Geometry of large random trees

According to Cayley's formula, there are n^{n-1} rooted labelled trees with n vertices. But how many of them have a diameter, say 100, or are of height 45? Enumeration questions of this sort lead one naturally to look at distribution of the diameter (resp. height) of a uniformly sampled tree with n vertices. For these questions, classical combinatorial approaches often rely upon generating functions; see in particular [5], [6].

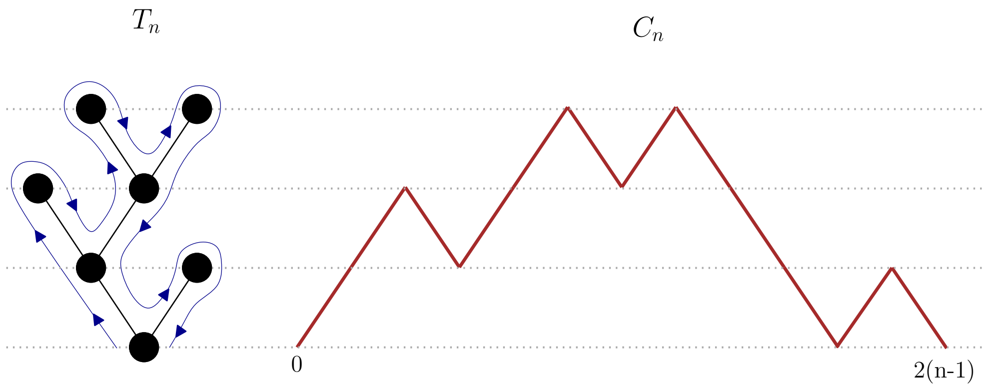

Figure 1: *The contour function of the tree T_n*

Alternatively, let us consider a function $C_n : [0, 2(n-1)] \to \mathbb{R}_+$, which depicts the "contour" of T_n in the following sense: imagine an exploration of T_n by a particle which starts from the root and visits continuously all edges at unit speed (assuming that every edge has unit length), backtracking as little as possible, and let $C_n(t)$ be the distance to the root from the particle's position at time t (see Fig. 1). The distribution of C_n is not simple in general. Nevertheless, it can be expressed as a functional of a random walk. Appealing to the invariance principal for random walks, one can show that

$$\Big(\tfrac{1}{\sqrt{n}} C_n\big(2(n-1)t\big)\Big)_{0\le t\le 1} \xrightarrow[n\to\infty]{\text{(in law)}} 2B^{\text{ex}} \ \text{ uniformly}, \tag{1}$$

where B^{ex} is the standard normalised Brownian excursion (of length 1). Now the height of T_n corresponds to $\max C_n$. As for the diameter, noting that the distance between two vertices visited respectively at time s and t is given by $d_C(s,t) := C_n(s) + C_n(t) - 2\min_{s\le u\le t} C_n(u)$, we easily see it can be written as $\max_{s,t} d_C(s,t)$. Then the convergence in (1) yields that

$$\tfrac{1}{\sqrt{n}}\,\text{Height}(T_n) \xrightarrow[n\to\infty]{\text{(in law)}} \max_s 2B^{\text{ex}}(s), \qquad \tfrac{1}{\sqrt{n}}\,\text{Diam}(T_n) \xrightarrow[n\to\infty]{\text{(in law)}} \max_{s,t} 2d_{B^{\text{ex}}}(s,t).$$

The distribution of $\max B^{\text{ex}}$ is well known. On the other hand, the work [7] explains a way to identify the distribution of $\max_{s,t} d_{B^{\text{ex}}}(s,t)$.

The convergence in (1) also has a "geometric" interpretation. Just as C_n is the contour function of T_n, the Brownian excursion $2B^{\text{ex}}$ is the contour function of the so-called *Brownian Continuum Random Tree*, so that (1) actually asserts a convergence of T_n to this tree, and the random variables $\max B^{\text{ex}}$ and $\max_{s,t} d_{B^{\text{ex}}}(s,t)$ can be identified as its height and diameter.

The above can be extended in various ways. For example, instead of uniformly sampled trees, one can consider trees sampled with probabilities proportional to some prescribed weight function. If we take the weight of a rooted tree t to be $w(t) = \prod_{i\geq 1} w_i^{n_i(t)}$, where $(w_i)_{i\geq 0}$ is a sequence of positive real numbers and $n_i(t)$ is the number of the vertices having out-degrees i in t, then very often this amounts to sampling a Galton–Watson trees conditioned to have n nodes. A celebrated theorem of Aldous [1] shows that if the offspring distribution of the Galton–Watson tree has finite second moment, then its contour function still converges in law to $2B^{\text{ex}}$ (up to a multiplicative constant). If, instead, the offspring distribution has infinite variance and belongs to the attraction domain of some α-stable law, $\alpha \in (1, 2]$, then the height (resp. diameter) of a such n-vertex Galton–Watson tree is of order $n^{1-1/\alpha}$; see [2]. The limit process replacing the Brownian excursion is the height process of α-stable process and it can be viewed as the contour function of the so-called α-stable tree, a sub-family of the Lévy trees of Le Gall & Le Jan [4]. The work [3] studies in particular the height and diameter of an α-stable tree.

It is also natural to look at *unrooted* trees. The work [8] considers an analogous model for (unrooted) random trees. As it turns out, the scaling limits of these trees are given by an unrooted version of Lévy trees.

References

[1] ALDOUS, D. The continuum random tree. III. *Ann. Probab. 21*, 1 (1993), 248–289.

[2] DUQUESNE, T. A limit theorem for the contour process of conditioned Galton-Watson trees. *Ann. Probab. 31*, 2 (2003), 996–1027.

[3] DUQUESNE, T., AND WANG, M. Decomposition of lévy trees along their diameter. *Ann. Inst. H. Poincaré Probab. Statist. 53*, 2 (2017), 539–593.

[4] LE GALL, J.-F., AND LE JAN, Y. Branching processes in Lévy processes: the exploration process. *Ann. Probab. 26*, 1 (1998), 213–252.

[5] RÉNYI, A., AND SZEKERES, G. On the height of trees. *J. Austral. Math. Soc. 7* (1967), 497–507.

[6] SZEKERES, G. Distribution of labelled trees by diameter. In *Combinatorial mathematics, X (Adelaide, 1982)*, vol. 1036 of *Lecture Notes in Math.* Springer, Berlin, 1983, pp. 392–397.

[7] WANG, M. Height and diameter of Brownian tree. *Electron. Commun. Probab. 20*, 88 (2015), 1–15.

[8] WANG, M. Scaling limits for a family of unrooted trees. *ALEA Lat. Am. J. Probab. Math. Stat. 13*, 2 (2016), 1039–1067.