The scaling limit of the minimum spanning tree of the complete graph

Christina Goldschmidt

(joint work with L. Addario-Berry, N. Broutin, G. Miermont)

The minimum spanning tree. Consider the complete graph, K_n, on vertices labelled by $\{1, 2, \ldots, n\}$. Put independent random weights on the edges which are uniformly distributed on $[0, 1]$ and find the spanning tree M_n of smallest total weight; this is the so-called *minimum spanning tree (MST)*. Now think of M_n as a metric space by taking the metric to be the graph distance divided by $n^{1/3}$. We also endow M_n with a probability measure by placing mass $1/n$ on each vertex. Our main result is the following theorem.

Theorem 1. *There exists a random compact measured metric space $\mathcal{M}$ such that, as $n \to \infty$,*

$$M_n \to \mathcal{M}$$

in distribution.

The convergence here is in the sense of the Gromov–Hausdorff–Prokhorov distance, which we now define.

The Gromov–Hausdorff–Prokhorov distance. Let (X, d, μ) and (X', d', μ') be measured metric spaces. A *correspondence* between X and X' is defined to be a measurable subset R of $X \times X'$ such that for every $x \in X$ there exists $x' \in X'$ such that $(x, x') \in R$ and vice versa. A *partial coupling* of μ and μ' is a finite Borel measure on $X \times X'$ such that $p_*\pi \leq \mu$ and $p'_*\pi \leq \mu'$, where $p : X \times X' \to X$ and $p' : X \times X' \to X'$ are the canonical projections. The *distortion* of R is

$$\sup\{|d(x, y) - d'(x', y')| : (x, x') \in R, (y, y') \in R\}$$

and the *discrepancy* of π is

$$\max\{(\mu - p_*\pi)(X), (\mu' - p'_*\pi)(X')\}.$$

The Gromov–Hausdorff–Prokhorov distance, $\mathrm{d_{GHP}}((X, d, \mu), (X', d', \mu'))$, between (X, d, μ) and (X', d', μ') is then the infimum of the values $\epsilon > 0$ such that there exist a correspondence R and a partial coupling π such that the distortion of R and the discrepancy of π are both strictly smaller than ϵ and, furthermore, $\pi(R^c) < \epsilon$.

Write $\mathscr{M}$ for the set of measured isometry-equivalence classes of compact separable measured metric spaces. Then $(\mathscr{M}, \mathrm{d_{GHP}})$ is a complete separable metric space.

Scaling limits of random discrete trees. Theorem 1 is very much in the spirit of recent work on the scaling limits of a wide variety of random discrete trees, which began with the seminal work of Aldous [4, 5, 6] in the early 1990's. The prototypical result [4, 8] is that the uniform random tree on vertices labelled by $\{1, 2, \ldots, n\}$, considered as a measured metric space, converges in distribution to the *Brownian continuum random tree (CRT)*. The Brownian CRT is an example of a *random* $\mathbb{R}$*-tree* (see Le Gall [9]).

Kruskal's algorithm and the Erdős–Rényi random graph. The proof of Theorem 1 relies on *Kruskal's algorithm* for building the MST.

- Start from a forest of isolated vertices. List the edges of the graph as $e_1, e_2, \ldots, e_{\binom{n}{2}}$ in increasing order of weight.
- At step i, add edge e_i as long as it does not create a cycle.
- Stop when all vertices are connected.

Now consider the *Erdős–Rényi random graph process*, a natural coupling of the different Erdős–Rényi random graphs (see, for example, [7] and the references therein). This process may be obtained as follows: for a fixed parameter value $p \in [0, 1]$, keep all edges whose weight is at most p. By using the same edge-weights, we can easily couple the random graph process and a continuous-time version of Kruskal's algorithm so that, at a fixed time p, the components of the two processes have the same vertex-sets (and, indeed, so that the components of the Kruskal process are the MST's of the components of the Erdős–Rényi process). In particular, it is straightforward to see that the Kruskal process also undergoes the Erdős–Rényi phase transition at $p = 1/n$. In particular, for $p = (1 + \epsilon)/n$, there is a unique *giant component* containing a positive proportion of the vertices. It turns out that the metric structure of the MST has already essentially been built by this point (although the vast majority of the mass still lies outside the giant component).

In order to gain a finer understanding of the metric structure, we need to look earlier in the evolution of the process, within the *critical window* i.e. when $p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$ for some $\lambda \in \mathbb{R}$. Results from [1, 2] (described by Nicolas Broutin elsewhere in this report) tell us that for fixed λ, the Erdős–Rényi graph possesses a scaling limit when the graph distance is rescaled by $n^{-1/3}$ within each component. This scaling limit is a collection of random $\mathbb{R}$-trees in each of which a finite number of points have been identified to create cycles. By breaking these cycles appropriately, we are able to obtain that, at a fixed point λ in the critical window, the Kruskal process also has a scaling limit as a sequence of $\mathbb{R}$-trees. Starting from a large enough value of λ we are then able to control the way that the mass and metric structure evolve as the smaller trees gradually attach to the largest tree, in order to obtain our scaling limit.

The limit metric space. The limit metric space $\mathcal{M}$ is a random measured $\mathbb{R}$-tree, which is almost surely binary and whose mass measure is concentrated on its leaves. It shares all of these properties with the Brownian CRT, $\mathcal{T}$. However, they are not the same object. Suppose that $N(\epsilon)$ is the number of balls of radius $\epsilon > 0$ needed to cover $\mathcal{M}$. Then

$$\frac{N(\epsilon)}{\log(1/\epsilon)} \to 3$$

in probability, as $\epsilon \to 0$. Since $\mathcal{T}$ has box-counting dimension 2, it follows that the laws of $\mathcal{M}$ and $\mathcal{T}$ are mutually singular.

The results presented here are the subject of the paper [3] which is currently in preparation.

References

[1] L. Addario-Berry, N. Broutin, and C. Goldschmidt, *The continuum limit of critical random graphs*, Probability Theory and Related Fields **152**, 3-4 (2012), 367–406.

[2] L. Addario-Berry, N. Broutin, and C. Goldschmidt, *Critical random graphs: limiting constructions and distributional properties*, Electronic Journal of Probability **15** (2010), 741–775, Paper no. 25.

[3] L. Addario-Berry, N. Broutin, C. Goldschmidt, and G. Miermont, *The scaling limit of the minimum spanning tree of the complete graph*, in preparation.

[4] D. Aldous, *The continuum random tree. I*, Annals of Probability **19** (1991), 1–28.

[5] D. Aldous. *The continuum random tree II: an overview.* In M.T. Barlow and N.H. Bingham, editors, *Stochastic Analysis*, pages 23–70. Cambridge University Press, 1991.

[6] D. Aldous, *The continuum random tree. III*, Annals of Probabilty **21** (1993), 248–289.

[7] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[8] J.-F. Le Gall, *Random real trees*, Annales de la Faculté des Sciences de Toulouse Série 6 **XV** (2006), 35–62.

[9] J.-F. Le Gall, *Random trees and applications*, Probability Surveys **2** (2005), 245–311.