Successive minimum spanning trees

SVANTE JANSON
(joint work with Gregory Sorkin)

Consider the complete graph K_n with edge costs that are i.i.d. random variables, with a uniform distribution $U(0,1)$ (or, alternatively, an exponential distribution $\mathrm{Exp}(1)$). A well-known problem is to find the minimum (cost) spanning tree T_1, and its cost $c(T_1)$. A famous result by Frieze [2] shows that as $n \to \infty$, $c(T_1)$ converges in probability to $\zeta(3)$. (In both the uniform and exponential cases.)

Suppose now that we want a second spanning tree T_2, edge-disjoint from the first, and that we select it in a greedy fashion by first finding the minimum spanning tree T_1, and then the minimum spanning tree T_2 using only the remaining edges. (I.e., the minimum spanning tree in $K_n \setminus T_1$, meaning the graph with edge set $E(K_n) \setminus E(T_1)$.) We then continue and define T_3 as the minimum spanning tree in $K_n \setminus (T_1 \cup T_2)$, and so on. We show that the costs $c(T_2)$, $c(T_3)$, ... also converge in probability to some constants.

Theorem 1. *For each $k \geq 1$, there exists a constant γ_k such that, as $n \to \infty$, $c(T_k) \xrightarrow{\text{p}} \gamma_k$ (for both uniform and exponential cost distributions).*

The result extends easily to other distributions of the edge costs, by standard arguments, but we consider in here only the uniform and exponential cases.

By Frieze [2], $\gamma_1 = \zeta(3)$. The constants γ_k for larger k are given by some expressions in the proof, but not in a form that is easily evaluated since they involve solutions of some non-linear functional equations (which furthermore involve a parameter). We can show the following bounds, which imply that γ_k is roughly $2k$ for large k:

$$k^2 \leq \sum_{i=1}^{k} \gamma_i \leq k^2 + k, \qquad k \geq 1 \tag{1}$$

and

$$2k - 2k^{1/2} < \gamma_k < 2k + 2k^{1/2}, \qquad k \geq 1. \tag{2}$$

A minor technical problem is that T_2 (and T_3, ...) does not always exist; it may happen that T_1 is a star and then $K_n \setminus T_1$ is disconnected. This happens only with a small probability, and w.h.p. (with high probability, i.e., with probability $1-o(1)$ as $n \to \infty$), T_k is defined for every fixed k. However, we avoid this problem completely by modifying the model: we assume that we have a multigraph with an infinite number of copies of each edge in K_n, and that these have the costs given by the points in a Poisson process with intensity 1 on $[0, \infty)$. (The Poisson processes for different edges are, of course, independent.) Note that when finding T_1, we only care about the cheapest copy of each edge, and its cost has an $\mathrm{Exp}(1)$ distribution, so the problem for T_1 is the same as the original one. However, we now never run out of edges and we can define T_k for all integers $k = 1, 2, 3, \ldots$. Asymptotically, the three models are equivalent, and Theorem 1 holds for any of the models.

The multigraph model, moreover, is useful in our proofs because of the added independence.

Frieze [2] also proved that the expectation $\mathbb{E}\, c(T_1)$ converges to $\zeta(3)$. For the multigraph model just described, this too extends.

Theorem 2. *For the multigraph model,* $\mathbb{E}\, c(T_k) \to \gamma_k$ *for each* $k \geq 1$ *as* $n \to \infty$.

Remark 3. However, for the simple graph K_n with, say, exponential costs, there is as said above a small but positive probability that T_k does not exist for $k \geq 2$. Hence, either $\mathbb{E}\, c(T_k)$ is undefined for $k \geq 2$, or (better) we define $c(T_k) = \infty$ when T_k does not exist, and then $\mathbb{E}\, c(T_k) = \infty$ for $k \geq 2$ and every n. Hence Theorem 2 does not hold for simple graphs, and the multigraph model is essential for studying the expectation.

Remark 4. Frieze and Johansson [3] recently considered a related problem, where instead of choosing spanning trees $T_1, T_2, \dots$ greedily one by one, they choose k edge-disjoint spanning trees with minimum total cost. It is easy to see, by small examples, that selecting k spanning trees greedily one by one does not always give a set of k edge-disjoint spanning trees with minimum cost, so the problems are different. We can also show that, at least for $k = 2$, the two problems also asymptotically have different answers, in the sense that the limiting values of the minimum cost (which exist for both problems) are different.

The proofs are, as in many other previous papers on the random minimum spanning tree problem, based on *Kruskal's algorithm* which processes the edges in order of increasing cost and keeps the ones that join two different components in the forest obtained so far. (I.e., it keeps the edges that do not form a cycle together with some previously chosen edges.) The second minimum spanning tree can then be found by another application of the same algorithm to the remaining edges, and so on.

The results are proved by considering a random (multi)graph process, where copies of each edge ij arrive as a Poisson process with intensity $1/n$; an edge arriving at time t has cost t/n. We let $G_1(t)$ be the multigraph formed by the edges that have arrived at time t. We run Kruskal's algorithm and let $F_1(t)$ be the forest formed by the edges selected up to time t for the minimum spanning tree T_1. We let $G_2(t)$ be the multigraph consisting of the edges in $G_1(t) \setminus F_1(t)$, and let $F_2(t)$ be the forest formed by the edges selected up to time t by Kruskal's algorithm applied to $G_2(t)$, and so on. We show, by induction in k, that each $G_k(t)$ is an example of an inhomogeneous random graph of the type studied in [1]; results from [1] thus yield results on the (asymptotic) structure of $G_k(t)$, in particular on the existence and size of a giant component, and these structural results are used to show the theorems above on the cost $c(T_k)$.

References

[1] Béla Bollobás, Svante Janson and Oliver Riordan, The phase transition in inhomogeneous random graphs. *Random Struct. Alg.* **31** (2007), 3–122.

[2] Alan M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Applied Mathematics* **10** (1985), 47–56.

[3] Alan Frieze and Tony Johansson, On edge disjoint spanning trees in a randomly weighted complete graph. Preprint, 2015. `arXiv:1505.03429`