The Annals of Applied Probability
1992, Vol. 2, No. 1, 113–130

THE MINIMUM SPANNING TREE CONSTANT IN GEOMETRICAL PROBABILITY AND UNDER THE INDEPENDENT MODEL: A UNIFIED APPROACH

BY FLORIN AVRAM AND DIMITRIS BERTSIMAS

Northeastern University and Massachusetts Institute of Technology

Given n uniformly and independently distributed 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 $\beta_{\mathrm{MST}}(d)n^{(d-1)/d}$, where the constant $\beta_{\mathrm{MST}}(d)$ depends only on the dimension d. It has been a major open problem to determine the constant $\beta_{\mathrm{MST}}(d)$. In this paper we obtain an exact expression for the constant $\beta_{\mathrm{MST}}(d)$ on a torus as a series expansion. Truncating the expansion after a finite number of terms yields a sequence of lower bounds; the first five 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.

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 $X_1, \dots, X_n$ 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) ([14]), the minimum matching (M) ([11]), 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 $\lim_{n\to\infty}(L_F/n^{(d-1)/d}) = \beta_F(d)a^{1/d}$ almost surely. Unfortunately the

Received April 1990; revised April 1991.
[1]Research supported by NSF Grants DDM-90-10332 and DDM-90-14751.
AMS 1980 *subject classifications*. 60D05, 90C27.
Key words and phrases. Geometrical probability, minimum spanning tree constant, Euclidean and independent random models.

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 [9] introduced the model and analyzed the TSP and the assignment problem in [10]. Frieze [5] and Steele [13] analyzed the MST and proved that the MST converges in probability as $n \to \infty$ to $\zeta(3) = \sum_{k=1}^{\infty}(1/k^3)$ 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 $\beta_{\mathrm{MST}}(d)$ under the Euclidean toroidal model as a series expansion. 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 main reason we have used the toroidal model is to avoid disturbing boundary effects. In an earlier version of the present paper we have conjectured that the constant is the same with the usual Euclidean model. Indeed, Jaillet [8] has proved our conjecture. In this way the expansion we have obtained is also valid for the usual Euclidean model.

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 $\beta_{\mathrm{MST}}(d)$. In Section 4 we prove that the independent model also satisfies these conditions and thus we find an expansion for $\beta_{\mathrm{MST}}(d)$ in the independent model. 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 random distances d_{ij}, $1 \le i, j \le n$ with $d_{ii} = 0$ and $d_{ij} = d_{ji}$. We assume that the distribution $F(x) = \Pr\{d_{ij} \le x\}$ satisfies

$$\lim_{x \to 0} \frac{F(x)}{c_d x^d} = 1,$$

and that there exists a constant M so that $d_{ij} \le M$. In the Euclidean model d represents the dimension and c_d represents the volume of a sphere of unit radius in dimension d. Note that in the case when the d_{ij} are also independent we get an independent model with the same marginal distribution for the distances as those of the d-dimensional Euclidean model, thus creating a

d-dimensional independent model. In this unified model the distances d_{ij} satisfy the following conditions:

1. (Isotropy of the points). The matrix $\{d_{ij}\}_{i,j=1}^{n}$ has the same distribution as $\{d_{\pi(i)\pi(j)}\}_{i,j=1}^{n}$ for all permutations π.
2. 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 points$\}$. Fix k.
   We assume that there exists $f_k(y)$ such that
$$\lim_{n\to\infty} P_{k,n}\left[\left(\frac{y}{nc_d}\right)^{1/d}\right] = f_k(y).$$
3. For any $n > k$,
$$P_{k,n}\left[\left(\frac{y}{nc_d}\right)^{1/d}\right] \le l_k(y), \quad \text{where } \int_0^\infty l_k(y) y^{1/d-1}\,dy < \infty.$$
4. For all K,
$$\left| n^{1/d} \int_0^\infty \left[\sum_{k=K}^{n} \frac{1}{k} P_{k,n}(z) - \frac{1}{n} \right] dz \right| = O(K^{-(d-1)/d}),$$
   where the constant in the $O(K^{-(d-1)/d})$ is independent of n.

Condition 1 is not crucial but convenient to work with. Both the independent and the toroidal model clearly satisfy it. Condition 2 will be seen below to be the natural scaling condition, which indeed leads to an expansion of $\beta_{\text{MST}}(d)$ in the parameters k and y. Conditions 3 and 4 stipulate that the contribution of large k and y becomes negligible, ensuring thereby the validity of the expansion.

For example, in the case of the Euclidean toroidal model, the expected number of points in a ball of radius $(y/nc_d)^{1/d}$ is equal to y; the same holds for the independent model as $n \to \infty$. In the case of the Euclidean model, it is also 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, that is, 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 $n^{1/d}$, our point process becomes the restriction to the torus $[0, n^{1/d}]^d$ of a Poisson point process with intensity 1. For this model it is clear that $P_{k,n}[(y/c_d)^{1/d}]$ converges to $f_k(y)$, where $f_k(y)$ now represents the probability that a given point belongs to a component of the graph of all distances smaller than $(y/c_d)^{1/d}$ with exactly k points. 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 value of the MST constant to the problem of computing the average edge in the minimum tree defined on points generated by a Poisson point process of intensity 1.

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* 1–4 *above. Then T_n satisfies*

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

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

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

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

$$\int_0^{\infty} [C_n(z) - 1]\, dz = (n-1)z_{n-1} + (n-2)[z_{n-2} - z_{n-1}] + \cdots + 1[z_1 - z_2]$$

$$= \sum_{j=1}^{n-1} z_j = T_n.$$

Note that in the last equation we used the fact that the greedy algorithm solves the MST so that indeed $\Sigma_{j=1}^{n-1} z_j = T_n$. Note that even if there are ties $z_{i+1} = z_i$ for some i's the expansion is still valid.

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

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

Introducing the indicator random variables $X_{i,k}(z)$, which are 1 if point i belongs to a component of $G_n(z)$ with exactly k points 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 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 points$\}$.

Therefore, from (2),

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

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

Taking limits and introducing the limit inside the finite summation, we obtain

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

By making the change of variables $z = (y/nc_d)^{1/d}$ and using condition 4, we obtain

$$\lim_{n\to\infty}\frac{E[T_n]}{n^{(d-1)/d}} = \frac{1}{d(c_d)^{1/d}}\sum_{k=1}^{K-1}\frac{1}{k}\lim_{n\to\infty}\int_0^\infty P_{k,n}\left[\left(\frac{y}{nc_d}\right)^{1/d}\right]$$
$$\times y^{1/d-1}\,dy + \Theta(K^{-(d-1)/d}).$$

Using condition 3, we apply the dominated convergence theorem, exchanging the limit and the integration operation. Using condition 2 and taking $K\to\infty$ we obtain (1). □

REMARK. Steele [14] considers the asymptotic behavior of the Euclidean MST with power weighted edges, that is, $T_n(a) = \sum_{i=1}^{n-1} z_i^a$, with $0 < a < d$. Using a straightforward modification of our method we prove:

THEOREM 2. *Under conditions* 1 *and* 2 *and the following modifications of conditions* 3 *and* 4:

3′. *for any* $n > k$, $P_{k,n}[(y/nc_d)^{1/d}] \le l_k(y)$, *where* $\int_0^\infty l_k(y)y^{a/d-1}\,dy < \infty$;
4′. *for all* K (*independent of* n),

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

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

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

For $a = d$, the subadditivity techniques of Steele [14] do not seem to work and it required the new techniques of Aldous and Steele [1] to prove that indeed $E[T_n(d)]$ converges (the result was first conjectured by Bland when $a = d = 2$ based on experimental evidence). Our estimates below show that 3′

and 4′ indeed hold for $a < d$. Although our estimates break down when $a = d$, expansion (3) still makes sense for $a = d$ and it is probably correct.

We now prove in the next section that the Euclidean toroidal model satisfies the conditions 1–4.

3. The MST in the Euclidean toroidal model. We consider now the Euclidean toroidal model, that is, the metric space $[-\frac{1}{2}, \frac{1}{2}]^d$, 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 Lebesgue measure. Condition 1 obviously holds. Note that c_d is the volume of a ball in dimension d with unit radius. In Lemma 3 below we prove that the conditions 2 and 3 are also satisfied.

LEMMA 3. *In the Euclidean toroidal model, conditions* 2 *and* 3 *hold with* $f_k(y)$ *and* $l_k(y)$ *defined in* (7) *and* (8), *respectively, below.*

PROOF. Let $x_0 = 0$. For $k \ge 2$, let $B'_k(z)$ denote the set of all $\{x_1, x_2, \dots, x_{k-1}\}$ such that the torus spheres $S'(x_j, z/2)$, with center x_j and radius $z/2$, $j = 0, \dots, k-1$ form a connected set. Another way to define $B'_k(z)$ is that it is the set of all points $\{x_1, x_2, \dots, x_{k-1}\}$ such that there exists a tree on $\{x_0, x_1, x_2, \dots, x_{k-1}\}$ 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 $g'_{k,z}(x_1, x_2, \dots, x_{k-1})$ denote the volume of $\bigcup_{j=0}^{k-1} S'(x_j, z)$, where $x_0 = 0$. By definition, $P_{k,n}(z)$ is the probability that a given point belongs to a component of $G_n(z)$ having exactly k points. Then $P_{1,n}(z) = (1 - c_d z^d)^{n-1}$. For $k \ge 2$, since there are $\binom{n-1}{k-1}$ choices for the other $k - 1$ points and all the other points should be outside of $\bigcup_{j=0}^{k-1} S'(x_j, z)$, we have that

$$(4) \quad P_{k,n}(z) = \binom{n-1}{k-1} \int'_{B'_k(z)} [1 - g'_{k,z}(x_1, x_2, \dots, x_{k-1})]^{n-k} dx_1 \cdots dx_{k-1},$$

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

Sets which do not touch the torus boundary are identical on the torus and in R^d and thus for $z \le 1/2k$, $k \ge 2$,

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

where $B_k(z)$ is the set of all $\{x_1, x_2, \dots, x_{k-1}\}$ such that the spheres $S(x_j, x/2)$, $j = 0, \dots, k-1$, form a connected set, $g_{k,z}(x_1, x_2, \dots, x_{k-1})$ is the volume of $\bigcup_{j=0}^{k-1} S(x_j, z)$ and the integral is a usual $k-1$-dimensional multiple integral in R^d. For example $g_{1,z} = c_d z^d$.

By changing variables in (5) to $u_j = x_j/z$, $j = 1, \ldots, k-1$, and noting that $g_{k,z}(zu_1, zu_2, \ldots, zu_{k-1}) = z^d g_{k,1}(u_1, u_2, \ldots, u_{k-1})$ we get that for $z \le 1/2k$:

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

Rescaling to $z = (y/nc_d)^{1/d}$ we obtain for $k \ge 2$,

$$P_{k,n}\left(\left[\frac{y}{nc_d}\right]^{1/d}\right) = \frac{y^{k-1}}{c_d^{k-1}(k-1)!} \prod_{j=1}^{k-1}\left(1 - \frac{j}{n}\right) \times \int_{B_k(1)} \left[1 - \frac{y}{nc_d} g_{k,1}(u_1, u_2, \ldots, u_{k-1})\right]^{n-k} du_1 \cdots du_{k-1}. \tag{6}$$

Since the integrand and the domain in the RHS of (6) are bounded, we get for $k \ge 2$,

$$\begin{aligned} f_k(y) &= \lim_{n\to\infty} P_{k,n}\left(\left[\frac{y}{nc_d}\right]^{1/d}\right) \\ &= \frac{y^{k-1}}{c_d^{k-1}(k-1)!} \times \int_{B_k(1)} \exp[-(y/c_d) g_{k,1}(u_1, u_2, \ldots, u_{k-1})]\, du_1 \cdots du_{k-1} \end{aligned} \tag{7}$$

and

$$f_1(y) = \lim_{n\to\infty} P_{1,n}\left(\left[\frac{y}{nc_d}\right]^{1/d}\right) = \lim_{n\to\infty}\left[1 - c_d\left(\frac{y}{nc_d}\right)\right]^{n-1} = e^{-y},$$

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

We now turn to condition 3. Note that

$$P_{1,n}\left(\left[\frac{y}{nc_d}\right]^{1/d}\right) = \left[1 - c_d\left(\frac{y}{nc_d}\right)\right]^{n-1} \le e^{-y(n-1)/n} \le l_1(y) = e^{-y/2}.$$

Fix $k \ge 2$ and y. As we mentioned before, when $z \ge \sqrt{d}/2$, $P_{k,n}(z) = 0$. For $z < \sqrt{d}/2$, we will use

$$g'_{k,z}(x_1, \ldots, x_{k-1}) \ge c_d z^d,$$

and since $1 - x \le e^{-x}$, we have

$$P_{k,n}(z) \le \binom{n-1}{k-1} \int'_{B'_k(z)} \exp[-(n-k) g'_{k,z}(x_1, x_2, \dots, x_{k-1})]\, dx_1 \cdots dx_{k-1}$$

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

Since $B'_k(z) \subseteq B_k(z)$ and $(n-k)/n \ge 1/(k+1)$, we get that

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

since $k^{k-2}(c_d z^d)^{k-1}$ is an upper bound on the integral (k^{k-2} is the number of trees on k points.) As a result,

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

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

We now prove condition 4.

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

Proof. Let $C_{K,n}(z)$ be the number of components having at least K points in the graph $G_n(z)$. As in the proof of Theorem 1,

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

As a result,

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

since the distances are bounded by M. Then

$$(9) \qquad n^{1/d} \int_0^\infty \left[\sum_{k=K}^{n} \frac{1}{k} P_{k,n}(z) - \frac{1}{n} \right] dz \ge -\frac{M}{n^{(d-1)/d}} \ge -\frac{M}{K^{(d-1)/d}}.$$

As z increases, it is clear that $C_{K,n}(z)$ will vary (both increase and decrease). Let z_i^+, $i \in J^+$, be the lengths of those edges l_i^+ whose addition causes $C_{K,n}(z)$ to increase. Let z_i^-, $i \in J^-$, be the lengths of those edges l_i^- whose addition causes $C_{K,n}(z)$ to decrease. Summation by parts yields that

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

Our goal is to bound $\sum_{i \in J^-} z_i^-$. The edges l_i^- connect components with at least K points. The edges l_i^- do not form a tree, but rather a pseudotree in the sense that they connect clusters of points rather than individual points. The clusters are defined as follows. Let us number the indices $i \in J^-$ in increasing order. This means that edges with a higher index were added later. The last edge $l^-_{|J^-|}$ connects two components of $G_n(z^-_{|J^-|})$ each of which has at least K points. In each of these components, find the edge that was added later (i.e., with higher index). If a component contains no further edges it is a cluster. Apply this division recursively, always subdividing into two components, until a component contains only an edge. In this case the only two components this edge joins are defined to be clusters. The pseudotree among the clusters has the smallest length among all possible pseudotrees. From each component in J^-, choose an arbitrary point. Form the MST among the representatives. This tree clearly has larger length than $\sum_{i \in J^-} z_i^-$, 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 $k_d r^{(d-1)/d}$ for some constant k_d (Steele [14], Lemma 2.2). Therefore, $\sum_{i \in J^-} z_i^- \le k_d |J^-|^{(d-1)/d}$. Since $|J^-| \le (n/K)$, then

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

Combining the above inequality with (9) we obtain condition 4. □

Combining Theorem 1 and Lemmas 3 and 4, we can now find a series expansion for the MST constant as follows.

THEOREM 5. *In the Euclidean toroidal model,*

$$(10)\quad \beta_{MST}(d) = \lim_{n\to\infty}\frac{E[T_n]}{n^{(d-1)/d}} = \frac{1}{d(c_d)^{1/d}}\sum_{k=1}^{\infty}\frac{1}{k}\int_0^\infty f_k(y)y^{1/d-1}\,dy,$$

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

$$f_1(y) = e^{-y},$$

$$f_k(y) = \frac{y^{k-1}}{c_d^{k-1}(k-1)!}$$

$$\int_{B_k(1)}\exp[-(y/c_d)g_{k,1}(u_1,u_2,\dots,u_{k-1})]\,du_1\cdots du_{k-1},\quad k\ge 2,$$

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

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

$$(11) \qquad = \frac{\Gamma(1/d)}{d(c_d)^{1/d}} + \sum_{k=2}^{\infty} \frac{\Gamma(k+1/d-1)}{dk!}$$

$$\times \int_{B_k(1)} g_{k,1}(u_1, u_2, \dots, u_{k-1})^{-(k+1/d-1)}\, du_1 \cdots du_{k-1}.$$

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 first five terms of this expansion to improve the best known lower bounds for the MST constant for $d = 2$.

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

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

Since $\beta_{\mathrm{MST}}(d)$ is the same in both the independent and Euclidean model (Jaillet [9]), Theorem 5 is valid for the usual Euclidean model as well.

4. The MST in the independent model. In this section we consider the case when d_{ij}, $1 \le i, j, \le n, i = j$, are i.i.d. random variables, whose distribution $F(x)$ satisfies

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

Again condition 1 trivially holds. In the following lemma we prove that conditions 2 and 3 are also satisfied.

LEMMA 6. *In the independent model, conditions* 2 *and* 3 *hold with* $f_k(y)$ *and* $l_k(y)$ *defined in* (13) *and* (14), *respectively, below.*

PROOF. In order for a given point to belong in a component of $G_n(z)$ with k points, we need that among the $\binom{k}{2}$ edges joining the k points, j of them ($j = k-1, \dots, \binom{k}{2}$) are less than or equal to z and they form a connected set. Moreover, all the $k(n-k)$ edges joining these k points to all the remaining $n-k$ points should have lengths at least z. Since there are $\binom{n-1}{k-1}$ choices for the other $k-1$ points, then conditioning on j, we obtain

$$(12)\qquad P_{k,n}(z)=\binom{n-1}{k-1}\sum_{j=k-1}^{\binom{k}{2}}[F(z)]^j[1-F(z)]^{k(n-k)+\binom{k}{2}-j}N_{k,j},$$

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

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

Since the contribution to the limit of the other terms in the RHS of (12) is 0 we obtain

$$(13)\qquad f_k(y)=\lim_{n\to\infty}P_{k,n}\left(\left[\left(\frac{y}{nc_d}\right)^{1/d}\right]\right)=\frac{k^{k-2}}{(k-1)!}y^{k-1}e^{-ky},$$

yielding condition 2.

To establish condition 3 we note again that for any $k<n$, $P_{k,n}([(y/nc_d)^{1/d}])=0$ for $(y/nc_d)^{1/d}\ge M$. Since $n\to\infty$, for $z=(y/nc_d)^{1/d}\to 0$ we can find two constants a, A such that

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

Let $M_k=\sum_{j=k-1}^{\binom{k}{2}}N_{k,j}$. From (12) we obtain

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

from which

$$(14)\quad P_{k,n}\left[\left(\frac{y}{nc_d}\right)^{1/d}\right]\le l_k(y)=\frac{1}{(k-1)!}\left(\frac{A}{c_d}\right)^{k-1}e^{k^2}M_ky^{k-1}e^{-aky/c_d}$$

and thus condition 3 holds. □

We now consider condition 4.

LEMMA 7. *In the independent model, condition 4 is satisfied.*

PROOF. As in Lemma 4, let $C_{K,n}(z)$ be the number of components having at least K points in the graph $G_n(z)$. Then (9) is still valid and also

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

Taking expectations we obtain that

$$\int_0^{\infty}(E[C_{K,n}(z)] - 1)\,dz \le E\left[\sum_{i\in J^-} z_i^-\right].$$

Using the same arguments as in the proof of Lemma 4, we obtain that $E[\sum_{i\in J^-} z_i^-]$ is smaller than the expected length of the MST on the representatives. But, the expected length of the MST on r points is less than $k_d r^{(d-1)/d}$ for some constant k_d (Timofeev [15]). Therefore,

$$E\left[\sum_{i\in J^-} z_i^-\right] \le k_d E\left[|J^-|^{(d-1)/d}\right].$$

Since $|J^-| \le n/K$, then

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

Combining the above inequality with (9) we obtain condition 4. □

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

THEOREM 8. *Under the independent model,*

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

PROOF. From (1) and (13), we have

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

REMARK. The MST with power weights has the following expansion for $0 < a \le d$:

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

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

For $d = 1$ and $c_d = 1$, the distances are uniformly distributed and thus for $a = 1$, we get the result due to Frieze [5] that $\lim_{n\to\infty} E[T_n] = \zeta(3)$. For $a = 1$ and general d, we obtain the same expansion for the independent case as in Timofeev [16] 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 five terms in the expansion (11).

THEOREM 9. *For $d = 2$, the Euclidean MST constant satisfies*

$$\beta_{MST}(2) \geq 0.600822. \tag{15}$$

PROOF. The first term in the series expansion of (11) gives $\beta_{\text{MST}} \geq 0.5$.

For $k = 2$, $B_2(1)$ is the set of points u_1 such that the two circles 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 $g_{2,1}(u_1) = 2\pi - \phi(r)$, where $\phi(r)$ is the area of intersection of two circles of unit radius at distance r apart. From simple trigonometry, we can derive $\phi(r)$ as follows:

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

As a result, using only the first two terms in the expansion (11), we obtain

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

Using the software package *Mathematica* to perform the integral numerically, we find that the contribution of the first two terms $\beta_{\text{MST}} \geq 0.575957$.

We now compute the third term in the series expansion. $B_3(1)$ is the set of points u_1, u_2 such that the three circles with centers u_1, u_2 and 0 of radius $\frac{1}{2}$ intersect. Let $u_1 = (r_1, \theta_1)$, $u_2 = (r_2, \theta_2)$ be the polar coordinates of u_1 and u_2. Then the contribution of the third term is

$$\begin{aligned} I_3 &= \frac{1}{2}\frac{\Gamma(5/2)}{6}\int_{B_3(1)} g_{3,1}(u_1, u_2)^{-5/2}\, du_1\, du_2 \\ &= \frac{1}{2}\frac{\Gamma(5/2)}{6}\left[\int_{B_3(1)\cap\{r_1 \leq r_2\}} g_{3,1}(u_1, u_2)^{-5/2}\, du_1\, du_2 \right. \\ &\qquad \left. + \int_{B_3(1)\cap\{r_1 \geq r_2\}} g_{3,1}(u_1, u_2)^{-5/2}\, du_1\, du_2\right] \\ &= \frac{\Gamma(5/2)}{6}\int_{B_3(1)\cap\{r_1 \leq r_2\}} g_{3,1}(u_1, u_2)^{-5/2}\, du_1\, du_2, \end{aligned}$$

because of symmetry. Since $r_1 \leq r_2$, this implies that $r_1 \leq 1$ since the circle $S(u_1, \frac{1}{2})$ has to intersect the circle $S(0, \frac{1}{2})$. In order for the circle $S(u_2, \frac{1}{2})$ to intersect at least one of $S(u_1, \frac{1}{2})$ or $S(0, \frac{1}{2})$, u_2 has to lie either in

$$A = S(0,1) - S(0, r_1) = \{(r_2, \theta_2) | r_1 \leq r_2 \leq 1\}$$

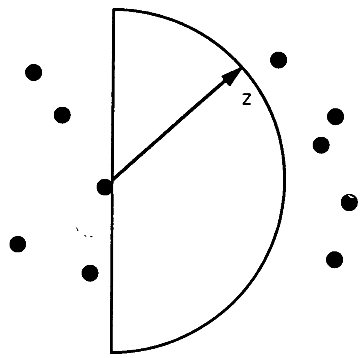

FIG. 1. *Three intersecting circles.*

or in

$$B = S(u_1, 1) - S(0,1)$$
$$= \left\{ (r_2, \theta_2) | 1 \le r_2 \le r_1 + 1, -\cos^{-1}\left(\frac{r_1^2 + r_2^2 - 1}{2r_1 r_2}\right) \le \theta_2 \le \cos^{-1}\left(\frac{r_1^2 + r_2^2 - 1}{2r_1 r_2}\right)\right\}.$$

Moreover, from simple geometry (see also Figure 1),

$$g_{3,1}(u_1, u_2) \le 2\pi - \phi(r_1) + \pi - \phi(r_2) = 3\pi - \phi(r_1) - \phi(r_2).$$

Then

$$I_3 \ge \frac{\Gamma(5/2)}{6}\left[\int_{r_1=0}^{1}\int_{\theta_1=0}^{2\pi}\int_{r_2=r_1}^{1}\int_{\theta_2=0}^{2\pi} r_1 r_2 (3\pi - \phi(r_1) - \phi(r_2))^{-5/2}\, dr_1\, d\theta_1\, dr_2\, d\theta_2 \right.$$
$$+\int_{r_1=0}^{1}\int_{\theta_1=0}^{2\pi}\int_{r_2=1}^{1+r_1}$$
$$\left.\times\int_{\theta_2=-\cos^{-1}((r_1^2+r_2^2-1)/2r_1r_2)}^{\cos^{-1}((r_1^2+r_2^2-1)/2r_1r_2)} r_1 r_2 (3\pi - \phi(r_1) - \phi(r_2))^{-5/2}\, dr_1\, d\theta_1\, dr_2\, d\theta_2\right]$$
$$= \frac{\Gamma(5/2)}{6}\left[4\pi^2 \int_{r_1=0}^{1}\int_{r_2=r_1}^{1} r_1 r_2 (3\pi - \phi(r_1) - \phi(r_2))^{-5/2}\, dr_1\, dr_2\right.$$
$$\left.+4\pi\int_{r_1=0}^{1}\int_{r_2=1}^{1+r_1}\cos^{-1}\left(\frac{r_1^2+r_2^2-1}{2r_1r_2}\right) r_1 r_2 (3\pi - \phi(r_1) - \phi(r_2))^{-5/2}\, dr_1\, dr_2\right].$$

Calculating the integrals numerically using *Mathematica* we find that the third term in the expansion is $I_3 \ge 0.021874$ and thus $\beta_{\text{MST}} \ge 0.597831$.

Building on the idea used to compute a lower bound on the third term in the series expansion we can compute a lower bound on the kth term in expansion (11) as follows: The area of integration $B_k(1)$ is larger than the sphere of radius larger than 1, since in the unit sphere all the spheres $S(u_i, \frac{1}{2})$ intersect. Then if $u_i = (r_i, \theta_i)$ in polar coordinates, then

$$g_{k,1}(u_1, u_2, \ldots, u_{k-1}) \leq k\pi - \sum_{i=1}^{k-1} \phi(r_i).$$

As a result, a lower bound on the kth term is

$$I_k \geq \frac{\Gamma(k + 1/2 - 1)}{2k!}(2\pi)^{k-1}$$
$$\times \int_{r_1=0}^{1} \cdots \int_{r_{k-1}=0}^{1} \prod_{i=1}^{k-1} r_i \left[k\pi - \sum_{i=1}^{k-1} \phi(r_i) \right]^{-(k-(1/2))} \prod_{i=1}^{k-1} dr_i.$$

Using numerical integration we find that $I_4 \geq 0.002591$ and $I_5 \geq 0.000399$. The first five terms then give

$$\beta_{\mathrm{MST}} \geq 0.600822. \qquad \square$$

With more work, one can potentially calculate better bounds using 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, that is, $\beta(2)$. Then the contribution of the first two terms gives $\beta_{\mathrm{MST}}(2) \geq 0.401$.

5.1. *Bounds for general dimensions.* In higher dimensions one can use the series expansion for $\beta_{\mathrm{MST}}(d)$ to find that

$$\frac{\Gamma(1/d)}{dc_d^{1/d}} \leq \beta_{\mathrm{MST}}(d) \leq \frac{2^{1/d}\Gamma(1/d)}{dc_d^{1/d}}. \tag{16}$$

The lower bound corresponds to the first term in the expansion (11), while the upper bound, which uses a technique of Hall [7] (pages 264–265), is derived as follows. Starting with the expansion (2) we observe that $C_n(z) \leq \sum_{i=1}^{n} Y_i(z)$, where $Y_i(z)$ are 1 when point i does not have a neighbor within z and towards the right (see Figure 2), and 0 otherwise. This is true since in every component, there exists at least one rightmost point.

Taking expectations and using the isotropy of points in the toroidal model we obtain that $E[C_n(z)] \leq n$ Pr{a given point has no neighbors closer than $(y/nc_d)^{1/d}$ and towards the right} $= n(1 - (c_d/2)z^d)^{n-1}$. Substituting this bound in (2) and making the change of variables $z = (2y/nc_d)^{1/d}$, we obtain

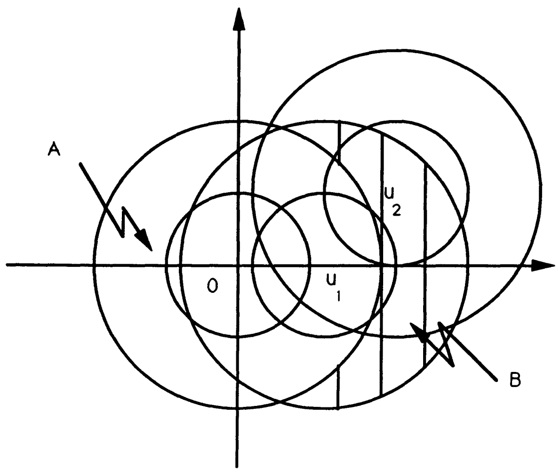

FIG. 2. *Upper bound for general dimension d.*

that

$$\beta_{\mathrm{MST}}(d) = \lim_{n\to\infty} \frac{E[T_n]}{n^{(d-1)/d}}$$
$$\le \frac{2^{1/d}}{d(c_d)^{1/d}} \int_0^\infty \lim_{n\to\infty} \left(1 - \frac{y}{n}\right)^{n-1} y^{1/d-1}\,dy$$
$$= \frac{2^{1/d}\Gamma(1/d)}{dc_d^{1/d}}.$$

Note that Bertsimas and van Ryzin [4] find that the exodic tree achieves this upper bound exactly.

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 $G((y/c_d)^{1/d})$ in the Poisson model converges to a forest, that is, 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 (12) for the independent and the Euclidean toroidal model, respectively. From (13), $f_k^{(I)}(y)$

is independent of the dimension d, whereas $f_k^{(E)}(y)$ depends on d. From the conjectured structure of $G((y/c_d)^{1/d})$ we expect that the following connection exists between the two models.

CONJECTURE 1.

$$\lim_{d\to\infty} f_k^{(E)}(y) = f_k^{(I)}(y). \tag{17}$$

CONJECTURE 2.

$$\sum_{k\ge K} \frac{f_k^{(E)}(y)}{k} \ge \sum_{k\ge K} \frac{f_k^{(I)}(y)}{k},$$
$$\lim_{d\to\infty} \sum_{k\ge K} \frac{f_k^{(E)}(y)}{k} = \sum_{k\ge K} \frac{f_k^{(I)}(y)}{k}. \tag{18}$$

Conjecture 1 may be easily checked for the cases $k = 1$ and 2 by direct computation. Furthermore, there are some interesting corollaries of Conjecture 2, for example, $\beta_{\mathrm{MST}}^{(E)}(d) \ge \beta_{\mathrm{MST}}^{(I)}(d)$. We can check this for $d = 2$. In this case, Theorem 8 for the independent model gives $\beta_{\mathrm{MST}}^{(I)}(2) = 0.568$, that is, the constant for the independent model provides a lower bound for the Euclidean model. Also using Conjecture 2 with $K = 4, 5$, we obtain that the contributions of the fourth and fifth terms, respectively, in the Euclidean model are larger than in the independent, which allows us to slightly improve our bounds to $I_4 \ge 0.004888$ and $I_5 \ge 0.002445$. Using the Euclidean model for the first three terms and the independent model for the remaining terms allows us to improve our previous lower bound (15) to $\beta_{\mathrm{MST}}^{(E)}(2) \ge 0.608701$.

Acknowledgments. We would like to thank Professors D. Aldous and M. Steele for making available to us their unpublished work and for useful discussions. We would also like to thank the referees of the paper for several insightful comments.

REFERENCES

[1] ALDOUS, D. and STEELE, J. M. (1990). Asymptotics for Euclidean minimal spanning trees on random points. Unpublished manuscript.

[2] AVRAM, F. and BERTSIMAS, D. (1990). The greedy algorithm in combinatorial optimization and its probabilistic analysis. Unpublished manuscript.

[3] BEARDWOOD, J., HALTON, J. H. and HAMMERSLEY, J. M. (1959). The shortest path through many points. *Proc. Cambridge Philos. Soc.* **55** 299–327.

[4] BERTSIMAS, D. and VAN RYZIN, G. (1990). An asymptotic determination of the minimum spanning tree and minimum matching constants in geometrical probability. *Oper. Res. Lett.* **9** 223–231.

[5] FRIEZE, A. (1985). On the value of a minimum spanning tree problem. *Discrete Appl. Math.* **10** 47–56.

[6] GOEMANS, M. and BERTSIMAS, D. (1991). Probabilistic analysis of the Held–Karp relaxation for the Euclidean traveling salesman problem. *Math. Oper. Res.* **1** 72–89.

[7] HALL, P. (1988). *Introduction to the Theory of Coverage Processes*. Wiley, New York.
[8] JAILLET, P. (1990). Personal communication.
[9] KARP, R. (1977). Probabilistic analysis of partitioning algorithms for the traveling salesman in the plane. *Math. Oper. Res.* **2** 209–224.
[10] KARP, R. (1977). On the value of random assignment. Unpublished manuscript.
[11] PAPADIMITRIOU, C. H. (1978). The probabilistic analysis of matching heuristics. *Proc. Fifteenth Ann. Conf. Comm. Contr. Computing*. Univ. Illinois.
[12] STEELE, J. M. (1981). Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probab.* **9** 365–376.
[13] STEELE, J. M. (1987). On Frieze's $\zeta(3)$ limit for lengths of minimal spanning trees. *Discrete Appl. Math.* **18** 99–103.
[14] STEELE, J. M. (1988). Growth rates of Euclidean minimal spanning trees with power weighted edges. *Ann. Probab.* **16** 1767–1787.
[15] TIMOFEEV, E. A. (1984). Random minimal trees. *Theory Probab. Appl.* **29** 134–142.
[16] TIMOFEEV, E. A. (1988). On finding the expected length of a random minimal tree. *Theory Probab. Appl.* **32** 361–365.

DEPARTMENT OF MATHEMATICS
NORTHEASTERN UNIVERSITY
BOSTON, MASSACHUSETTS 02115

SLOAN SCHOOL OF MANAGEMENT
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139