$P_{k,n}(z) = \binom{n-1}{k-1} \sum_{j=k-1}^{\binom{k}{2}} [F(z)]^j [1 - F(z)]^{k(n-k)+\binom{k}{2}-j} N_{k,j}, \tag{10}$$

where $N_{k,j}$ is the number of connected graphs with j edges and k vertices. For example $N_{k,k-1} = k^{k-2}$. Let

$$h_{k,n}(z) = \binom{n-1}{k-1}[F(z)]^{k-1}(1-F(z))^{k(n-k)+\binom{k}{2}-k+1}N_{k,k-1}$$

be the first term in the sum above. Then, it is easy to establish that

$$\lim_{n\to\infty} h_{k,n}[(\frac{y}{nc_d})^{1/d}] = \frac{k^{k-2}}{(k-1)!}y^{k-1}e^{-ky}.$$

Since the contribution to the limit of the other terms in the RHS of (10) is 0 we obtain:

$$f_k(y) = \lim_{n\to\infty} P_{k,n}([(\frac{y}{nc_d})^{1/d}]) = \frac{k^{k-2}}{(k-1)!}y^{k-1}e^{-ky}, \tag{11}$$

yielding condition 4.

To establish condition 5 we note again that $P_{k,n}([(\frac{y}{nc_d})^{1/d}]) = 0$ for $(\frac{y}{nc_d})^{1/d} \geq M$. For $z = (\frac{y}{nc_d})^{1/d} \leq M$ we can find 2 constants a, A such that

$$az^d \leq F(z) \leq Az^d.$$

Let $M_k = \sum_{j=k-1}^{\binom{k}{2}} N_{k,j}$. From (10) we obtain

$$P_{k,n}(z) \leq \frac{1}{(k-1)!}[nF(z)]^{k-1}e^{-knF(z)}e^{k^2}M_k,$$

from which

$$P_{k,n}[(\frac{y}{nc_d})^{1/d}] \leq l_k(y) = \frac{1}{(k-1)!}(\frac{A}{c_d})^{k-1}e^{k^2}M_k y^{k-1}e^{-aky/c_d} \tag{12}$$

and thus condition 5 holds. □

We now turn our attention to condition 6.

Lemma 7 *In the independent model, condition 6 is satisfied.*

Proof

This proof can be done along the lines of Frieze [5] by splitting

$$n^{1/d}\int_0^{\infty}[\sum_{k=K}^{n}\frac{1}{k}P_{k,n}(z)-\frac{1}{n}]dz$$

in three terms:

$$A_1 = n^{1/d}\int_0^{M}[\sum_{k=K}^{\gamma n^{1/2}}\frac{1}{k}P_{k,n}(z)-\frac{1}{n}]dz,$$

$$A_2 = n^{1/d}\int_0^{(\frac{\lambda\log n}{nc_d})^{1/d}}[\sum_{k=\gamma n^{1/2}}^{n}\frac{1}{k}P_{k,n}(z)-\frac{1}{n}]dz,$$

$$A_3 = n^{1/d}\int_{(\frac{\lambda\log n}{nc_d})^{1/d}}^{M}[\sum_{k=\gamma n^{1/2}}^{n}\frac{1}{k}P_{k,n}(z)-\frac{1}{n}]dz,$$

where $\gamma = \frac{a}{2A}$ and $\lambda > 4+\frac{2}{d}$. Using the techniques of Frieze [5] we prove that each of the terms A_1, A_2, A_3 can be made arbitrarily small for large K. We omit the details since they are similar to the paper by Frieze [5].□

Combining lemmas 6, 7 and theorem 1 we can find the following expression for the MST under the independent model:

Theorem 8 *Under the independent model*

$$\lim_{n\to\infty}\frac{E[T_n]}{n^{(d-1)/d}} = \frac{1}{d(c_d)^{1/d}}\sum_{k=1}^{\infty}\frac{\Gamma(k+\frac{1}{d}-1)}{k!k^{\frac{1}{d}+1}}.$$

Proof

From (1) and (11) we have

$$\lim_{n\to\infty}\frac{E[T_n]}{n^{(d-1)/d}} = \frac{1}{d(c_d)^{1/d}}\sum_{k=1}^{\infty}\frac{1}{k!k^{1/d+1}}\int_0^{\infty}(ky)^{k-1}e^{-ky}(ky)^{1/d-1}d(ky) =$$

$$\frac{1}{d(c_d)^{1/d}}\sum_{k=1}^{\infty}\frac{\Gamma(k+\frac{1}{d}-1)}{k!k^{\frac{1}{d}+1}}.\square$$

Remark: The MST with power weights has the following expansion for $0 < a < d$

$$\lim_{n\to\infty} \frac{E[T_n(a,d)]}{n^{(d-a)/d}} = \frac{a}{d(c_d)^{a/d}} \sum_{k=1}^{\infty} \frac{\Gamma(k+\frac{a}{d}-1)}{k!k^{\frac{a}{d}+1}}.$$

For $a = d = 2$ the expansion gives $\zeta(3)/\pi$.

For $d = 1$ and $c_1 = 1$, the distances are uniformly distributed and thus we get the result due to Frieze [5] that $\lim_{n\to\infty} E[T_n] = \zeta(3)$. For general d the same result is obtained by Timofeev [14] who analyzes Prim's algorithm, while we analyzed Kruskal's algorithm.

5 Improved lower bounds in the Euclidean model

We now turn our attention to the derivation of better bounds for the MST constant under the Euclidean model for $d = 2$. Using theorem 5 we compute the contribution of the first three terms in the expansion (9).

Theorem 9 *For $d = 2$ the Euclidean MST constant satisfies :*

$$\beta_{MST}(2) \geq 0.599. \tag{13}$$

Proof

For $k = 1$, $B_1(1)$ is the entire area and $g_1(u_0) = c_d$. Thus from Theorem 5 $f_1(y) = e^{-y}$.

For $k = 2$, $B_2(1)$ is the set of points x_1 such that the two spheres with centers u_1 and 0 of radius $\frac{1}{2}$ intersect. If $u_1 = (r, \theta)$, $r \leq 1$, are the polar coordinates of u_1, then $g_{2,1}(u_0, u_1) = 2\pi - \phi(r)$, where $\phi(r)$ is the area of intersection of two spheres of unit radius at distance r apart. From simple trigonometry we can derive $\phi(r)$ as follows:

$$\phi(r) = 2cos^{-1}(\frac{r}{2}) - r\sqrt{1-(\frac{r}{2})^2}.$$

As a result,

$$f_2(y) = \frac{y}{\pi}\int_{\theta=0}^{2\pi} d\theta \int_{r=0}^{1} e^{-y(2\pi-\phi(r))/\pi} r dr = 2y\int_{r=0}^{1} e^{-y(2\pi-\phi(r))/\pi} r dr.$$

Using only the first two terms in the expansion we obtain

$$\beta_{MST} \geq \frac{1}{2\sqrt{\pi}}[\int_{y=0}^{\infty} e^{-y} y^{-\frac{1}{2}} dy + \int_{y=0}^{\infty}\int_{r=0}^{1} e^{-y\frac{2\pi-\phi(r)}{\pi}} y^{\frac{1}{2}} r dr dy].$$

Performing the integration of the first term and interchanging integrals in the second term we obtain

$$\beta_{MST} \geq \frac{1}{2} + \frac{\pi^{3/2}}{4}\int_{r=0}^{1} \frac{r}{(2\pi - \phi(r))^{3/2}} dr.$$

Using the software package *Mathematica* to perform the integral numerically we find that the contribution of the first two terms gives $\beta_{MST} \geq 0.576$.

We now compute the third term in the series expansion. $B_3(1)$ is the set of points x_1, x_2 such that the three spheres with centers u_1, u_2 and 0 of radius $\frac{1}{2}$ intersect. Without loss of generality we assume that the sphere with center u_1 intersects the sphere with center 0. Let $u_1 = (t, \theta)$, $t \leq 1$, be the polar coordinates of u_1. If we rotate so that $\theta = 0$, let $u_2 = (r, \phi)$ be the polar coordinates of the second point in the new rotated coordinate system. Then the region can be partitioned into three areas (see figure 1).

In areas A, C the third sphere intersects only one other sphere, while in area B all the three spheres intersect. As a result, the function $f_3(y)$ can be written as follows:

$$f_3(y) = \frac{y^2}{2\pi^2}\int_{\theta=0}^{2\pi} d\theta \int_{t=0}^{1}\int_{A\cup B\cup C} e^{-y g_{3,1}(t,r,\phi)/\pi} r t dr dt d\phi.$$

From symmetry the integrals over A and C are equal. Moreover, in region $A = \{(r,\phi)/1 \leq r \leq t+1, -cos^{-1}(\frac{r^2+t^2-1}{2rt}) \leq \phi \leq cos^{-1}(\frac{r^2+t^2-1}{2rt})\}$

$$g_{3,1}(t, r, \phi) = 3\pi - \phi(r) - \phi(t),$$

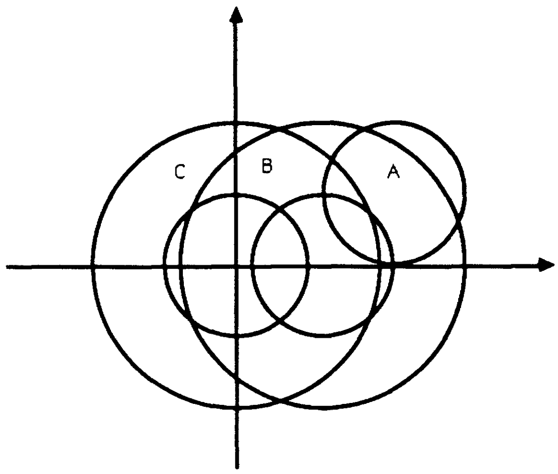

Figure 1: Three Intersecting Spheres

where $\phi(x) = 2cos^{-1}(\frac{x}{2}) - x\sqrt{1-(\frac{x}{2})^2}$, is the area of intersection of 2 spheres of unit radius, whose centers are x apart.

In region $B = \{(r,\phi)/0 \le r \le 1, -cos^{-1}(\frac{r^2+t^2-1}{2rt}) \le \phi \le cos^{-1}(\frac{r^2+t^2-1}{2rt})\}$

$$g_{3,1}(t,r,\phi) = 3\pi - \phi(r) - \phi(t) - \phi(\sqrt{r^2+t^2-2rtcos\phi}) + h(t,r,\phi),$$

where $h(t,r,\phi)$ is the common area of intersection of all the three spheres. The computation of $h(t,r,\phi)$ is tedious but it can easily be done.

The third term is $\frac{1}{6\sqrt{\pi}}\int_0^\infty f_3(y)y^{-\frac{1}{2}}dy$, which after some algebraic manipulations becomes:

$$\frac{\pi^{\frac{3}{2}}}{8}[4\int_{t=0}^{1} t \int_{r=1}^{t+1} \frac{cos^{-1}(\frac{r^2+t^2-1}{2rt})}{(3\pi-\phi(r)-\phi(t))^{\frac{5}{2}}} r\, dr\, dt +$$

$$\int_{t=0}^{1} t \int_{r=0}^{1} r \int_{\phi=-cos^{-1}(\frac{r^2+t^2-1}{2rt})}^{cos^{-1}(\frac{r^2+t^2-1}{2rt})} \frac{1}{(3\pi-\phi(r)-\phi(t)-\phi(\sqrt{r^2+t^2-2rtcos\phi})+h(t,r,\phi))^{\frac{5}{2}}} drdtd\phi].$$

The computation of these integrals was a challenge for *Mathematica*. Using a Taylor expansion of the integrand up to the sixth term we evaluated these integrals to find that the contribution of the three first terms gives $\beta_{MST} \geq 0.599$, which is in excellent agreement with the experimental data. □

With more work one can potentially calculate more terms in the series expansion. The best known previous bound for β_{MST} was $\frac{1}{2}$ and it is based on the distance to the nearest neighbor, (see Bertsimas and van Ryzin [4]). Note that the previous bound corresponds to the contribution of the first term in the series expansion.

Remark: We can use the expansion to find a lower bound for the Bland constant as well, i.e. $\beta(2,2)$. Then the contribution of the first two terms gives $\beta_{MST}(2,2) \geq 0.401$.

5.1 Bounds for general dimensions

In higher dimensions one can use the series expansion for $\beta_{MST}(d)$ to find that

$$\frac{\Gamma(\frac{1}{d})}{dc_d^{1/d}} \leq \beta_{MST}(d) \leq \frac{2^{1/d}\Gamma(\frac{1}{d})}{dc_d^{1/d}}. \tag{14}$$

The lower bound corresponds to the first term in the expansion, while the upper bound uses a technique of Hall [7] (p. 264-265) that $\sum_{k=1}^{\infty} \frac{f_k(y)}{k} \leq \lim_{n\to\infty} Pr\{$a given point has no neighbors closer than $(\frac{y}{nc_d})^{1/d}$ and towards the right$\} = e^{-\frac{y}{2}}$. Note that Bertsimas and van Ryzin [4] find that the exodic tree achieves this upper bound.

6 Concluding Remarks

Our analysis for the MST constants under both the Euclidean and independent model was made possible by analyzing directly the greedy algorithm, which solves the MST exactly. We have also analyzed in [2] the greedy algorithm applied to TSP and matching in both models and found series expansions for both problems.

Another interesting observation is the relation of the two models. As $d \to \infty$ we expect that the graph $G((\frac{y}{c_d})^{1/d})$ in the Poisson model converges to a forest, i.e to a graph whose clusters have no cycles (the clusters are then branching trees with Poisson distribution of offsprings with parameter y). This is also the limit as $n \to \infty$ of the independent model. Furthermore, we expect that the number of cycles is stochastically decreasing in d (and as $d \to \infty$ it becomes zero).

Let $f_k^{(I)}(y)$, $f_k^{(E)}(y)$ be the corresponding functions in (7) and (11) for the independent and the Euclidean model respectively. From (11), $f_k^{(I)}(y)$ is independent from the dimension d, whereas $f_k^{(E)}(y)$ depends on d. From the conjectured structure of $G((\frac{y}{c_d})^{1/d})$ we expect that the following connection exists between the two models.

Conjecture

1. $$\lim_{d\to\infty} f_k^{(E)}(y) = f_k^{(I)}(y). \tag{15}$$

2. Moreover, the function $\sum_{k\geq K} \frac{f_k^{(E)}(y)}{k}$ is decreasing in d and

$$\lim_{d\to\infty} \sum_{k\geq K} \frac{f_k^{(E)}(y)}{k} = \sum_{k\geq K} \frac{f_k^{(I)}(y)}{k}. \tag{16}$$

Conjecture (1) may be easily checked for the cases $k = 1$ and 2 by direct computation. Furthermore, there are some interesting corollaries of the conjecture. For example, $\beta_{MST}^{(E)}(d) \geq \beta_{MST}^{(I)}(d)$. We can check this for $d = 2$. In this case theorem 8 for the independent model gives $\beta_{MST}^{(I)}(2) = 0.568$, i.e. the constant for the independent model provides a lower bound for the

Euclidean model. Also using conjecture (2) with $K = 4$ allows to improve our previous lower bound to $\beta^{(E)}_{MST}(2) \geq 0.61$.

Acknowledgment:

We would like to thank Professors D. Aldous and M. Steele for making available to us their unpublished work and for useful discussions.

References

[1] D. Aldous and J.M. Steele, (1990) "Asymptotics for Euclidean Minimal Spanning Trees on Random Points", personal communication.

[2] F. Avram and D. Bertsimas, (1990) "The greedy algorithm in combinatorial optimization and its probabilistic analysis", in preparation.

[3] J. Beardwood, J.H. Halton and J.M. Hammersley, (1959), "The shortest path through many points", *Proc. Cambridge Philos. Soc.*, 55, 299-327.

[4] D. Bertsimas and G. van Ryzin, (1990) "An Asymptotic Determination of the Minimum Spanning Tree and Minimum Matching Constants in Geometrical Probability", to appear in *Oper. Res. Letters.*

[5] A. Frieze, (1985) "On the Value of a Minimum Spanning Tree Problem", *Disc. Appl. Math.*, 10, 47-56.

[6] M. Goemans and D. Bertsimas, (1990) "Probabilistic analysis of the Held-Karp relaxation for the Euclidean traveling salesman problem", MIT Sloan School Working Paper, to appear in *Math. Oper. Res.*

[7] P. Hall, (1988) *Introduction to the Theory of Coverage Processes*, Wiley, New York.

[8] R. Karp , (1977) "Probabilistic Analysis of Partitioning Algorithms for the Traveling Salesman in the Plane", *Math. Oper. Res.*, 2, 209-224.

[9] R. Karp , (1977) "On the value of random assignment", manuscript.

[10] C.H. Papadimitriou, (1978) "The probabilistic analysis of matching heuristics", *Proc. 15th Annual Conference Comm. Contr. Computing*, Univ. Illinois.

[11] J.M. Steele, (1987) "On Frieze's $\zeta(3)$ Limit for Lengths of Minimal Spanning Trees", *Disc. Appl. Math.*, 18, 99-103.

[12] J.M. Steele, (1981) "Subadditive Euclidean Functionals and Nonlinear Growth in Geometric Probability", *Ann. Prob.*, 9, 365-376.

[13] J.M. Steele, (1988) "Growth Rates of Euclidean Minimal Spanning Trees with Power Weighted Edges", Annals of Prob., 1767-1787.

[14] E.A. Timofeev, (1987) "On finding the expected length of a random minimal tree", *Theory Probab. Appl.*, 361-365.