




2005

## Exact Expectations of Minimal Spanning Trees for Graphs With Random Edge Weights

James A. Fill  
*Johns Hopkins University*

J Michael Steele  
*University of Pennsylvania*

Follow this and additional works at: [https://repository.upenn.edu/statistics\\_papers](https://repository.upenn.edu/statistics_papers)

 Part of the [Business Commons](#), and the [Statistics and Probability Commons](#)

---

### Recommended Citation

Fill, J. A., & Steele, J. M. (2005). Exact Expectations of Minimal Spanning Trees for Graphs With Random Edge Weights. *Stein's Method and Applications*, 1-10. Retrieved from [https://repository.upenn.edu/statistics\\_papers/17](https://repository.upenn.edu/statistics_papers/17)

This paper is posted at ScholarlyCommons. [https://repository.upenn.edu/statistics\\_papers/17](https://repository.upenn.edu/statistics_papers/17)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Exact Expectations of Minimal Spanning Trees for Graphs With Random Edge Weights

## Abstract

Two methods are used to compute the expected value of the length of the minimal spanning tree (MST) of a graph whose edges are assigned lengths which are independent and uniformly distributed. The first method yields an exact formula in terms of the Tutte polynomial. As an illustration, the expected length of the MST of the Petersen graph is found to be  $34877/12012 = 2.9035 \dots$ . A second, more elementary, method for computing the expected length of the MST is then derived by conditioning on the length of the shortest edge. Both methods in principle apply to any finite graph. To illustrate the method we compute the expected lengths of the MSTs for complete graphs.

## Disciplines

Business | Statistics and Probability

# EXACT EXPECTATIONS OF MINIMAL SPANNING TREES FOR GRAPHS WITH RANDOM EDGE WEIGHTS

JAMES ALLEN FILL  
AND  
J. MICHAEL STEELE

ABSTRACT. Two methods are used to compute the expected value of the length of the minimal spanning tree (MST) of a graph whose edges are assigned lengths which are independent and uniformly distributed. The first method yields an exact formula in terms of the Tutte polynomial. As an illustration, the expected length of the MST of the Petersen graph is found to be  $34877/12012 = 2.9035\dots$ . A second, more elementary, method for computing the expected length of the MST is then derived by conditioning on the length of the shortest edge. Both methods in principle apply to any finite graph. To illustrate the method we compute the expected lengths of the MSTs for complete graphs.

## 1. INTRODUCTION TO A FORMULA

Given a finite, connected, simple graph  $G$ , we let  $v(G)$  denote the set of vertices and let  $e(G)$  denote the set of edges. For each edge  $e \in e(G)$  we introduce a non-negative random variable  $\xi_e$  which we view as the length of  $e$ , and we assume that the edge lengths  $\{\xi_e : e \in e(G)\}$  are independent with a common distribution  $F$ . If  $\mathcal{S}(G)$  denotes the set of spanning trees of  $G$  and  $\mathbb{I}$  denotes the indicator function, then the random variable

$$L_{\text{MST}}(G) = \min_{T \in \mathcal{S}(G)} \sum_{e \in G} \xi_e \mathbb{I}(e \in T)$$

is the length of the minimal spanning tree of  $G$ . In Steele (2002) a general formula was introduced for the expected value of  $L_{\text{MST}}(G)$ .

**Theorem 1.** *If  $G$  is a finite connected graph and the Tutte polynomial of  $G$  is  $T(G; x, y)$ , then for independent edge lengths that are uniformly distributed on  $[0, 1]$ , one has*

$$(1) \quad \mathbb{E}[L_{\text{MST}}(G)] = \int_0^1 \frac{(1-p)}{p} \frac{T_x(G; 1/p, 1/(1-p))}{T(G; 1/p, 1/(1-p))} dp,$$

where  $T_x(x, y)$  denotes the partial derivative of  $T(x, y)$  with respect to  $x$ .

The next few sections will provide a self-contained proof of this result. The formula is then applied to several concrete examples, including—for novelty's sake—the famous Petersen graph. We then focus on  $K_n$ , the complete graph on  $n$  vertices. In particular, we show how a conditioning argument based on first principles can

---

The research of J. A. Fill was supported by NSF Grant DMS-0104167 and by The Johns Hopkins University's Acheson J. Duncan Fund for the Advancement of Research in Statistics. April 13, 2004.

be used to compute several expectations which were first found from Tutte polynomials.

## 2. MSTs AND CONNECTED COMPONENTS

For any finite graph  $G$  and any subset  $A$  of the edge set  $e(G)$ , we let  $k(G, A)$  denote the number of connected components of the graph with vertex set  $v(G)$  and edge set  $A$ . If each edge  $e \in G$  is assigned length  $\xi_e$ , it is often useful to take  $A$  to be the set of short edges defined by

$$e_t(G) := \{e \in e(G) : \xi_e \leq t\}.$$

For continuously distributed edge lengths, the minimum spanning tree is (almost surely) unique, and the sum

$$N_{\text{MST}}(G, t) = \sum_{e \in \text{MST}(G)} \mathbb{I}(\xi_e \leq t),$$

gives us the number of edges of the MST of  $G$  that are elements of  $e_t(G)$ .

Now, if  $G$  is a connected graph, then by counting the number of elements of  $e_t(G)$  in each connected component of  $(G, e_t(G))$  one finds the formula

$$N_{\text{MST}}(G, t) + k(G, e_t(G)) = n,$$

and this yields a convenient representation for  $L_{\text{MST}}(G)$ . Specifically, we have

$$\begin{aligned} L_{\text{MST}}(G) &= \sum_{e \in G} \xi_e \mathbb{I}(e \in \text{MST}(G)) = \sum_{e \in G} \int_0^1 \mathbb{I}(t < \xi_e, e \in \text{MST}(G)) dt \\ &= \int_0^1 \sum_{e \in G} [\mathbb{I}(e \in \text{MST}(G)) - \mathbb{I}(\xi_e \leq t, e \in \text{MST}(G))] dt \\ &= \int_0^1 [n - 1 - N_{\text{MST}}(G, t)] dt = \int_0^1 [k(G, e_t(G)) - 1] dt, \end{aligned}$$

so we find

$$(2) \quad 1 + L_{\text{MST}}(G) = \int_0^1 k(G, e_t(G)) dt.$$

Versions of this formula go back at least to Avram and Bertsimas (1992), but here we take away a new message. For us the main benefit of the formula (2) is that through  $k(G, e_t(G))$  it suggests how one should be able to calculate  $\mathbb{E}[L_{\text{MST}}(G)]$  with help from the Tutte polynomial of  $G$ .

## 3. THE TUTTE POLYNOMIAL: A REVIEW OF BASIC FACTS

Given a graph  $G$  (which may have loops or parallel edges), the Tutte polynomial  $T(G; x, y)$  may be computed by successively applying the four rules:

- (1) If  $G$  has no edges, then  $T(G; x, y) = 1$ .
- (2) If  $e$  is an edge of  $G$  that is neither a loop nor an isthmus<sup>1</sup>, then

$$T(G; x, y) = T(G'_e; x, y) + T(G''_e; x, y),$$

where  $G'_e$  is the graph  $G$  with the edge  $e$  *deleted* and  $G''_e$  is the graph  $G$  with the edge  $e$  *contracted*.

---

<sup>1</sup>A loop is an edge from a vertex to itself, and an isthmus is an edge whose removal will disconnect the graph.

- (3) If  $e$  is an isthmus, then  $T(G; x, y) = xT(G'_e; x, y)$ .
- (4) If  $e$  is a loop, then  $T(G; x, y) = yT(G''_e; x, y)$ .

SOME INSTRUCTIVE EXAMPLES

Rules (1) and (3) tell us that the Tutte polynomial of  $K_2$  is just  $x$ . In fact, by  $n - 1$  applications of Rule (3) one finds that the Tutte polynomial of any tree with  $n$  vertices is  $x^{n-1}$ . The rules are more interesting when contractions are required, and it is particularly instructive to check that the Tutte polynomial of the triangle graph  $K_3$  is  $x + x^2 + y$ .

For a more complicated example, one can consider the Tutte polynomial of a *bow tie graph*  $G$  which is defined by joining two copies of  $K_3$  at a single vertex. For the bow tie one finds

$$T(G; x, y) = (x + x^2 + y)^2,$$

and this formula has an elegant (and easily proved) generalization. Specifically, if  $G$  and  $H$  are two graphs that have one vertex and no edges in common, then

$$(3) \quad T(G \cup H; x, y) = T(G; x, y)T(H; x, y).$$

As a corollary of this observation we see that the product of Tutte polynomials is always a Tutte polynomial.

THE RANK FUNCTION AND MEASURES OF CONNECTEDNESS

The rank function  $r(\cdot)$  of a graph  $G$  is a function on the subsets of  $e(G)$  which associates to each  $A \subset e(G)$  the value

$$r(A) = |v(G)| - k(G, A),$$

where, as before,  $k(G, A)$  is the number of connected components of the graph with vertex set  $v(G)$  and edge set  $A \subset e(G)$ . The rank function provides a measure of the extent to which the graph  $(v(G), A)$  is *disconnected*, and it also provides an explicit formula for the Tutte polynomial,

$$(4) \quad T(G; x, y) = \sum_{A \subset e(G)} (x - 1)^{r(e(G)) - r(A)} (y - 1)^{|A| - r(A)}.$$

One simple consequence of this formula is that  $T(G; 2, 2) = 2^{|e(G)|}$ , a fact that is sometimes used to check a Tutte polynomial that has been computed by hand.

4. CONNECTION TO THE PROBABILITY MODEL

For a connected graph  $G$ , one has

$$r(A) = |v(G)| - k(G, A) = n - k(G, A) \quad \text{and} \quad r(e(G)) = n - 1,$$

so if we set  $m = |e(G)|$  then the sum formula (4) may be written as

$$\begin{aligned} T(G; x, y) &= \frac{1}{(x - 1)(y - 1)^n} \sum_{A \subset e(G)} (y - 1)^{|A|} [(x - 1)(y - 1)]^{k(G, A)} \\ &= \frac{y^m}{(x - 1)(y - 1)^n} \sum_{A \subset e(G)} \left(\frac{y - 1}{y}\right)^{|A|} \left(\frac{1}{y}\right)^{m - |A|} [(x - 1)(y - 1)]^{k(G, A)}. \end{aligned}$$

If one now sets

$$(5) \quad p = \frac{y - 1}{y} \quad \text{and} \quad 1 - p = \frac{1}{y},$$

one gets a sum of the form

$$(6) \quad \sum_{A \subseteq e(G)} p^{|A|} (1-p)^{m-|A|} f(A).$$

Any such sum has an obvious interpretation as a mathematical expectation, and such reinterpretations of the Tutte polynomial are often used in statistical physics. More recently they have been used in studies of the computational complexity of the Tutte polynomial (see, for example, Welsh (1999) or Lemma 1 of Alon, Frieze, and Welsh (1994)).

## 5. A MOMENT GENERATING FUNCTION AND A MOMENT

The reformulation in formula (6) also turns out to give an almost automatic path to a formula for  $\mathbb{E}[L_{\text{MST}}(G)]$ . The first factor in the sum is equal to the probability (under the uniform model) that one has  $\xi_e \leq p$  for exactly those edges in the set  $A$ , so if one takes

$$A = e_p(G) = \{e : e \in e(G), \xi_e \leq p\}$$

then the moment generating function

$$\varphi(t) \equiv \varphi_p(t) := \mathbb{E}[\exp\{t k(G, e_p(G))\}]$$

can be written in terms of  $T(G; x, y)$ . Specifically one has

$$(7) \quad \varphi(t) = p^{n-1} (1-p)^{m-n+1} e^t T\left(G; 1 + e^t \frac{1-p}{p}, \frac{1}{1-p}\right),$$

and this formula tells almost everything there is to say about the distribution of  $k(G, e_p(G))$ . In particular, it yields a quick way to calculate  $\mathbb{E}[k(G, e_p(G))]$ .

If we retain the abbreviations (5), then (taking  $1 + [e^t(1-p)/p]$  for  $x$ ) we have

$$\varphi'(t) = \varphi(t) \left\{ 1 + e^t \frac{1-p}{p} \frac{T_x(G; x, y)}{T(G; x, y)} \right\},$$

When we let  $t = 0$ , we find for  $x = 1/p$  and  $y = 1/(1-p)$  that

$$(8) \quad \mathbb{E}[k(G, e_p(G))] = 1 + \frac{1-p}{p} \frac{T_x(G; x, y)}{T(G; x, y)}.$$

Finally, from the definitions of  $x$  and  $y$  together with the representation (2) for  $L_{\text{MST}}(G)$  in as an integral of  $k(G, e_p(G))$ , we find

$$(9) \quad \mathbb{E}[L_{\text{MST}}(G)] = \int_0^1 \frac{(1-p)}{p} \frac{T_x(G; 1/p, 1/(1-p))}{T(G; 1/p, 1/(1-p))} dp,$$

and the proof of Theorem 1 is complete.

## 6. SIMPLE—AND NOT-SO-SIMPLE—EXAMPLES

As noted in Steele (2002), there are some easy examples that help to familiarize the formula (9). In particular, if  $G$  is a tree with  $n$  vertices, then  $T(G; x, y) = x^{n-1}$  and the integral works out to be  $(n-1)/2$ , which is obviously the correct value of  $\mathbb{E}[L_{\text{MST}}(G)]$ . This fact may also be seen as a corollary of the product rule (3); if graphs  $G$  and  $H$  share only one a common vertex and no edges, then the graph  $G \cup H$  has Tutte polynomial  $T(G; x, y)T(H; x, y)$  and formula (9) tells us that

$$\mathbb{E}[L_{\text{MST}}(G \cup H)] = \mathbb{E}[L_{\text{MST}}(G)] + \mathbb{E}[L_{\text{MST}}(H)],$$

a fact which can also be seen directly from the definition of the MST.

For the complete graph on three vertices we have already seen that one has  $T(K_3) = x + x^2 + y$ , and the integral (9) yields  $\mathbb{E}[L_{\text{MST}}(K_3)] = 3/4$ , which one can also check independently. Also, for  $K_4$  the Tutte polynomial can be found by hand to be

$$T(K_4; x, y) = 2x + 2y + 3x^2 + 3y^2 + 4xy + x^3 + y^3,$$

and when this polynomial is used in the integral formula (9), we find

$$\mathbb{E}[L_{\text{MST}}(K_4)] = \frac{31}{35}.$$

A still more compelling example than  $K_4$  is given by the famous Petersen graph that is illustrated by Figure 1.

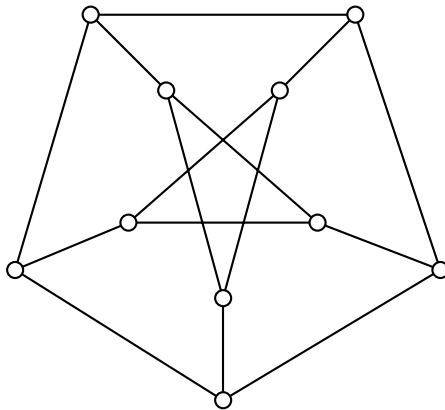


FIGURE 1. The Petersen graph can be used to illustrate several graph-theoretic concepts. In particular, it is non-Hamiltonian and nonplanar, yet it can be embedded in the projective plane.

The Tutte polynomial  $T(G; x, y)$  of the Petersen graph is not easily found by hand, but the **Maple** program *tuttepoly* reveals that it is given by

$$\begin{aligned} &36x + 120x^2 + 180x^3 + 170x^4 + 114x^5 + 56x^6 + 21x^7 + 6x^8 + x^9 + \\ &36y + 84y^2 + 75y^3 + 35y^4 + 9y^5 + y^6 + 168xy + 240x^2y + 170x^3y + \\ &70x^4y + 12x^5y + 171xy^2 + 105x^2y^2 + 30x^3y^2 + 65xy^3 + 15x^2y^3 + 10xy^4. \end{aligned}$$

From the formula (1) for the expectation, symbolic calculation then gives

$$(10) \quad \mathbb{E}[L_{\text{MST}}(\text{Petersen})] = \frac{34877}{12012} = 2.90351\dots$$

### 7. COMPLETE GRAPHS AND TWO PROBLEMS

For the complete graph  $K_n$  the Tutte polynomial  $T(K_n; x, y)$  is known for moderate values of  $n$ , but the complexity of the polynomials grows rapidly with  $n$ . The polynomials for  $n = 2, 3, \dots, 8$  are given by Gessel and Sagan (1996) and Gessel (personal communication) has extended the computation to  $n = 15$ . These polynomials and formula (1) can be used to compute  $\mathbb{E}[L_{\text{MST}}(K_n)]$ .

$n$	$\mathbb{E}[L_{\text{MST}}(K_n)]$	Numerical Value	Forward Difference
2	1/2	0.50000	0.250000
3	3/4	0.75000	0.135714
4	31/35	0.88571	0.080736
5	893/924	0.96645	0.051864
6	278/273	1.01832	0.035401
7	30739/29172	1.05372	0.025343
8	199462271/184848378	1.07906	0.018844
9	126510063932/115228853025	1.09790	—————

From the tabled values it is natural to conjecture that  $\mathbb{E}[L_{\text{MST}}(K_n)]$  forms a monotone increasing concave sequence. One might hope to prove this result with help from the integral formula (1) and the known properties of the Tutte polynomials for  $K_n$ , but this does not seem to be easy. This problem was raised at the conference *Mathematics and Computer Science II* at Versailles in 2002, but no progress has been made.

The tabled values may also tell us something about a remarkable result of Frieze (1985) which asserts that

$$\lim_{n \rightarrow \infty} \mathbb{E}[L_{\text{MST}}(K_n)] = \zeta(3) = 1.202\dots$$

The rate of convergence in Frieze's limit is not known, but the declining forward differences in this table for  $\mathbb{E}[L_{\text{MST}}(K_n)]$  suggest that the rate of convergence in Frieze's theorem may be rather slow.

Unfortunately, the computational complexity of the Tutte polynomial makes it unlikely that any of these problems will be resolved by the integral formula (1). As a consequence, it seems useful to consider alternative approaches to the calculation of  $\mathbb{E}[L_{\text{MST}}(G)]$  which are closer to first principles.

## 8. A RECURSION AND ITS APPLICATIONS

If the random variables  $U_1, U_2, \dots, U_n$  are independent and uniformly distributed on  $[0, 1]$  and if  $E$  denotes the event  $\{U_n = \min(U_1, U_2, \dots, U_n) \text{ and } U_n = y\}$ , then conditional on  $E$  the random variables  $U_1, U_2, \dots, U_{n-1}$  are independent and uniformly distributed on  $[y, 1]$ . This elementary observation suggests a recursive approach to the computation of  $\mathbb{E}[L_{\text{MST}}(G)]$ .

The recursion is perhaps most easily understood by first considering  $K_3$ . For the moment, let  $D$  denote the graph with two vertices  $v_1$  and  $v_2$  and two parallel edges between  $v_1$  and  $v_2$ . Under the uniform model  $U[0, 1]$  for edge lengths we obviously have

$$\mathbb{E}^{U[0,1]}[L_{\text{MST}}(D)] = \mathbb{E}^{U[0,1]}[\min(U_1, U_2)] = 1/3.$$

On the other hand, if the edge lengths of  $D$  are chosen uniformly in the interval  $[y, 1]$ , then we have

$$\mathbb{E}^{U[y,1]}[L_{\text{MST}}(D)] = y + (1 - y)\mathbb{E}^{U[0,1]}[L_{\text{MST}}(D)] = (1 + 2y)/3.$$

Since the shortest edge in a graph is always in the minimal spanning tree, these observations suggest that if we condition on the length of the shortest edge, then we might well expect to find a useful recurrence relation.

If  $Y$  denotes the length of the shortest edge in  $K_3$  under the uniform model, then the density of  $Y$  on  $[0, 1]$  is just  $3(1 - y)^2$ . Also, after we take the shortest edge



$(u, v)$  of  $K_3$  as an element of our MST, the rest of the cost of the MST of  $K_3$  is simply the cost of the MST of the graph obtained from  $K_3$  by identifying  $u$  and  $v$  and removing the resulting loop. This observation gives us the identity

$$(11) \quad \begin{aligned} \mathbb{E}[L_{\text{MST}}(K_3)] &= \int_0^1 \mathbb{E}[L_{\text{MST}}(K_3)|Y = y] \{3(1-y)^2\} dy \\ &= \int_0^1 \left\{ y + \mathbb{E}^{U[y,1]}[L_{\text{MST}}(D)] \right\} 3(1-y)^2 dy = \frac{3}{4}, \end{aligned}$$

which fortunately agrees with the value given twice before. To apply this idea more broadly one just needs to introduce appropriate notation.

#### NOTATION AND A GENERAL RECURSION

If  $\mathbf{A} = (a_{ij})_{n \times n}$  is a symmetric  $n$ -by- $n$  matrix with  $a_{ij} \in \{0, 1, \dots\}$  and all diagonal entries equal to zero, then we let  $G(\mathbf{A})$  denote the loopless graph with vertex set  $\{v_1, v_2, \dots, v_n\}$  such that for each  $1 \leq i < j \leq n$  there are  $a_{ij}$  parallel edges between  $v_i$  and  $v_j$ . Also, for each  $1 \leq i < j \leq n$  such that  $a_{ij} \geq 1$ , we let  $G(\mathbf{A}^{(ij)})$  denote the graph with  $n-1$  vertices which is derived from  $G(\mathbf{A})$  by first identifying vertices  $i$  and  $j$  and then removing the resulting  $a_{ij}$  loops. If  $G(\mathbf{A})$  is connected, then the length of the MST is well-defined for both  $G(\mathbf{A})$  and  $G(\mathbf{A}^{(ij)})$ , and the quantities

$$\phi(\mathbf{A}) = \mathbb{E}^{U[0,1]}[L_{\text{MST}}(G(\mathbf{A}))] \quad \text{and} \quad \phi(\mathbf{A}^{(ij)}) = \mathbb{E}^{U[0,1]}[L_{\text{MST}}(G(\mathbf{A}^{(ij)}))],$$

are connected by a simple linear identity.

**Theorem 2.** *For all  $n \geq 2$ , one has the recursion*

$$(12) \quad \phi(\mathbf{A}) = \frac{n-1 + \sum_{1 \leq i < j \leq n} a_{ij} \phi(\mathbf{A}^{(ij)})}{1 + \sum_{1 \leq i < j \leq n} a_{ij}}.$$

*Proof.* To follow the pattern that we used for  $K_3$ , we first let  $Y$  denote the length of the shortest edge in  $G(\mathbf{A})$ . We then let  $m = \sum_{i,j} a_{ij} \geq 1$  denote the number of edges in  $G(\mathbf{A})$  and condition on both the value of  $Y$  and the event that it is edge  $e$  which has length  $Y$ . This gives us the integral formula

$$\phi(\mathbf{A}) = \frac{1}{m} \sum_e \int_0^1 \mathbb{E}^{U[0,1]}[L_{\text{MST}}(G(\mathbf{A}))|Y = y, \xi_e = y] m(1-y)^{m-1} dy.$$

Now, if  $e$  is one of the  $a_{ij}$  edges between  $v_i$  and  $v_j$ , then we have

$$\begin{aligned} \mathbb{E}^{U[0,1]}[L_{\text{MST}}(G(\mathbf{A}))|Y = y, \xi_e = y] &= y + \mathbb{E}^{U[y,1]}[L_{\text{MST}}(G(\mathbf{A}^{(ij)}))] \\ &= y + \{(n-2)y + (1-y)\phi(\mathbf{A}^{(ij)})\}. \end{aligned}$$

Thus, after working out the integral, we find that  $\phi(\mathbf{A})$  is equal to

$$\frac{1}{m} \sum_{1 \leq i < j \leq n} a_{ij} \left\{ \frac{n-1}{m+1} + \frac{m}{m+1} \phi(\mathbf{A}^{(ij)}) \right\} = \frac{n-1 + \sum_{1 \leq i < j \leq n} a_{ij} \phi(\mathbf{A}^{(ij)})}{1 + \sum_{1 \leq i < j \leq n} a_{ij}}.$$

□

#### A PRELIMINARY EXAMPLE

It is obvious from first principles that for a tree  $T_n$  with  $n$  vertices one has  $\mathbb{E}^{U[0,1]}[L_{\text{MST}}(T_n)] = (n-1)/2$ , but it is instructive to see how the recursion (12)

recovers this fact. If  $G(\mathbf{A}) = T_n$ , then for each  $1 \leq i < j \leq n$  such that  $a_{ij} = 1$  the graph  $G(\mathbf{A}^{(ij)})$  is tree with  $n - 1$  vertices. By the recursion (12) and induction, one then sees that  $\mathbb{E}^{U[0,1]}[L_{\text{MST}}(T_n)]$  depends only on  $n$ —and not on the particular structure of the tree  $T_n$ . The recursion (12) therefore asserts

$$\mathbb{E}^{U[0,1]}[L_{\text{MST}}(T_n)] = \frac{(n-1)(1 + \mathbb{E}^{U[0,1]}[L_{\text{MST}}(T_{n-1})])}{n},$$

and this leads us directly to  $\mathbb{E}^{U[0,1]}[L_{\text{MST}}(T_n)] = (n-1)/2$ .

#### MORE REVEALING EXAMPLES

The analysis of the complete graphs calls on more of the machinery of the recursion (12) for  $\phi(\mathbf{A})$ , but the story begins simply enough. For  $n = 2$  the strict upper triangle of  $\mathbf{A}$  has just the single entry  $a \equiv a_{12}$ , and in this special case write  $\phi(a)$  instead of  $\phi(\mathbf{A})$ . From first principles it is immediate that  $\phi(a) = 1/(1+a)$ , and as a special case we see  $\mathbb{E}^{U[0,1]}[L_{\text{MST}}(K_2)] = \phi(1) = 1/2$ .

The triangle graph  $K_3$  is more interesting. Here the associated matrix is

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \text{ which we abbreviate by its upper triangle } \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}.$$

From the recursion (12) we then find

$$\mathbb{E}^{U[0,1]}[L_{\text{MST}}(K_3)] = \phi \left( \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix} \right) = [2 + 3\phi(2)]/4 = 3/4,$$

a result that is already quite familiar.

The analogous calculation for  $K_4$  is only a little more complicated. One application of the recursion (12) gives

$$\mathbb{E}^{U[0,1]}[L_{\text{MST}}(K_4)] = \phi \left( \begin{pmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{pmatrix} \right) = \left[ 3 + 6\phi \left( \begin{pmatrix} 2 & 2 \\ & 1 \end{pmatrix} \right) \right] / 7,$$

and a second application gives

$$\phi \left( \begin{pmatrix} 2 & 2 \\ & 1 \end{pmatrix} \right) = [2 + 4\phi(3) + \phi(4)]/6 = 8/15,$$

so in the end we find  $\mathbb{E}^{U[0,1]}[L_{\text{MST}}(K_4)] = 31/35$ , which fortunately agrees with the table of Section 7.

For  $K_5$  the computation is not so brief, but the pattern is familiar. To begin we have

$$\mathbb{E}^{U[0,1]}[L_{\text{MST}}(K_5)] = \phi \left( \begin{pmatrix} 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 \\ & & 1 & 1 \\ & & & 1 \end{pmatrix} \right),$$

and one application of the recursion (12) brings us to

$$\mathbb{E}^{U[0,1]}[L_{\text{MST}}(K_5)] = \frac{1}{11} \left[ 4 + 10\phi \left( \begin{pmatrix} 2 & 2 & 2 \\ & 1 & 1 \\ & & 1 \end{pmatrix} \right) \right].$$

The second application now gives

$$\phi \begin{pmatrix} 2 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix} = \frac{1}{10} \left[ 3 + 6\phi \begin{pmatrix} 3 & 3 \\ 1 & 1 \end{pmatrix} + 3\phi \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \right],$$

and we also have

$$\phi \begin{pmatrix} 3 & 3 \\ 1 & 1 \end{pmatrix} = \frac{2 + 6\phi(4) + \phi(6)}{8} = \frac{2 + (6/5) + (1/7)}{8} = \frac{117}{280}$$

and

$$\phi \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} = \frac{2 + 4\phi(4) + 4\phi(6)}{9} = \frac{2 + (4/5) + (4/7)}{9} = \frac{118}{315},$$

from which we find

$$\phi \begin{pmatrix} 2 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix} = \frac{3 + (351/140) + (118/105)}{10} = \frac{557}{840}.$$

After feeding these intermediate quantities into our first identity, we find at last

$$\mathbb{E}^{U[0,1]}[L_{\text{MST}}(K_5)] = \frac{4 + (557/84)}{11} = \frac{893}{924}.$$

## 9. CONCLUDING REMARKS

Neither the elementary recursion (12) nor the integral formula (1) seem to provide one with an easy path to the qualitative features of  $\mathbb{E}[L_{\text{MST}}(K_n)]$ , though they do provide tantalizing hints. The sequence  $\mathbb{E}[L_{\text{MST}}(K_n)]$  is quite likely monotone and concave, but a proof of these properties would seem to call for a much clearer understanding of either the Tutte polynomials  $T(K_n; x, y)$  or the intermediate quantities  $\phi(\mathbf{A}) = \mathbb{E}^{U[0,1]}[L_{\text{MST}}(G(\mathbf{A}))]$  associated with the recursion (12).

ACKNOWLEDGEMENTS: Ira Gessel kindly provided an extension to  $n = 15$  of the table of Tutte polynomials for  $K_n$  given in Gessel and Sagan (1996). The Tutte polynomial for the Petersen graph was obtained from the web page of S.C. Locke at <http://www.math.fau.edu/locke/TPol.htm>.

## REFERENCES

- [1] N. Alon, A. Frieze, and D. Welsh (1994) *Polynomial time randomized approximation schemes for the Tutte polynomial of dense graphs*, pp. 24–35 in the Proceedings of the 35th Annual Symposium on the Foundations of Computer Science (S. Goldwasser, ed.), IEEE Computer Society Press, 1994.
- [2] F. Avram and D. Bertsimas (1992) *The minimum spanning tree constant in geometric probability and under the independent model: a unified approach*, *Annals of Applied Probability* **2**, 113–130.
- [3] A. M. Frieze (1985) *On the value of a random minimum spanning tree problem*, *Discrete Appl. Math.* **10**, 47–56.
- [4] I. M. Gessel and B. E. Sagan (1996) *The Tutte polynomial of a graph, depth-first search, and simplicial complex partitions*, *Electronic Journal of Combinatorics* **3**, Research Paper 9, (36 pp.).
- [5] J. M. Steele (2002) *Minimal Spanning Trees for Graphs with Random Edge Lengths*, in *Mathematics and Computer Science II: Algorithms, Trees, Combinatorics and Probabilities*, B. Chauvin, Ph. Flajolet, D. Gardy, and A. Mokkadem (eds.), Birkhäuser, Boston, 2002.
- [6] D. Welsh (1999) *The Tutte polynomial*, *Random Structures and Algorithms* **15**, 210–228.

JAMES ALLEN FILL, DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS, THE JOHNS HOPKINS UNIVERSITY, BALTIMORE MD 21218-2682, [jimfill@jhu.edu](mailto:jimfill@jhu.edu)

J. MICHAEL STEELE, DEPARTMENT OF STATISTICS, WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA PA 19104, [steele@wharton.upenn.edu](mailto:steele@wharton.upenn.edu)