The lower tail of the random minimum spanning tree

Abraham D. Flaxman
abie@microsoft.com

September 22, 2006

Technical Report
MSR-TR-2006-132

Consider a complete graph K_n where the edges have costs given by independent random variables, each distributed uniformly between 0 and 1. The cost of the minimum spanning tree in this graph is a random variable which has been the subject of much study. This note considers the large deviation probability of this random variable. Previous work has shown that the log-probability of deviation by ε is $-\Omega(n)$, and that log-probability of Z exceeding $\zeta(3)$ this bound is correct; $\log \Pr[Z \geq \zeta(3) + \varepsilon] = -\Theta(n)$. The purpose of this note is to provide a simple proof that the scaling of the lower tail is also linear, $\log \Pr[Z \leq \zeta(3) - \varepsilon] = -\Theta(n)$.

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
http://www.research.microsoft.com

1 Introduction

If the edge costs of the complete graph K_n are independent random variables, each uniformly distributed between 0 and 1, then the cost of a minimum spanning tree is a random variable which has expectation asymptotically equal to $\zeta(3) = \sum_{i=1}^{\infty} i^{-3}$ [3]. Furthermore, after an appropriate rescaling, this random variable converges in distribution to a Gaussian distribution with an explicitly known variance of about 1.6857 [5]. This note considers the large deviation probability of this random variable, denoted Z_n.

In [6], as an example application of Talagrand's Inequality, it is shown that Z_n satisfies an exponential tail inequality of the form

$$\Pr[|Z_n - \zeta(3)| \geq \varepsilon] \leq e^{-C_\varepsilon n}.$$

(See also [2] for an alternative approach with additional details). Simple considerations show that for the log-probability of Z_n exceeding $\zeta(3)$ this bound is correct, which is to say that $\log \Pr[Z_n \geq \zeta(3) + \varepsilon] = -\Theta(n)$. For example, the probability that every edge incident to vertex 1 has cost at least 1/2 is $(1/2)^{n-1}$, and conditioned on this event, **whp** $Z_n = (1 + o(1))(\zeta(3) + 1/2)$.

The behavior of the lower tail is not as simple to identify. A casual inspection may lead to the conjecture that the lower tail is even more tightly concentrated than the upper tail. The previous paragraph described how an overly large value of Z_n can be "blamed" on a single vertex which has only expensive edges. However, for a single vertex to be similarly responsible for the cost of the tree being significantly lower than expected, it needs to have a lot of edges with cost less than $\zeta(3)/n$. This occurs with log-probability of $-\Theta(n \log n)$.

The purpose of this note is to show that the lower tail of Z_n is at least e^{-Cn} for any constant deviation less than $\zeta(3)$. (Note that, for example, $\Pr[Z_n \leq \zeta(3) - (\zeta(3) - n^{-10})]$, is not at least e^{-Cn}.)

Theorem 1 *Let the random variable Z_n be the cost of the minimum spanning tree when the edges of the complete graph K_n have costs selected independently and uniformly at random in the interval $[0, 1]$. Then, for any $\varepsilon \in (0, 1)$, there exists a constant δ, such that for all sufficiently large n,*

$$\Pr[Z_n \leq (1 - \varepsilon)\zeta(3)] \geq e^{-\delta n}.$$

Though this scaling behavior is not terribly surprising, it does rule out the possibility of a surprise. This is in contrast with, for example, the surprising result on the concentration of the eigenvalues in random matrix due to Alon, Krivelevich, and Vu [1]. That paper considers how tightly an eigenvalue of a random matrix is concentrated around its mean, and shows that, for example, the log-probability of deviation of the first eigenvalue of the adjacency matrix of $\mathbb{G}_{n,1/2}$ of scales like $-\Omega(n^2)$.

2 Lower bound

The argument establishing a lower bound is based on the observation that if the weights on the edges are independent and given by the minimum of 2 random variabless selected uniformly at random from $[0, 1]$ then the expected cost is $\zeta(3)/2$ (this is proved by Steele in [7] and extended by Frieze and McDiarmid in [4]; in fact, the only feature of the edge weight distribution that is important to the expected value of Z_n is the behavior of the density function at 0.)

To make use of this observation, consider the following complicated way to generate Z_n: Look first at a larger probability space, where each edge has 2 values, X_e^+ and X_e^-, and each vertex has a polarity chosen uniformly at random, $\Phi(v) \in \pm 1$. Then, to obtain Z_n, consider the graph where edge $e = \{u, v\}$ has weight $Y_e = X_e^{\Phi(u)\Phi(v)}$.

Edge weights generated in this manner are identically distributed with the original model, and so the cost of the minimium spanning tree is distributed identically with Z_n. But with this generative procedure it is easy to obtain a lower bound on the log-probability of the event $\{Z_n \leq 3\zeta(3)/4\}$. Consider the minimum spanning tree in the graph where edge e has weight $\min\{X_e^+, X_e^-\}$. Since this is a tree, there is a function ψ which assigns every vertex a polarity so that $X_e^{\psi(u)\psi(v)}$ is the minimum of the 2 values. (To see this, designate some vertex the root, and start by arbitrarially assigning a polarity to the root, and then assigning the polarity of additional vertices in the order given by a breadth-first search of the minimum spanning tree.) If this function is the one that comes up, then the expected cost of Z_n is asymptotic to $\zeta(3)/2$, and, by Markov's inequality, $\Pr[Z_n \geq 3/2(\zeta(3)/2) \mid \Phi = \psi] \leq 2/3$. The event $\{\Phi = \psi\}$ has the same probability as the event that Φ equals any other polarity function, so unconditionally, $\Pr[Z_n \leq 3\zeta(3)/4] \geq (1/3)2^{-n}$.

For values of $\varepsilon > 1/4$, repeat this argument but with the larger probability space containing k different weights for each edge, and $\Phi(v)$ chosen uniformly from k complex roots of unity. Again, considering as a weight the minimum of the k weights on each edge leads to the expected value $\zeta(3)/k$, and probability that this random variable exceeds $2\zeta(3)/k$ is at most $1/2$. Since there is again a function ψ that results in selecting the minimum value for each edge in the minimum spanning tree, an upper-bound on the unconditional probability is

$$\Pr[Z_n \leq 2\zeta(3)/k] \geq (1/2)k^{-n}.$$

Note that this argument also works when k is a function of n, showing that

$$\log \Pr[Z_n = \mathcal{O}(1/k)] = -\Omega(n \log k).$$

References

[1] Alon, N., Krivelevich, M., and Vu, V. H. On the concentration of eigenvalues of random symmetric matrices. *Israel J. Math. 131* (2002), 259–267.

[2] Flaxman, A. D., Frieze, A., and Krivelevich, M. On the random 2-stage minimum spanning tree. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms* (Philadelphia, PA, USA, 2005), Society for Industrial and Applied Mathematics, pp. 919–926.

[3] Frieze, A. M. On the value of a random minimum spanning tree problem. *Discrete Appl. Math. 10*, 1 (1985), 47–56.

[4] Frieze, A. M., and McDiarmid, C. J. H. On random minimum length spanning trees. *Combinatorica 9*, 4 (1989), 363–374.

[5] Janson, S. The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. *Random Structures Algorithms 7*, 4 (1995), 337–355.

[6] McDiarmid, C. On the method of bounded differences. In *London Mathematical Society Lecture Note Series*, vol. 141. Cambridge University Press, 1989, pp. 148–188.

[7] Steele, J. M. On Frieze's $\zeta(3)$ limit for lengths of minimal spanning trees. *Discrete Appl. Math. 18*, 1 (1987), 99–103.