A Modification of the Random Cutting Model

Fabian Burghart ✉

Department of Mathematics, Uppsala University, Sweden

── Abstract ──────────────────────────────────

We propose a modification to the random destruction of graphs: Given a finite network with a distinguished set of sources and targets, remove (cut) vertices at random, discarding components that do not contain a source node. We investigate the number of cuts required until all targets are removed, and the size of the remaining graph. This model interpolates between the random cutting model going back to Meir and Moon [28] and site percolation. We prove several general results, including that the size of the remaining graph is a tight family of random variables for compatible sequences of expander-type graphs, and determine limiting distributions complete binary trees.

1 Introduction and structure of the paper

Investigating the behaviour of trees when randomly removing vertices was first done by Meir and Moon in [28]. The process starts with a rooted tree, where at every step a uniformly chosen vertex is deleted, and all remaining components that do not include the root vertex are discarded. Since the process naturally stops once the root node has been cut, the question of interest is the random number of cuts needed to reach this state. In [28], the expected value and the variance of this random variable for a random labelled tree are found.

Since this initial paper, the model has been considered for different types of random and deterministic graphs, such as in the works of Panholzer [29, 30], Janson [20, 21] (see also [2] for an alternative approach by Addario-Berry, Broutin and Holmgren), or Holmgren [18, 19].

The random cutting model relates to both the record problem (as was observed in [21]) and to fragmentation processes and their Cut-trees, see e.g. [4], [5], or [8]. Moreover, it connects to Union-Find algorithms, see [31] and the further work in [13, 15, 22, 23].

In recent years, modifications of the original cutting model have been discussed: Kuba and Panholzer regarded the case of isolating a leaf, a general node, or multiple nodes simultaneously (instead of isolating the root) in [24, 25, 26], and Cai, Holmgren et al. proposed and investigated the k-cut model in [6, 10, 11], where a node is only removed after it has been cut for the k-th time.

In this paper, a different modification of the cutting model is introduced, where, additionally to one or several root vertices (which we will call sources) a second set of vertices (targets) are given. This allows for defining a stopping time for the cutting procedure on the graph by looking at the first moment when all of the targets have been removed (i.e. the

sources have been *separated* from the targets) – see Section 2 for the detailed definitions. We can then ask several natural questions, such as about the number of cuts necessary to separate the two sets and about the size of the remaining graph.

Section 3 will contain several basic estimates, and we will formalise the imprecise notion that separation interpolates between the cutting model and site percolation (Propositions 1 and 3). This requires the right definition that enables us to approximate such a graph by finite graphs, all while respecting the choice of sources and targets, see Definition 2.

In Section 4 we obtain the probability for a fixed subgraph to be the remaining graph at the time of separation. This leads to Theorem 9, which could be regarded as the main result of the paper and gives sufficient conditions for the size of the graph at separation to be a tight sequence of random variables when the graph approximates a locally finite infinite graph in the sense of Definition 2.

For the final section, our scope will focus on rooted trees, since their recursive nature can be used to simplify many of the arguments and calculations. The arguments here relate to earlier work by Devroye et al., [12, 14]. We consider the separation sizes and separation times for complete binary trees as an illustrating example, and finish by indicating potential future research directions.

In order to comply with the requirements for this extended abstract, proofs of auxiliary statements have been omitted, but can be found in the full version of this paper, [9]. Moreover, the full version contains several examples that were not included here.

2   Cutting procedures

2.1   Some notation

We will always use $G = (V(G), E(G))$ to denote a graph, consisting of its vertex and edge set, but will shorten the notation to $V = V(G)$ and $E = E(G)$ if there is no ambiguity from the context. Since most subgraphs we will consider are induced and therefore uniquely determined by their vertex set, we will not distinguish between an induced subgraph and its vertex (sub-)set.

If two vertices v, w are neighbours, we also use the notation $v \sim w$. More generally, we write $\mathrm{dist}(v, w)$ for the graph distance between vertices v, w. In the case where $A, B \subseteq V(G)$ are subsets, $\mathrm{dist}(A, B)$ is to be understood as $\min\{\mathrm{dist}(v, w) : v \in A, w \in B\}$.

Given any set $A \subseteq V(G)$ and a fixed set S of source nodes, we define the closure of A to be $\mathrm{clos}_S(A) := A \cup S \cup \{v \in V(G) : v \sim w \text{ for some } w \in A\}$. The (exterior) boundary of A is defined as $\partial_S A := \mathrm{clos}_S(A) \setminus A$. In other words, the vertices in $\partial_S A$ are precisely the vertices not in A that are in S or neighbour some vertex in A. Note that this implies e.g. $\partial_S \emptyset := S$.

2.2   Cutting and separation

Consider a finite simple connected graph $G = (V, E)$ with a distinguished subset $S \subseteq V$ whose vertices are referred to as sources. Now, proceed as follows:

1. Choose a vertex $v \in V$ uniformly at random, and remove it – together with all edges incident to v – from the graph. This will potentially split the graph into connected components, in which case we only keep the components containing sources, regarding them jointly as a new (potentially disconnected graph).

2. Iterate step 1, where the randomness in choosing the node is assumed to be independent from everything that happened previously.

3. The process terminates once the graph contains no more vertices. Equivalently, this happens as soon as the last source node has been removed.

This defines a finite sequence $G =: G_0 \supsetneq G_1 \supsetneq ... \supsetneq G_{r-1} \supsetneq G_r = \emptyset$ of random induced subgraphs, where we denoted the empty subgraph consisting of no vertices with $\emptyset$. We will denote this process by $\mathsf{Cut}(G)$.

Introducing a second set of distinguished vertices, T, whose vertices we refer to as targets, we can now consider the following functionals of the cutting process:

- The cutting number $\mathfrak{C}(G)$. This is merely the number of cuts until the last source node is cut, or equivalently, until the remaining graph is empty, i.e. $\mathfrak{C}(G) = \inf\{i \geq 0 : G_i = \emptyset\}$. Note that this does not rely on T.
- The separation number $\mathfrak{S}(G)$, defined to be the number of cuts until the remaining graph does not contain target nodes anymore (independently of how many sources are still present). In other words, $\mathfrak{S}(G) = \inf\{i \geq 0 : V(G_i) \cap T = \emptyset\}$. We say that at this time, separation (of S and T) occurs.
- The separation subgraph $G_\mathfrak{S} := G_{\mathfrak{S}(G)}$, defined to be the random subgraph of G at separation.
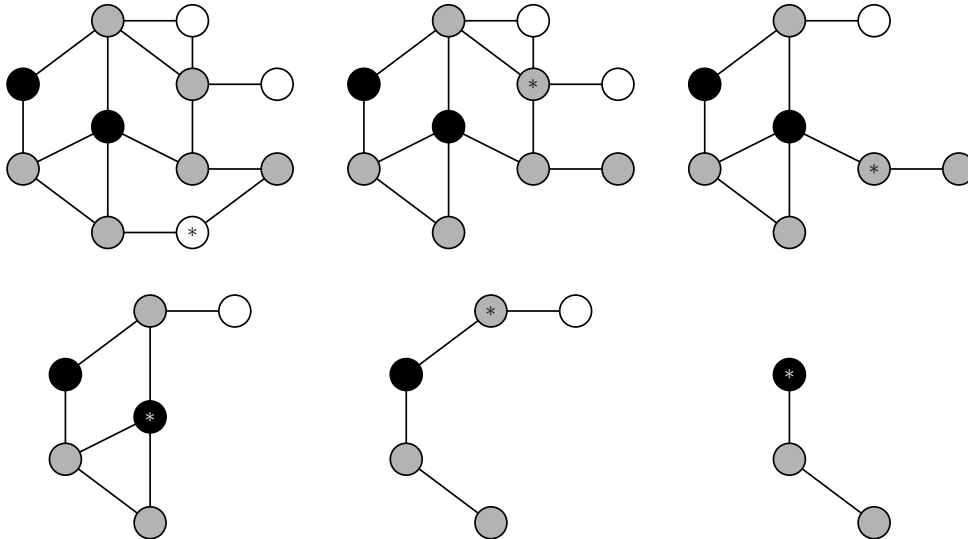
Figure 1 The cutting procedure on a graph G from top left to bottom right, with source nodes in black and target nodes in white. The symbol $*$ in a vertex indicates that this is the vertex about to be cut in the next step. In this example, $\mathfrak{S}(G) = 5$, $\mathfrak{C}(G) = 6$ (after the last source node has been cut) and $G_\mathfrak{S}$ is a path on three vertices.

2.3 The continuous-time model

As has been observed previously by [21] and since been brought to effective use, the above cutting model is equivalent to a model where each node is equipped with a random alarm clock whose alarm triggers after time X_v, $v \in V$. Whenever an alarm rings, and the corresponding node is still in the graph at this time, that node will be removed together with any new components that do not contain source nodes. Here, to ensure equivalence to the discrete-time cutting model above, we assume that $(X_v)_{v \in V}$ is an i.i.d. family of $\mathsf{Exp}(1)$-distributed random variables – while any i.i.d. family of continuous random variables would suffice, the memoryless property will be useful later.

This once again yields a monotone stochastic process of subgraphs of G, but now parametrised by continuous time, $(G_t^c)_{t \in [0,\infty)}$. We will denote this process by $\mathsf{Cut}^c(G)$. However, G_t will only attain finitely many different graphs, and we will still denote those graphs by $G_0, G_1, G_2, \dots$ in order of occurrence, as before. Hence, we can denote by G_{t-} the graph that was attained by $\mathsf{Cut}^c(G)$ immediately before time t; so, $G_{t-} = G_t$ iff no cut happened at time t.

Note that there are two ways of generalising the random variables $\mathfrak{C}$ and $\mathfrak{S}$ to the continuous-time setting: By default, $\mathfrak{C}$ and $\mathfrak{S}$ respectively denote the quantities

$$\mathfrak{C} = \inf\{i \in \mathbb{N} : G_i = \emptyset\} \quad \text{and} \quad \mathfrak{S} = \inf\{i \in \mathbb{N} : V(G_i) \cap T = \emptyset\},$$

exactly as before, while $\mathfrak{C}_c$ and $\mathfrak{S}_c$ denote

$$\mathfrak{C}_c = \inf\{t \geq 0 : G_t = \emptyset\} \quad \text{and} \quad \mathfrak{S}_c = \inf\{t \geq 0 : V(G_t) \cap T = \emptyset\},$$

respectively.

3 Cutting and site percolation

The following proposition asserts that the additional freedom of choosing target nodes for the separation number can be used to obtain the original cutting number. In other words, $\mathfrak{S}(G)$ can be understood as a generalisation of $\mathfrak{C}(G)$.

▶ **Proposition 1.** *Let $G = (V, E)$ be a finite connected graph, and let $S, T \subseteq V$ be the sets of source and target nodes, respectively. Then, we have $\mathfrak{S}(G) \leq \mathfrak{C}(G)$ deterministically, with equality if $S \subseteq T$. Conversely, if $\mathfrak{S}(G) = \mathfrak{C}(G)$ in distribution, then $S \subseteq T$. Moreover, all of those statements also hold true for $\mathfrak{S}_c(G)$ and $\mathfrak{C}_c(G)$ in the continuous-time model.*

We remark that therefore, $\mathfrak{S}(G) = \mathfrak{C}(G)$ holds in distribution if and only if it holds deterministically.

Proof. At time $\mathfrak{C}(G)$, the remaining graph is empty, so separation must have occurred already. Thus $\mathfrak{S}(G) \leq \mathfrak{C}(G)$.

If $S \subseteq T$, then separation will occur as soon as the last source node has been removed, at which time the remaining graph will be empty. Thus $\mathfrak{S}(G) = \mathfrak{C}(G)$. Conversely, if there exists $v \in S \setminus T$ then $G_\mathfrak{S}$ contains v with some positive probability p_0. If this happens, $\mathfrak{S}(G) \leq \mathfrak{C}(G) - 1$, so $\mathbf{E}[\mathfrak{C}(G)] - \mathbf{E}[\mathfrak{S}(G)] \geq p_0$, and equality in distribution cannot hold.

For the continuous-time model, only the last argument requires modification: Once again, if $v \in S \setminus T$ then $G_\mathfrak{S}$ contains v with probability $p_0 > 0$. In this case, $\mathfrak{S}_c(G) < X_v \leq \mathfrak{C}_c(G)$, and we have

$$\mathbf{E}[\mathfrak{C}_c(G)] - \mathbf{E}[\mathfrak{S}_c(G)] \geq p_0 \mathbf{E}[X_v - \mathfrak{S}_c(G) \mid v \in V(G_\mathfrak{S})] = p_0$$

since $X_v \sim \mathsf{Exp}(1)$ is memoryless.   ◀

We now show that in a certain sense, the continuous-time separation model on an infinite graph G with infinite distance $\mathrm{dist}(S, T)$ contains the site percolation model on G.

More precisely, recall that for Bernoulli site percolation in an infinite graph G, every node is independently kept with some probability $p \in [0, 1]$ and otherwise rejected, thus giving a random subgraph of G. We denote by $\mathrm{perc}_S(p)$ the probability that the $\mathsf{Ber}(p)$-site percolation on G exhibits an infinite cluster containing at least one vertex of S.

▶ **Definition 2.** *Let G be a locally finite, infinite connected graph, containing two subsets $S, T \subseteq V(G)$. We say that the sequence $(G^{(n)})_{n \geq 1}$ of finite induced subgraphs of G exhausts G if the following conditions are satisfied:*

(i) *The $G^{(n)}$ are connected subgraphs satisfying*

$$G^{(1)} \subseteq G^{(2)} \subseteq ... \subseteq G \quad and \quad \bigcup_{n \geq 1} V\left(G^{(n)}\right) = V(G).$$

(ii) *The set S is entirely contained in $G^{(n)}$ for all n (and understood to be the set of source nodes of $G^{(n)}$), and each subgraph $G^{(n)}$ is endowed with the target set $T^{(n)} := \left(T \cap V\left(G^{(n)}\right)\right) \cup \partial_\emptyset \left(G \setminus G^{(n)}\right)$.*

Observe that the target nodes $T^{(n)}$ are indeed a subset of $V\left(G^{(n)}\right)$ and will be non-empty even if $T = \emptyset$. Moreover, condition (ii) necessitates that S is finite.

▶ **Proposition 3.** *Let G be a locally finite, infinite graph with a finite set of source nodes, S, and let $T = \emptyset$. Assume that $(G^{(n)})_{n \geq 1}$ exhausts G.*

Then,

$$\lim_{n \to \infty} \mathbf{P}\left[\mathfrak{S}_c\left(G^{(n)}\right) \geq x\right] = \mathrm{perc}_S(e^{-x}). \tag{1}$$

Proof. Note first that if $T = \emptyset$, then (using the notation of Definition 2) $\mathrm{dist}\left(S, T^{(n)}\right) \to \infty$ as $n \to \infty$. Indeed, there would otherwise be a bound $C \in \mathbb{N}$ with $\mathrm{dist}\left(S, T^{(n)}\right) < C$. Consider the neighbourhood $B_C(S) := \{v \in V(G) : \mathrm{dist}(v, S) \leq C\}$ of S. Since $B_C(S)$ is finite, it will eventually be contained in all $G^{(n)}$, contradicting the notion that $T^{(n)} = \partial_\emptyset \left(G \setminus G^{(n)}\right)$ contains vertices in $B_{C-1}(S)$.

Recall next that independently removing each vertex v of G at a random time $X_v \sim \mathsf{Exp}(1)$ gives rise to the monotonous coupling of Bernoulli site percolation for all parameters $p \in [0, 1]$ (cf. [27, p.138]). Indeed, at time $x \in [0, \infty]$, the graph we observe that way is a sample of $\mathsf{Ber}(e^{-x})$-site percolation on G. We can couple the process obtained in this way to the continuous-time cutting model by restricting our attention to the intersection of $G^{(n)}$ with those percolation clusters that intersect S.

To show $\geq$ in (1), assume that $\mathsf{Ber}(e^{-x})$-site percolation exhibits an infinite cluster which intersects S, such cluster necessarily intersects $T^{(n)}$ as well and hence, for each n, contains a path connecting S with $T^{(n)}$. By the coupling indicated above, this path must then also be present in the sample of the continuous-time cutting model on $G^{(n)}$ at time x. Therefore, $\mathrm{perc}_S(e^{-x}) \leq \mathbf{P}\left[\mathfrak{S}_c\left(G^{(n)}\right) \geq x\right]$, and letting n tend to ∞ yields $\liminf_{n \to \infty} \mathbf{P}\left[\mathfrak{S}_c\left(G^{(n)}\right) \geq x\right] \geq \mathrm{perc}_S(e^{-x})$.

For the other inequality, suppose now that $\mathsf{Ber}(e^{-x})$-site percolation does not exhibit an infinite cluster intersecting S, so that the total mass of clusters intersecting S is bounded by some finite integer, say k. By the second assumption, we have $\mathrm{dist}(S, T^{(n)}) > k$ for all but finitely many n. However, this implies that eventually, the clusters intersecting S cannot intersect $T^{(n)}$, which, for the coupled cutting procedure, means that separation in $G^{(n)}$ must have occurred before time x. So,

$$\mathbf{P}\left[|V(\text{Clusters intersecting } S)| \leq k\right] \leq \mathbf{P}\left[\mathfrak{S}_c\left(G^{(n)}\right) < x\right].$$

Taking the limit superior for $n \to \infty$ yields

$$\limsup_{n \to \infty} \mathbf{P}\left[\mathfrak{S}_c\left(G^{(n)}\right) \geq x\right] \leq 1 - \mathbf{P}\left[|V(\text{Clusters intersecting } S)| \leq k\right],$$

which implies the existence of the limit and $\leq$ in (1) after passing to the limit for $k \to \infty$ as well. ◀

4 Visiting probability of subgraphs and size of the separation graph

Consider a finite simple connected graph G with $S, T \subseteq V$ as usual. The aim of this section is to determine the probability that at some time $i \geq 1$, the cutting procedure $\mathsf{Cut}(G)$ will produce a specific subgraph G_*.

▶ **Lemma 4.** *Fix an induced subgraph G_* of G with every component of G_* containing at least one source node. Then, for all times $t \geq 0$ in the continuous-time cutting model, we have*

$$\mathbf{P}\left[G_t = G_*\right] = e^{-|G_*|t}\left(1 - e^{-t}\right)^{|\partial_S G_*|}. \tag{2}$$

Moreover, consider $v_* \in \partial_S G_*$. Then

$$\mathbf{P}\left[G_{X_{v_*}} = G_* \mid X_{v_*}\right] = e^{-|G_*|X_{v_*}}\left(1 - e^{-X_{v_*}}\right)^{|\partial_S G_*|-1}. \tag{3}$$

▶ **Corollary 5.** *Fix an induced subgraph G_* of G with every component of G_* containing at least one source node. Let $v_* \in \partial_S G_*$. Denote the i-th graph obtained in the cutting process by G_i and the i-th cut node by v_i. Then*

$$\mathbf{P}\left[\exists i \geq 1 : G_i = G_* \text{ and } v_i = v_*\right] = \frac{1}{|\partial_S G_*|}\binom{|G_*| + |\partial_S G_*|}{|\partial_S G_*|}^{-1} \tag{4}$$

and therefore

$$\mathbf{P}\left[\exists i \geq 1 : G_i = G_*\right] = \binom{|G_*| + |\partial_S G_*|}{|\partial_S G_*|}^{-1}. \tag{5}$$

▶ **Definition 6.** *Let G be a finite connected graph equipped with sources S and targets $T \neq \emptyset$. An induced subgraph G_* is called* admissible *if the probability $\mathbf{P}[G_{\mathfrak{S}} = G_*]$ is positive. We denote by $\mathscr{A}_m(G)$ the set of all admissible subgraphs G_* of G of size $|V(G_*)| = m$.*

Assuming that $T \neq \emptyset$, it can be shown (Lemma 12 in [9]) that $G_* \subseteq G$ is admissible if and only if G_* contains no target nodes and every component of G_* contains at least one source node.

Relying on the preceding results, we can establish the following connection between the graph $G_{\mathfrak{S}}$ at separation and the continuous-time separation number $\mathfrak{S}_c$:

▶ **Proposition 7.** *Let G_* be an admissible subgraph of G. Then,*

$$\mathbf{P}[G_{\mathfrak{S}} = G_*] = \sum_{v_* \in \partial_S G_*} \int_0^1 u^{|G_*|-1}(1-u)^{|\partial_S G_*|-1}\mathbf{P}\left[\mathfrak{S}_c(H[v_*]) \geq -\ln u\right]\,\mathrm{d}u, \tag{6}$$

where $H[v_*]$ denotes the graph obtained from G in the following way: Remove all vertices in G_* and in $\partial_S G_* \setminus \{v_*\}$ from G, and from what remains, let $H[v_*]$ be the connected component containing v_*. We endow $H[v_*]$ with source node v_* and target nodes $T \cap V(H[v_*])$.

Proof. Fix a vertex $v_* \in \partial_S G_*$, and assume that this is the last node to be removed for separation to occur. We observe first that by definition of the separation number, any graphs obtained by $\mathsf{Cut}(G)$ before separation must have contained a path from v_* to T. In particular, the last graph before separation occurred contained such a path, which additionally was not

passing through any other nodes in G_* or $\partial_S G_*$ and must have therefore been contained in $H[v_*]$. The existence of such a path means, however, that the graph $H[v_*]$ is not yet separated. Thus, by transitioning from the discrete to the continuous-time model, we obtain

$$\mathbf{P}[G_{\mathfrak{S}} = G_*, v_{\mathfrak{S}} = v_* \mid X_{v_*}]$$
$$= \mathbf{P}\left[G_{X_{v_*}} = G_* \text{ and } v_* \text{ connects to } T \text{ in } G_{X_{v_*}^-} \mid X_{v_*}\right]$$
$$= \mathbf{P}\left[G_{X_{v_*}} = G_* \mid X_{v_*}\right] \mathbf{P}\left[\mathfrak{S}_c(H[v_*]) \geq X_{v_*} \mid X_{v_*}\right], \tag{7}$$

where conditional independence holds true because

$$\{G_{X_{v_*}} = G_*\} \quad \text{and} \quad \{\mathfrak{S}_c(H[v_*]) \geq X_{v_*}\}$$

are events on vertex sets which only share v_*. Conditioned on X_{v_*} being x, the event $\{\mathfrak{S}_c(H[v_*]) \geq X_{v_*}\}$ amounts to the existence of a path from v_* to the set of target nodes in $H[v_*]$, none of whose clocks have rung yet at time X_{v_*}. On the other hand, without the conditioning, the same event is equivalent to the existence of a path from a neighbour of v_* to the set of target nodes in $H[v_*]$. Hence,

$$\mathbf{P}\left[\mathfrak{S}_c(H[v_*]) \geq X_{v_*} \mid X_{v_*} = x\right] = \frac{\mathbf{P}\left[\mathfrak{S}_c(H[v_*]) \geq x\right]}{e^{-x}}.$$

In light of (3) from Lemma 4, we can now rewrite equation (7) as

$$\mathbf{P}[G_{\mathfrak{S}} = G_*, v_{\mathfrak{S}} = v_* \mid X_{v_*} = x] = e^{-|G_*|x}\left(1 - e^{-x}\right)^{|\partial_S G_*|-1} \frac{\mathbf{P}\left[\mathfrak{S}_c(H[v_*]) \geq x\right]}{e^{-x}}. \tag{8}$$

Finally, observe that, with μ_X denoting the distribution of X_{v_*},

$$\mathbf{P}[G_{\mathfrak{S}} = G_*] = \sum_{v_* \in \partial_S G_*} \int_0^\infty \mathbf{P}[G_{\mathfrak{S}} = G_*, v_{\mathfrak{S}} = v_* \mid X_{v_*} = x]\, \mathrm{d}\mu_X(x),$$

so that, after plugging in the expression from (8) and using $X_{v_*} \sim \mathsf{Exp}(1)$, we obtain

$$\mathbf{P}[G_{\mathfrak{S}} = G_*] = \sum_{v_* \in \partial_S G_*} \int_0^\infty (e^{-x})^{|G_*|} \left(1 - e^{-x}\right)^{|\partial_S G_*|-1} \frac{\mathbf{P}\left[\mathfrak{S}_c(H[v_*]) \geq x\right]}{e^{-x}} \cdot e^{-x}\, \mathrm{d}x,$$

which only differs from (6) by the substitution $e^{-x} = u$. ◄

▶ Remark 8. The (unconditioned) probability $\mathbf{P}[\mathfrak{S}_c(H[v_*]) \geq X_{v_*}]$, has a number of equivalent versions. Indeed, if we consider any G with $S = \{v_*\}$, then we have the following equalities:

$$\mathbf{P}[\mathfrak{S}_c(G) \geq X_{v_*}] = \mathbf{P}[\mathfrak{S}_c(G) \geq \mathfrak{C}_c(G)] = \mathbf{P}[\mathfrak{S}(G) \geq \mathfrak{C}(G)] = \mathbf{P}[G_{\mathfrak{S}} = \emptyset] = \mathbf{P}[|G_{\mathfrak{S}}| = 0].$$

Moreover, in the first three formulations, the strict inequality ">" is impossible, so one could just as well write "=" there. Additionally, since $\mathfrak{S}_c(H[v_*]) \leq X_{v_*}$, we have the estimate

$$\mathbf{P}\left[\mathfrak{S}_c(H[v_*]) \geq -\ln u\right] \leq \mathbf{P}\left[X_{v_*} \geq -\ln u\right] = u \qquad \forall u \in [0,1]. \tag{9}$$

Recall that a family of real-valued random variables $X_i, i \in I$, is *tight* if for all $\varepsilon > 0$, there exists a constant M such that $\mathbf{P}[|X_i| \geq M] < \varepsilon$ for all $i \in I$, cf. [7, p.37]. We conclude this section by showing the following tightness result about the size of $G_{\mathfrak{S}}$:

▶ **Theorem 9.** *Let $(G^{(n)})_{n \geq 1}$ be a sequence of finite induced subgraphs exhausting the locally finite, infinite graph G equipped with subsets $S, T \subseteq V(G)$. Define $a_{m,n} = \sum_{A \in \mathscr{A}_m(G^{(n)})} |\partial_S A|$ and let $a_m := \lim_{n \to \infty} a_{m,n}$. Assume that there are constants $b > 0$ and $L \geq 0$ such that*

  (i) *$|\partial_S A| \geq bm + 1$ for all $A \in \mathscr{A}_m(G^{(n)})$ and all $m \geq L$, and*
  (ii) *the function $f(x) = \sum_{m=L}^{\infty} a_m x^m$ has radius of convergence at least $\frac{b^b}{(b+1)^{b+1}}$ and satisfies*

$$\int_0^1 f(x(1-x)^b)\, \mathrm{d}x < \infty \tag{10}$$

Then the sizes of the separation graphs, $\left|G_{\mathfrak{S}}^{(n)}\right|$, form a tight family of random variables.

Observe that, since $c := \max_{x \in [0,1]} x(1-x)^b = \frac{b^b}{(b+1)^{(b+1)}}$, condition (ii) requires that the radius of convergence, r, of the power series f is at least c. In case $r > c$, this condition is trivially satisfied as the integrand is bounded. However, in case of equality $r = c$, f has a singularity at c by virtue of Pringsheim's theorem, cf. [16, Theorem IV.6], and (10) is a non-trivial requirement. We also remark that condition (i) above is a variant of the notion of expander graphs, which play a crucial role in the theory of percolation, see e.g. [3].

Proof. We first show that for fixed m, the sequence $a_{m,n}$ is nondecreasing and eventually constant. Recall from Definition 2 that the sets $T^{(n)}$ consist of two parts, namely $T \cap V\left(G^{(n)}\right)$ and vertices that are incident to edges leaving $G^{(n)}$. By copying the respective argument from the proof of Proposition 3, we can show again that dist $\left(S, T^{(n)} \setminus T\right) \to \infty$ as $n \to \infty$. Hence, for n sufficiently large, the closed neighbourhood of S of radius m will become independent of n, and so will $a_{m,n}$. Moreover, since the $G^{(n)}$ are monotonously growing, any admissible subgraph in $G^{(n)}$ will also be admissible in $G^{(n+1)}$ and the number of boundary vertices will not decrease when changing from $G^{(n)}$ to $G^{(n+1)}$. Hence $a_{m,n}$ is nondecreasing in n. In particular, the limit a_m exists, is finite, and an upper bound to $a_{m,n}$ for all n.

Let $M \geq L$. We now apply Proposition 7, where we set $p_{v_*}(x) := \mathbf{P}[\mathfrak{S}_c(H[v_*]) \geq -\ln x]$ (note that this still depends on $G_{\mathfrak{S}}$ as well!) for brevity. Summing over all $m \geq M$ and all $A \in \mathscr{A}_m(G^{(n)})$ yields

$$\mathbf{P}\left[\left|G_{\mathfrak{S}}^{(n)}\right| \geq M\right] = \sum_{m=M}^{\infty} \sum_{A \in \mathscr{A}_m(G^{(n)})} \int_0^1 x^{m-1}(1-x)^{|\partial_S A|-1} \sum_{v_* \in \partial_S A} p_{v_*}(x)\, \mathrm{d}x$$

$$\leq \sum_{m=M}^{\infty} \sum_{A \in \mathscr{A}_m(G^{(n)})} |\partial_S A| \int_0^1 x^m (1-x)^{|\partial_S A|-1}\, \mathrm{d}x,$$

where we applied the estimate (9). Using assumption (i), we get

$$\mathbf{P}\left[\left|G_{\mathfrak{S}}^{(n)}\right| \geq M\right] \leq \sum_{m=M}^{\infty} \sum_{A \in \mathscr{A}_m(G^{(n)})} |\partial_S A| \int_0^1 (x(1-x)^b)^m\, \mathrm{d}x$$

$$= \sum_{m=M}^{\infty} \int_0^1 a_{m,n}(x(1-x)^b)^m\, \mathrm{d}x.$$

Then, by the above argument on the monotonicity of $a_{m,n}$, we obtain

$$\mathbf{P}\left[\left|G_{\mathfrak{S}}^{(n)}\right| \geq M\right] \leq \sum_{m=M}^{\infty} \int_0^1 a_m (x(1-x)^b)^m\, \mathrm{d}x.$$

By monotone convergence, we moreover obtain for all $M \geq L$ that

$$\sum_{m=M}^{\infty} \int_0^1 a_m(x(1-x)^b)^m \, \mathrm{d}x = \int_0^1 \sum_{m=M}^{\infty} a_m(x(1-x)^b)^m \, \mathrm{d}x$$

$$\leq \int_0^1 \sum_{m=L}^{\infty} a_m(x(1-x)^b)^m \, \mathrm{d}x = \int_0^1 f(x(1-x)^b) \, \mathrm{d}x$$

which is finite by assumption (ii). Hence the tail of the series on the left-hand side converges to zero, and therefore $\mathbf{P}\left[\left|G_{\mathfrak{S}}^{(n)}\right| \geq M\right] \to 0$ uniformly in n as $M \to \infty$ as well. ◄

▶ **Remark 10.** Tightness of the sizes of the separation graphs $\left|G_{\mathfrak{S}}^{(n)}\right|$ is a helpful property in order to translate limit laws from the cutting times $\mathfrak{C}\left(G^{(n)}\right)$ to the separation times $\mathfrak{S}\left(G^{(n)}\right)$: Assume that there exist sequences α_n and $\beta_n > 0$ such that

$$\frac{\mathfrak{C}\left(G^{(n)}\right) - \alpha_n}{\beta_n} \xrightarrow{\mathscr{D}} X \qquad \text{as } n \to \infty$$

for a random variable X with positive variance. If $\beta_n \to \infty$ and $\left|G_{\mathfrak{S}}^{(n)}\right|$ forms a tight family of random variables, then the deterministic estimate $\mathfrak{S}\left(G^{(n)}\right) \leq \mathfrak{C}\left(G^{(n)}\right) \leq \mathfrak{S}\left(G^{(n)}\right) + \left|G_{\mathfrak{S}}^{(n)}\right|$ implies

$$\frac{\mathfrak{S}\left(G^{(n)}\right) - \alpha_n}{\beta_n} \xrightarrow{\mathscr{D}} X \qquad \text{as } n \to \infty.$$

5 Separating trees

In this section, let $G = \mathcal{T}$ be a rooted tree, where we will always interpret the root node as the (unique) source vertex, and the leaves as targets.

To each node $w \in V(\mathcal{T})$, we assign a polynomial $p[w]$ from $\mathbb{Z}[x]$ recursively as follows: If w is a leaf, define $p[w](x) = x$. Otherwise, denote the children of w by $w_1, ..., w_r$ for $r \geq 1$. Then, define

$$p[w](x) := x\left(1 - \prod_{i=1}^{r}(1 - p[w_i](x))\right). \tag{11}$$

Observe that in the case where w only has a single child w_1, this simplifies to $p[w](x) = xp[w_1](x)$. Furthermore, if $\mathcal{T}_*$ is the fringe subtree of $\mathcal{T}$ rooted at w (i.e. the subtree consisting of w together with all its descendants), we also write $p[\mathcal{T}_*] := p[w]$. In particular, $p[\mathcal{T}] := p[\text{root}]$.

▶ **Proposition 11.** *For the continuous-time cutting model on a rooted tree $\mathcal{T}$, we have*

$$\mathbf{P}[\mathfrak{S}_c(\mathcal{T}) \geq x] = p[\mathcal{T}](e^{-x}) \tag{12}$$

for all $x \geq 0$. Equivalently, one can interpret $p[\mathcal{T}](q)$ for $q \in [0,1]$ as the probability that $\mathrm{Ber}(q)$-site percolation on $\mathcal{T}$ contains a path from the root to a leaf.
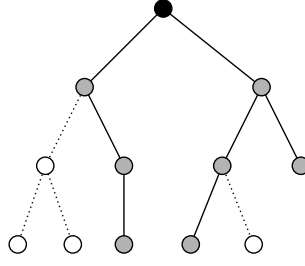
■ **Figure 2** A faithful subtree of an underlying rooted tree $\mathcal{T}$. The root is shown in black, with vertices belonging to the subtrees being shaded grey. Dotted edges and white vertices belong to $\mathcal{T}$, but not to the subgraph.

▶ **Corollary 12.** *We have*

$$\mathbf{P}[G_{\mathfrak{S}} = \emptyset] = \int_0^1 \frac{p[\mathcal{T}](u)}{u}\,\mathrm{d}u. \tag{13}$$

Moreover, we have for any subtree $\mathcal{T}_* \subseteq \mathcal{T}$ containing the root node but none of the leaves:

$$\mathbf{P}[G_{\mathfrak{S}} = \mathcal{T}_*] = \sum_{v_* \in \partial \mathcal{T}_*} \int_0^1 u^{|\mathcal{T}_*|-1}(1-u)^{|\partial \mathcal{T}_*|-1} p[v_*](u)\,\mathrm{d}u. \tag{14}$$

A *transversal* in a rooted tree $\mathcal{T}$ is defined to be a subset of vertices that intersect every path from the root to a leaf. It then follows from the proof of Proposition 11 that $1 - p[\mathcal{T}](1-q)$ yields the probability that a random set of vertices, containing each vertex independently with probability q, is a transversal of $\mathcal{T}$. It is this expression that was investigated in [14, 12].

▶ **Remark 13.** There are some straightforward conclusions to be drawn from the previous proposition: The polynomials $p[\mathcal{T}](x)$ map the interval $[0,1]$ to itself, with $p[\mathcal{T}](0) = 0$ and $p[\mathcal{T}](1) = 1$. Furthermore, they are monotonically increasing on this interval, and by (11), $p[\mathcal{T}](x) \leq x$ for $x \in [0,1]$.

▶ **Definition 14.** *Let $\mathcal{T}$ be a rooted tree. A subtree $\mathcal{T}_*$ is called* faithful *if it contains the root of $\mathcal{T}$ as its own root and if every leaf of $\mathcal{T}_*$ is a leaf of $\mathcal{T}$. We denote the set of all faithful subtrees of $\mathcal{T}$ by $\mathscr{F}(\mathcal{T})$, and the set of all faithful subtrees on n vertices by $\mathscr{F}_n(\mathcal{T})$. Equivalently, a faithful subtree $\mathcal{T}_* \subseteq \mathcal{T}$ can be seen as choosing a number of paths from the root to the leaves of $\mathcal{T}$. We hence denote the set of all faithful subtrees of $\mathcal{T}$ with exactly ℓ leaves by $\mathscr{P}_\ell(\mathcal{T})$.*

The following propositions gives an alternative, combinatorial characterisation of $p[\mathcal{T}]$ which might be of independent interest:

▶ **Proposition 15.** *Write $p[\mathcal{T}](x) = \sum_{j=0}^n a_j x^j$ with $a_j \in \mathbb{Z}$ for $j = 0, 1, ..., n$. Denote by $L(\mathcal{T}_*)$ the set of leaves of $\mathcal{T}_*$. Then,*

$$a_j = \sum_{\mathcal{T}_* \in \mathscr{F}_j(\mathcal{T})} (-1)^{|L(\mathcal{T}_*)|+1} \tag{15}$$

$$= \sum_{\ell \geq 1} (-1)^{\ell+1} |\mathscr{F}_j(\mathcal{T}) \cap \mathscr{P}_\ell(\mathcal{T})|. \tag{16}$$

In particular, $\deg(p[\mathcal{T}]) = |\mathcal{T}|$ and $a_j = 0$ for $j \leq \mathrm{dist}(\mathrm{root}, L(\mathcal{T}))$.

▶ **Example 16.** Denote by CBT_n the full complete rooted binary tree on $2^n - 1$ vertices, this being the rooted binary tree having 2^h vertices at every height $h = 0, ..., n-1$. Observe that CBT_1 is the tree consisting of only the root node, and that CBT_{n+1} splits into two copies of CBT_n upon removing the root node. Thus, the associated polynomials satisfy the recurrence relation

$$p[\mathrm{CBT}_{n+1}](x) = x(1 - (1 - p[\mathrm{CBT}_n](x))^2) = 2xp[\mathrm{CBT}_n](x) - xp[\mathrm{CBT}_n](x)^2$$

with $p[\mathrm{CBT}_1](x) = x$. It can be shown that $p[\mathrm{CBT}_n]$ converges pointwise and monotonically decreasing to

$$\varphi(x) := \max\left\{0, 2 - \frac{1}{x}\right\} \tag{17}$$

as $n \to \infty$, for $x \in [0, 1]$.

In light of Proposition 3, this result should not be surprising: The sequence of rooted trees CBT_n satisfies all the conditions, and the function $\varphi(x)$ indeed equals the probability that the root node is contained in an infinite cluster of $\mathsf{Ber}(x)$-site percolation, as can be verified independently, e.g. from [17, p.256].

Continuing with our analysis, the probability of the remaining tree at separation being empty now follows handily from Corollary 12:

$$\lim_{n \to \infty} \mathbf{P}[|\mathrm{CBT}_{n,\mathfrak{S}}| = 0] = \int_0^1 \frac{\varphi(x)}{x}\, \mathrm{d}x = \ln 4 - 1 \approx 0.3863$$

In a similar fashion, we can continue to determine the limiting probability of separation graphs of any size $m \geq 0$: Since there are $C_m = \frac{1}{m+1}\binom{2m}{m}$-many[1] subtrees of the infinite rooted binary tree on m vertices, and each of those has $m + 1$ boundary vertices, it can be shown using the dominated convergence theorem that

$$p_m := \lim_{n \to \infty} \mathbf{P}[|\mathrm{CBT}_{n,\mathfrak{S}}| = m] = \binom{2m}{m}\int_0^1 x^{m-1}(1-x)^m\varphi(x)\, \mathrm{d}x. \tag{18}$$

Observe that, in the notation of Theorem 9, the sequence $a_m = \binom{2m}{m}$ has generating function $\frac{1}{\sqrt{1-4x}}$ (cf. [1, OEIS A000984]), thus violating the integrability condition (10) for the proper constant $b = 1$. However, we can check by hand that (18) defines a probability distribution with the following computation, relying on the generating function of $\binom{2m}{m}$:

$$\sum_{m=0}^{\infty} p_m = \int_{1/2}^1 \left(\sum_{m=0}^{\infty} \binom{2m}{m}(x - x^2)^m\right) \cdot \frac{2x-1}{x^2}\, \mathrm{d}x$$

$$= \int_{1/2}^1 \frac{1}{\sqrt{1 - 4x + 4x^2}} \cdot \frac{2x-1}{x^2}\, \mathrm{d}x = \int_{1/2}^1 \frac{1}{x^2}\, \mathrm{d}x = 1. \tag{19}$$

Note also that a random variable X having the probability distribution defined by (18) does not have a finite first moment: Imitating the approach of (19) leads to

$$\mathbf{E}[X] = \sum_{m=0}^{\infty} m\binom{2m}{m}\int_0^1 x^{m-1}(1-x)^m\varphi(x)\, \mathrm{d}x = \int_{1/2}^1 \frac{2(1-x)}{x(2x-1)^2}\, \mathrm{d}x$$

where the integral on the right-hand side diverges.

---

[1] These are, of course, the Catalan numbers, [1, OEIS A000108].

A more detailed investigation of how $p[\mathcal{T}]$ depends on the tree $\mathcal{T}$ reveals that two trees that only differ in a fringe tree which is rooted far away from the root node will have approximately the same polynomial function $p[\,\cdot\,]$ over $[0,1]$, provided a technical condition holds, see [9, Theorem 25]. Using this, it is possible to verify that the same limiting distribution as in (18) also holds if we consider the sequence of complete binary trees on n vertices (of which the full complete binary trees are merely a subsequence):

Denote by $\mathcal{T}_n$ the complete binary tree on n vertices, this being the binary tree having 2^k vertices at height k for $0 \le k < \lfloor \lg n \rfloor =: m$, with the remaining $n - 2^m + 1$ vertices at height m in their left-most positions.

▶ **Proposition 17.** *With $\varphi(x)$ defined by equation (17) above, we have $p[\mathcal{T}_n](x) \to \varphi(x)$ uniformly over $[0,1]$, as $n \to \infty$.*

Consequently, the limiting distribution of $|\mathcal{T}_{n,\mathfrak{S}}|$ coincides with that of $|\mathrm{CBT}_{n,\mathfrak{S}}|$ and is given by equation (18).

Observe that by this proposition and by (19), the random variables $|\mathcal{T}_{n,\mathfrak{S}}|$ converge in distribution and are therefore tight. By Remark 10 this means that the limit law obtained by Janson in [20, Theorem 1.1] for $\mathfrak{C}(\mathcal{T}_n)$ holds also for $\mathfrak{S}(\mathcal{T}_n)$. More explicitly, if we denote by $\{x\} := x - \lfloor x \rfloor$ the fractional part of $x \in \mathbb{R}$ then we obtain the following

▶ **Corollary 18.** *Let $n \to \infty$ such that $\{\lg n - \lg \lg n\} \to \gamma \in [0,1]$. Write $f(\gamma) := 2^\gamma - 1 - \gamma$. Let W_γ be a random variable with an infinitely divisible distribution with characteristic function*

$$\mathbf{E}\left[e^{itW_\gamma}\right] = \exp\left(if(\gamma)t + \int_0^\infty \left(e^{itx} - 1 - itx\mathbb{1}_{\{x<1\}}\right)\,\mathrm{d}\nu_\gamma(x)\right),$$

with the Lévy measure ν_γ being supported on $[0,\infty)$ and having density $\mathrm{d}\nu_\gamma = 2^{\{\lg x + \gamma\}} x^{-2}\,\mathrm{d}x$. Then

$$\frac{\lg^2 n}{n}\left(\mathfrak{S}(\mathcal{T}_n) - \frac{n}{\lg n} - \frac{n \lg \lg n}{\lg^2 n}\right) \xrightarrow{\mathscr{D}} -W_\gamma$$

6   Further questions

We use this final section to present several, deliberately broad questions or remarks that could lead to interesting future research.

1. Determine the asymptotic distributions of $\mathfrak{S}(G^{(n)})$ and $\left|G_{\mathfrak{S}}^{(n)}\right|$ for other families of deterministic and random trees. The author hopes to answer this for conditioned Galton-Watson trees in a follow-up paper.
2. What happens if the roles of S and T are exchanged? For which graphs and which choices of S, T are the random variables $\mathfrak{S}(G; S, T)$ and $\mathfrak{S}(G; T, S)$ equal in distribution?
3. How to evaluate the asymptotic distribution of $\mathfrak{S}$ directly, without relying on previous knowledge of $\mathfrak{C}$ as in Corollary 18?
4. It is easy to see that the edge-cutting process on a graph G is exactly the vertex-cutting process on the line graph of G. This therefore raises the question: How is the separation time on G related to the separation time on the line graph of G?
5. For which sequences of graphs $G^{(n)}$ exhausting a locally finite infinite G (with fixed sources and targets) are the random variables $\left|G_{\mathfrak{S}}^{(n)}\right|$ *not* tight? In this case, what can be said about the structure of the remaining graph?

6. By definition, the separation number is the number of cuts required to separate two subsets $S, T \subseteq V$ from each other. Starting in a graph with a high connectivity (say, by having $k \geq 2$ vertex-disjoint paths from S to T in G) we can ask for the number of cuts required to reduce the connectivity to some $j \leq k$. The case $j = 0$ specializes to the separation number as we defined it.

References

1 OEIS Foundation Inc. (2021). The on-line encyclopedia of integer sequences. URL: `http://oeis.org`.

2 L. Addario-Berry, N. Broutin, and C. Holmgren. Cutting down trees with a Markov chainsaw. *Annals of Applied Probability*, 24(6):2297–2339, 2014.

3 N. Alon, I. Benjamini, and A. Stacey. Percolation on finite graphs and isoperimetric inequalities. *The Annals of Probability*, 32(3):1727–1745, 2004.

4 J. Bertoin. Fires on trees. In *Annales de l'IHP Probabilités et statistiques*, volume 48, 909–921, 2012.

5 G. Berzunza. The cut-tree of large trees with small heights. *Random Structures & Algorithms*, 51(3):404–427, 2017.

6 G. Berzunza, X. S. Cai, and C. Holmgren. The k-cut model in deterministic and random trees. *arXiv*, 2019. `arXiv:1907.02770`.

7 P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 1st edition, 1968.

8 N. Broutin and M. Wang. Cutting down **p**-trees and inhomogeneous continuum random trees. *Bernoulli*, 23(4A):2380–2433, 2017.

9 F. Burghart. A modification of the random cutting model, 2022. `arXiv:2111.02968`.

10 X. S. Cai and C. Holmgren. Cutting resilient networks-complete binary trees. *The Electronic Journal of Combinatorics*, 26(4), 2019.

11 X. S. Cai, C. Holmgren, L. Devroye, and F. Skerman. k-cut on paths and some trees. *Electronic Journal of Probability*, 24, 2019.

12 V. Campos, V. Chvatal, L. Devroye, and P. Taslakian. Transversals in trees. *Journal of graph theory*, 73(1):32–43, 2013.

13 P. Chassaing and R. Marchand. Merging costs for the additive Marcus-Lushnikov process, and Union-Find algorithms, 2004. `arXiv:math/0406094`.

14 L. Devroye. A note on the probability of cutting a Galton-Watson tree. *Electronic Journal of Probability*, 16:2001–2019, 2011.

15 J. A. Fill, N. Kapur, and A. Panholzer. Destruction of very simple trees. *Algorithmica*, 46(3):345–366, 2006.

16 P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

17 G. R. Grimmett. *Percolation*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer, 2nd edition, 1999.

18 C. Holmgren. Random records and cuttings in binary search trees. *Combinatorics, Probability and Computing*, 19(3):391–424, 2010.

19 C. Holmgren. A weakly 1-stable distribution for the number of random records and cuttings in split trees. *Advances in Applied Probability*, 43(1):151–177, 2011.

20 S. Janson. Random records and cuttings in complete binary trees. In *Mathematics and Computer Science III,* 241–253. Springer, 2004.

21 S. Janson. Random cutting and records in deterministic and random trees. *Random Structures & Algorithms*, 29(2):139–179, 2006.

22 D. E. Knuth and A. Schönhage. The expected linearity of a simple equivalence algorithm. *Theoretical Computer Science*, 6(3):281–315, 1978.

23 M. Kuba and A. Panholzer. Analysis of the total costs for variants of the Union-Find algorithm. In *Discrete Mathematics and Theoretical Computer Science*, pages 283–294. Discrete Mathematics and Theoretical Computer Science, 2007.

24   M. Kuba and A. Panholzer. Isolating a leaf in rooted trees via random cuttings. *Annals of Combinatorics*, 12(1):81–99, 2008.

25   M. Kuba and A. Panholzer. Isolating nodes in recursive trees. *Aequationes mathematicae*, 76(3):258–280, 2008.

26   M. Kuba and A. Panholzer. Multiple isolation of nodes in recursive trees. *Online Journal of Analytic Combinatorics*, 9(7), 2014.

27   R. Lyons and Y. Peres. *Probability on trees and networks*. Cambridge University Press, 2017.

28   A. Meir and J. W. Moon. Cutting down random trees. *Journal of the Australian Mathematical Society*, 11(3):313–324, 1970.

29   A. Panholzer. Non-crossing trees revisited: cutting down and spanning subtrees. In *Discrete Mathematics and Theoretical Computer Science,* 265–276, 2003.

30   A. Panholzer. Cutting down very simple trees. *Quaestiones Mathematicae*, 29(2):211–227, 2006.

31   A. C.-C. Yao. On the average behavior of set merging algorithms. In *Proceedings of the eighth annual ACM symposium on Theory of computing,* 192–195, 1976.