



A MODIFICATION OF THE RANDOM CUTTING MODEL

FABIAN BURGHART,* *Uppsala University*

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 (*J. Austral. Math. Soc.* **11**, 1970) 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 for binary caterpillar trees and complete binary trees.

Keywords: Random cutting model; random separation of graphs; percolation

2020 Mathematics Subject Classification: Primary 60C05

Secondary 05C30

1. Introduction

The behaviour of trees from which vertices are randomly removed was first investigated by Meir and Moon in [23]. The process in question starts with a rooted tree, where at every step a uniformly chosen vertex (or edge, but for the purposes of this paper we will focus on vertex-removals) 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 (or isolated), the question of interest is the random number of cuts needed to reach this state. In [23], the expected value and the variance of this random variable for a random labelled tree is found.

Panholzer [25, 26] used generating functions to determine the asymptotic distribution of the cutting number for non-crossing trees and for certain families of simply generated trees. Janson [18] generalised these results, obtaining the asymptotic distribution for conditioned Galton–Watson trees, by approximating the cutting number with a sum of independent random variables (see also the alternative proof in [2]). Using a similar strategy, a limit law for complete binary trees is proven in [17]. This work was later extended by Holmgren to binary search trees [15] and split trees [16].

The random cutting model relates to both the record problem (as was observed in [18]) and to fragmentation processes, where the genealogy of the procedure gives rise to the so-called cut tree, which has become both a useful tool in obtaining results on the cutting number and an object of independent interest. See e.g. [4], [8], or [5].

Received 16 May 2022; accepted 4 April 2023.

* Postal address: Department of Mathematics, Uppsala University, S-751 06 Uppsala, Sweden. Email address: fabian.burghart@math.uu.se

© The Author(s), 2023. Published by Cambridge University Press on behalf of Applied Probability Trust.

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 [19–21], and Cai *et al.* proposed and investigated the k -cut model, where a node is only removed after it has been cut for the k th time, in [6, 9, 10].

In this paper we introduce a different modification of the cutting model, where, in addition to one or several root vertices (which we will call sources), a second set of vertices (targets) is 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). We can then ask several natural questions, about the number of cuts necessary to separate the two sets, about which vertex was the last to be removed, and about the size of the remaining graph. The aim of the paper is to investigate how separation interpolates between the cutting model and site percolation, how these questions relate to each other, and what general properties hold for this modification of the cutting model. We also give some examples.

Plan of the paper

In Section 2, we will fix some notation, define our model in both continuous- and discrete-time settings, and present a first example.

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 3 and 4). Since the latter is commonly defined on a locally finite infinite graph, this requires the right definition (Definition 2) that enables us to approximate such a graph by finite graphs, all while respecting the choice of sources and targets.

Section 4 begins by determining the probability that a fixed subgraph occurs at some point in the cutting procedure, which is then used to obtain the probability that this subgraph is the remaining part at separation. This leads to Theorem 1, which could be regarded as the main result of the paper; it 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.

In Section 5, our scope will be limited to rooted trees, since their recursive nature can be used to simplify many of the arguments and calculations. Here, we will investigate a relevant polynomial that already came up (in a slightly different form) in the works of Devroye *et al.* [11, 12].

Section 6 will contain three examples of rooted trees: binary caterpillars, star-shaped graphs, and complete binary trees. These examples are selected to illustrate the techniques of the earlier sections (and the limitations thereof). Finally, in Section 7 we will give a brief compilation of further research questions.

2. Cutting procedures

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. If the graph in question is a tree, we prefer the symbol \mathcal{T} over G . Further notation for special graphs will be introduced as required. 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 $\text{dist}(v, w)$ for the graph distance between vertices v, w . In the case where $A, B \subseteq V(G)$ are subsets, $\text{dist}(A, B)$ is to be understood as $\min\{\text{dist}(v, w) : v \in A, w \in B\}$.

Given any set $A \subseteq V(G)$ and a fixed set S of source nodes, we define the closure of A to be

$$\text{clos}_S(A) := A \cup S \cup \{v \in V(G) : v \sim w \text{ for some } w \in A\}.$$

The (exterior) boundary of A is defined as $\partial_S A := \text{clos}_S(A) \setminus A$. 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. $\partial_S \emptyset := S$. (Attributing this special role to S in this definition turns out to be useful throughout the paper).

We use the symbols Γ and B to denote the gamma and beta functions, respectively, and will make use of the identities

$$B(x, y) := \int_0^1 u^{x-1}(1-u)^{y-1} du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

and

$$B(n+1, m+1) = \frac{1}{n+m+1} \binom{n+m}{n}^{-1} = \frac{1}{n+1} \binom{n+m+1}{n+1}^{-1}$$

for $x, y > 0$ and $n, m \in \mathbb{N}$ (where, in the last expression, the roles of n and m can be reversed thanks to the symmetry of the beta function); cf. [1, Chapter 6].

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 keep only the components containing sources.
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 random sequence of induced subgraphs

$$G =: G_0 \supseteq G_1 \supseteq \dots \supseteq G_{r-1} \supseteq G_r = \emptyset, \tag{1}$$

where, with a slight abuse of notation, we denote the empty subgraph consisting of no vertices by \emptyset . Observe that therefore we can view the cutting procedure as a discrete-time stochastic process on the state space consisting of subgraphs of G , and as such it is in fact a Markov chain. We will denote this process by $\text{Cut}(G)$.

Instead of describing the cutting procedure by the sequence of graphs we obtain, as in (1), we can equivalently keep track of the removed vertices, v_1, \dots, v_r . We will use both styles of bookkeeping interchangeably, depending on which one is more suitable for the task at hand.

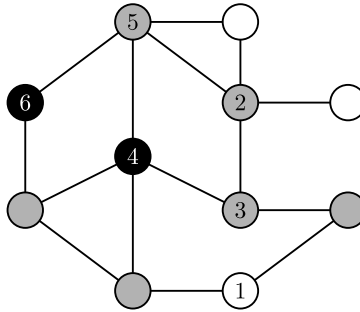


FIGURE 1. A graph G , with source nodes in black and target nodes in white. The number i at a vertex indicates that this vertex was chosen by the i th cut. Note that cuts 2 and 3 each remove an additional vertex. In this example, $\mathfrak{S}(G) = 5$, $\mathfrak{C}(G) = 6$, and $G_{\mathfrak{S}}$ is a path on three vertices, beginning with the source labelled 6.

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 (see Figure 1 for an example):

- 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 contains no more target nodes (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.
- The separation node $v_{\mathfrak{S}}$, denoting the last node that was removed before separation occurred.

The continuous-time model

As has previously been observed by [18] 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, if the corresponding node is still in the graph at that time, then 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 independent and identically distributed (i.i.d.) family of $\text{Exp}(1)$ -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 $\text{Cut}^c(G)$. 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 $\text{Cut}^c(G)$ immediately before time t —so $G_{t-} = G_t$ if and only if 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

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

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

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

respectively.

A first example

Denote by P_n for $n \geq 0$ a path on $2n + 1$ vertices. Let the middle vertex be the unique source of P_n , and let T be the two leaves.

Proposition 1. *Let $X \sim \text{Poi}(\ln 2)$, and let \mathbf{P} be the distribution of $1 + X$ conditioned on X being at least 1. Then $\mathfrak{S}[P_n] \xrightarrow{\mathcal{D}} \mathbf{P}$. In other words,*

$$\mathbb{P}[\mathfrak{S}[P_n] \leq k] \rightarrow \mathbb{P}[X \leq k - 1 | X \geq 1] = \sum_{j=1}^{k-1} \frac{(\ln 2)^j}{j!}$$

for all $k \geq 0$.

Proof. Embed $P_n, n \geq 1$, in $[0,1]$ by mapping the $2n + 1$ vertices to equidistant points $p_{n,i} = \frac{i}{2n}$ for $i = 0, 1, \dots, 2n$ such that neighbouring vertices have a distance of exactly $\frac{1}{2n}$. Note that independently of n , the root vertex will always be at $1/2$ and the leaves at 0 and 1 .

Given that a cut is chosen uniformly among the discrete vertices of the component containing the source, the limit of the cut position as n tends to infinity will be uniformly drawn from an interval containing $1/2$ under the above embedding, which follows from Theorem 7.8 in [7]. Thus, in the limit, we can investigate a space-continuous cutting model: we begin with the interval $[0,1]$, then choose a point from the interval uniformly at random, and consider the component containing $1/2$. Now we repeat this procedure, where the randomness in every step is only dependent on the current interval. In analogy to the discrete cutting number, we can define $\mathfrak{S}[0, 1]$ to be the time of the first cut where the remaining interval is of the form (a, b) with $0 < a < b < 1$. By this construction, we have $\mathfrak{S}[P_n] \xrightarrow{\mathcal{D}} \mathfrak{S}[0, 1]$.

In this continuous setting, $\mathfrak{S}[0, 1] = k$ for $k \geq 2$ means that the first $k - 1$ cuts all happened on the same side of $1/2$, i.e. in $(1/2, 1)$ or in $[0, 1/2)$, and the k th cut took place on the other side. Since the entire setup is symmetric with respect to $1/2$, assume without loss of generality that the first cut happens inside $[1/2, 1]$. The probability for $\{\mathfrak{S}[0, 1] = k\}$ to happen is thus expressed by the following iterated integral, where c_1, c_2, \dots denote the first cut, the second cut, etc.:

$$\mathbb{P}[\mathfrak{S}[0, 1] = k] = 2 \int_{1/2}^1 \int_{1/2}^{c_1} \dots \int_{1/2}^{c_{k-2}} \frac{1/2}{c_{k-1}} \cdot \frac{dc_{k-1}}{c_{k-2}} \dots \frac{dc_2}{c_1} dc_1. \tag{2}$$

Here, the innermost integrand $1/(2c_{k-1})$ is the probability of c_k falling into the interval $[0, 1/2]$ given that the last cut was c_{k-1} , or equivalently, given that the interval for c_k to choose from was $[0, c_{k-1})$. The nested integrals in (2) then arise from repeatedly applying the law of total probability and can be evaluated iteratively to yield $\frac{\ln(2)^{k-1}}{(k-1)!}$, as required. \square

This result is in stark contrast with $\mathfrak{C}(P_n)$, for which it is known that

$$\frac{\mathfrak{C}(P_n) - 2 \ln n}{\sqrt{2 \ln n}} \xrightarrow{\mathcal{D}} N(0, 1)$$

as $n \rightarrow \infty$ (cf. e.g. [18, Example 8.1] and references therein).

3. Monotonicity, cutting, and site percolation

The purpose of this section is to state and prove some basic properties concerning the behaviour of the separation time when changes are made to the underlying graph or the selection of source/target nodes. For this purpose, we will write $\mathfrak{S}(G; S, T)$ instead of $\mathfrak{S}(G)$ whenever there is potential ambiguity over the choice of source and target nodes. The following definition will prove useful.

Definition 1. Let $G = (V, E)$ be a finite graph, and let $S, T \subseteq V$ be distinguished sets of sources and targets. A vertex $v \in V$ is called *essential* if G contains a path $sv_1 \dots v_k t$ containing v , such that s and t are the unique vertices on this path that are contained in S and T , respectively. Call a vertex *non-essential* if no such path exists.

Observe that it is possible both for source and target nodes to be non-essential—this is the case precisely if every relevant path passes through another source or target node.

If two graphs G, G' on a common vertex set with respective source and target vertices $S, T, S', T' \subseteq V$ are given, and if they are such that any sequence of removed vertices v_1, \dots, v_r that separates S from T in G also separates S' from T' in G' , then $\mathfrak{S}(G'; S', T') \leq \mathfrak{S}(G; S, T)$ and $\mathfrak{S}_c(G'; S', T') \leq \mathfrak{S}_c(G; S, T)$ by using the same randomness for the cutting model on G' as on G . This argument proves each claim in the next proposition.

Proposition 2. Let $G = (V, E)$ be a finite connected graph, and let $S, T \subseteq V$ be the sets of source and target nodes, respectively. The following are true:

1. If G' is a subgraph of G endowed with source and target nodes inherited from G , then $\mathfrak{S}(G') \leq \mathfrak{S}(G)$.
2. If S and T are replaced by subsets $S' \subseteq S$ and $T' \subseteq T$ respectively, then $\mathfrak{S}(G; S', T') \leq \mathfrak{S}(G; S, T)$.
3. If T is replaced by a set T' of essential vertices such that every path from S to T intersects T' , then $\mathfrak{S}(G; S, T) \leq \mathfrak{S}(G; S, T')$.
4. If S is replaced by a set S' of essential vertices such that every path from S to T intersects S' , then $\mathfrak{S}(G; S, T) \leq \mathfrak{S}(G; S', T)$.

All four claims remain correct if \mathfrak{S} is replaced by its continuous-time counterpart \mathfrak{S}_c .

Remark 1. It follows from (i) in Proposition 2 that one has the following chain of inequalities for any connected graph $G = (V, E)$ with $S, T \subseteq V$:

$$\mathfrak{S}(\mathcal{T}) \leq \mathfrak{S}(G) \leq \mathfrak{S}(K_V), \tag{3}$$

where \mathcal{T} is a spanning tree of G and K_V is the complete graph on the vertex set V . Observe that for some graphs, both the upper and lower estimates may hold with equality. Indeed, assume K_V has a single source denoted as s , and set \mathcal{T} to be the spanning tree of K_V that is a star graph

centred on s . Running the cutting procedure on either K_V or \mathcal{T} will amount to removing one vertex at a time until the last source vertex is removed, thus showing equality.

It should also be noted that the continuous-time separation time is much more stable under changes to G , as demonstrated by the following lemma.

Lemma 1. *Let $G = (V, E)$ and $S, T \subseteq V$ as above. Denote by G' the graph obtained by removing all non-essential vertices from G . Then $\mathfrak{S}_c[G] = \mathfrak{S}_c[G']$ deterministically.*

Proof. Denote by $G = G_0, G_1, G_2, \dots$ and $G' = G'_0, G'_1, G'_2, \dots$ the continuous-time cutting procedures on G and G' , respectively. As $V(G') \subseteq V(G)$, we can use the same randomness for X_v in both graphs when $v \in V(G')$. Hence we certainly have $\mathfrak{S}_c[G] \geq \mathfrak{S}_c[G']$. Now assume that at time $t \geq \mathfrak{S}_c[G']$, there still is a path $sv_1 \dots v_k t$ connecting S and T in G_t (thus necessarily using non-essential nodes). This path must contain vertices v_i, v_j with $i \leq j$ such that $v_i \in S, v_j \in T$, and none of the vertices on the path between v_i and v_j are in either S or T . However, by definition all vertices v_i, v_{i+1}, \dots, v_j are essential and thus in G' , contradicting the assumption that separation in G' has already occurred. \square

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 3. *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 these statements also hold true for $\mathfrak{S}_c(G)$ and $\mathfrak{C}_c(G)$ in the continuous-time model.*

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 $\mathbb{E}[\mathfrak{C}(G)] - \mathbb{E}[\mathfrak{S}(G)] \geq p_0$, 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

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

since $X_v \sim \text{Exp}(1)$ is memoryless. \square

Separation and site percolation

In this section, we show that in a certain sense, the continuous-time separation model on an infinite graph G with infinite distance $\text{dist}(S, T)$ 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 $\text{perc}_S(p)$ the probability that the $\text{Ber}(p)$ -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:

1. The $G^{(n)}$ are connected subgraphs satisfying

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

2. 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 $T^{(n)} := (T \cap V(G^{(n)})) \cup \partial_{\emptyset}(G \setminus G^{(n)})$.

Observe that the set of target nodes $T^{(n)}$ is indeed a subset of $V(G^{(n)})$ and will be non-empty even if $T = \emptyset$. Moreover, Condition 2 necessitates that S is finite.

Proposition 4. *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

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\mathfrak{S}_c(G^{(n)}) \geq x \right] = \text{perc}_S(e^{-x}). \tag{4}$$

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

Recall next that independently removing each vertex v of G at a random time $X_v \sim \text{Exp}(1)$ gives rise to the monotonous coupling of Bernoulli site percolation for all parameters $p \in [0, 1]$ (cf. [22, p. 138]). Indeed, at time $x \in [0, \infty]$, the graph we observe in this way is a sample of $\text{Ber}(e^{-x})$ -site percolation on G . We can couple the process thus obtained 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 (4), assume that $\text{Ber}(e^{-x})$ -site percolation exhibits an infinite cluster which intersects S ; such a 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, $\text{perc}_S(e^{-x}) \leq \mathbb{P}[\mathfrak{S}_c(G^{(n)}) \geq x]$, and letting n tend to ∞ yields $\liminf_{n \rightarrow \infty} \mathbb{P}[\mathfrak{S}_c(G^{(n)}) \geq x] \geq \text{perc}_S(e^{-x})$.

For the other inequality, suppose now that $\text{Ber}(e^{-x})$ -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 $\text{dist}(S, T^{(n)}) > k$ 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

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

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

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

which implies the existence of the limit and \leq in (4) after passing to the limit for $k \rightarrow \infty$ 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 $\text{Cut}(G)$ will produce a specific subgraph G_* .

Lemma 2. 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

$$\mathbb{P} [G_t \supseteq G_*] = e^{-|G_*|t} \tag{5}$$

and

$$\mathbb{P} [G_t = G_*] = e^{-|G_*|t} (1 - e^{-t})^{|\partial_S G_*|} . \tag{6}$$

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

$$\mathbb{P} [G_{X_{v_*}} = G_* \mid X_{v_*}] = e^{-|G_*|X_{v_*}} (1 - e^{-X_{v_*}})^{|\partial_S G_*|-1} . \tag{7}$$

Proof. The random subgraph G_t contains G_* if and only if at time t , none of the clocks X_v for $v \in V(G_*)$ have rung yet and every component is still attached to a source node. Hence, as the X_v are independent $\text{Exp}(1)$ -distributed random variables, we obtain

$$\mathbb{P} [G_t \supseteq G_*] = \prod_{v \in V(G_*)} \mathbb{P}[X_v \geq t] = e^{-|G_*|t} .$$

Similarly, $G_t = G_*$ if and only if at time t , none of the clocks for $v \in V(G_*)$, but all of the clocks for $v \in \partial_S G_*$ (this includes the nodes in S which are not in G_* , by the definition of ∂_S in Section 2), have rung, which yields (6).

For the final claim, note that at time X_{v_*} , all nodes in G_* must still be intact, whereas all nodes in $\partial_S G_* \setminus \{v_*\}$ must already have been cut. Now proceed as in the case for a deterministic time. \square

Of course, Equations (5) and (6) in the previous proposition are equivalent. One can easily obtain (5) from (6) by summing over all graphs that contain G_* . For the other direction, applying the inclusion–exclusion principle suffices.

Corollary 1. 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

$$\mathbb{P} [\exists i \geq 1 : G_i = G_* \text{ and } v_i = v_*] = \frac{1}{|\partial_S G_*|} \binom{|G_*| + |\partial_S G_*|}{|\partial_S G_*|}^{-1} \tag{8}$$

and therefore

$$\mathbb{P} [\exists i \geq 1 : G_i = G_*] = \binom{|G_*| + |\partial_S G_*|}{|\partial_S G_*|}^{-1} . \tag{9}$$

Proof. We prove the result by considering the continuous-time model instead. There, the event $B(v_*) := \{\exists i \geq 1 : G_i = G_* \text{ and } v_i = v_*\}$ translates to the event $\{G_{X_{v_*}} = G_*\}$. Hence, we obtain the first result from integrating over $X_{v_*} \sim \text{Exp}(1)$ in (7) by substituting $e^{-x} = u$:

$$\begin{aligned} \mathbb{P}[B(v_*)] &= \int_0^\infty (e^{-x})^{|G_*|} (1 - e^{-x})^{|\partial_S G_*|-1} \cdot e^{-x} dx \\ &= \int_0^1 u^{|G_*|} (1 - u)^{|\partial_S G_*|-1} du \\ &= B(|G_*| + 1, |\partial_S G_*|) \\ &= \frac{1}{|\partial_S G_*|} \binom{|G_*| + |\partial_S G_*|}{|\partial_S G_*|}^{-1}, \end{aligned}$$

as required. Finally, the observation that

$$\{\exists i \geq 1 : G_i = G_*\} = \bigsqcup_{v_* \in \partial G_*} B(v_*),$$

where the right-hand side denotes the disjoint union over the events $B(v_*)$, lets us obtain (9) from (8). □

Definition 3. 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 $\mathbb{P}[G_\ominus = G_*]$ is positive. We denote by $\mathcal{A}_m(G)$ the set of all admissible subgraphs G_* of G of size $|V(G_*)| = m$.

Lemma 3. Assume $T \neq \emptyset$. An induced subgraph $G_* \subseteq G$ is admissible if and only if G_* contains no target nodes and every component of G_* contains at least one source node.

Proof. That the stated conditions for G_* are necessary for G_* to be admissible is evident from the definitions of the cutting procedure and the separation time. To show that they are also sufficient, label the vertices in $\partial_S G_*$ by v_1, \dots, v_r in such a way that after the removal of v_1, \dots, v_{r-1} from G , the remaining graph still contains a path connecting S and T (such a boundary vertex exists since there must be at least one such path in G , and it cannot be contained entirely in G_* by assumption). Then one way of realising the event $\{G_\ominus = G_*\}$ is by removing the vertices v_1, \dots, v_r in this order. Hence, assuming $|V(G)| = n$, we have

$$\mathbb{P}[G_\ominus = G_*] \geq \prod_{i=1}^r \frac{1}{n + 1 - i} > 0,$$

thus concluding the proof. □

Using similar methods as in the proofs of Lemma 2 and Corollary 1, we can establish the following connection between the graph G_\ominus at separation and the continuous-time separation number \mathfrak{S}_c .

Proposition 5. Let G_* be an admissible subgraph of G . Then

$$\mathbb{P}[G_\ominus = G_*] = \sum_{v_* \in \partial_S G_*} \int_0^1 u^{|G_*|-1} (1 - u)^{|\partial_S G_*|-1} \mathbb{P}[\mathfrak{S}_c(H[v_*]) \geq -\ln u] du, \quad (10)$$

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 $\text{Cut}(G)$ 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 did not pass 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

$$\begin{aligned} \mathbb{P}[G_{\mathfrak{S}} = G_*, v_{\mathfrak{S}} = v_* \mid X_{v_*}] &= \mathbb{P}\left[G_{X_{v_*}} = G_* \text{ and } v_* \text{ connects to } T \text{ in } G_{X_{v_*}^-} \mid X_{v_*}\right] \\ &= \mathbb{P}\left[G_{X_{v_*}} = G_* \mid X_{v_*}\right] \mathbb{P}\left[\mathfrak{S}_c(H[v_*]) \geq X_{v_*} \mid X_{v_*}\right], \end{aligned} \tag{11}$$

where conditional independence holds true because

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

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,

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

In light of (7) from Lemma 2, we can now rewrite Equation (11) as

$$\begin{aligned} \mathbb{P}[G_{\mathfrak{S}} = G_*, v_{\mathfrak{S}} = v_* \mid X_{v_*} = x] &= e^{-|G_*|x} (1 - e^{-x})^{|\partial_S G_*|-1} \frac{\mathbb{P}\left[\mathfrak{S}_c(H[v_*]) \geq x\right]}{e^{-x}}. \end{aligned} \tag{12}$$

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

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

so that, after plugging in the expression from (12) and using $X_{v_*} \sim \text{Exp}(1)$, we obtain

$$\begin{aligned} \mathbb{P}[G_{\mathfrak{S}} = G_*] &= \sum_{v_* \in \partial_S G_*} \int_0^\infty (e^{-x})^{|G_*|} (1 - e^{-x})^{|\partial_S G_*|-1} \frac{\mathbb{P}\left[\mathfrak{S}_c(H[v_*]) \geq x\right]}{e^{-x}} \cdot e^{-x} dx, \end{aligned}$$

which only differs from (10) by the substitution $e^{-x} = u$. □

Remark 2. The (unconditioned) probability $\mathbb{P}[\mathfrak{S}_c(H[v_*]) \geq X_{v_*}]$ has a number of equivalent versions. Indeed, if we consider any G with $S = \{v_*\}$ and arbitrary T , then we have the following equalities:

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

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

$$\mathbb{P}[\mathfrak{S}_c(H[v_*]) \geq -\ln u] \leq \mathbb{P}[X_{v_*} \geq -\ln u] = u \quad \forall u \in [0, 1]. \tag{13}$$

Proposition 5 enables us to make the upper bound $\mathfrak{S}(K_V)$ from (3) more explicit.

Corollary 2. *Let K_V be a complete graph on a vertex set V of cardinality $|V| = n$ with sources S and targets T . Then*

$$\mathbb{P}[|K_{V,\mathfrak{S}}| = m] = \frac{|T|}{m+1} \binom{n}{m+1}^{-1} \left[\binom{n-|T|}{m} - \binom{n-|S \cup T|}{m} \right], \tag{14}$$

for $m = 1, \dots, n$. Furthermore, for times $t = 1, \dots, n - 1$,

$$\begin{aligned} \mathbb{P}[\mathfrak{S}(K_V) = t] &= \frac{1}{n-t+1} \binom{n}{n-t+1}^{-1} \\ &\cdot \left[|T| \binom{n-|T|}{n-t} + |S| \binom{n-|S|}{n-t} - |S \cup T| \binom{n-|S \cup T|}{n-t} \right]. \end{aligned} \tag{15}$$

Proof. According to Proposition 5, we have

$$\begin{aligned} \mathbb{P}[|K_{V,\mathfrak{S}}| = m] &= \sum_{G_* \in \mathcal{A}_m(K_V)} \sum_{v_* \in \partial_S G_*} \int_0^1 u^{m-1} (1-u)^{|\partial_S G_*|-1} \mathbb{P}[\mathfrak{S}_c(H[v_*]) \geq -\ln u] \, du, \end{aligned}$$

where $\mathcal{A}_m(K_V)$ denotes the set of admissible subgraphs of K_V on m vertices, as in Definition 3. For $m \geq 1$, this can be evaluated: $\partial_S G_*$ always consists of all the vertices not in G_* , so $|\partial_S G_*| = n - m$, and since the graphs $H[v_*]$ only contain the vertex v_* , we obtain

$$\mathbb{P}[\mathfrak{S}_c(H[v_*]) \geq -\ln u] = \begin{cases} 0 & \text{if } v_* \notin T, \\ u & \text{if } v_* \in T. \end{cases}$$

Since at separation all targets must have been removed, this yields

$$\begin{aligned} \mathbb{P}[|K_{V,\mathfrak{S}}| = m] &= \sum_{G_* \in \mathcal{A}_m(K_V)} |T| \int_0^1 u^m (1-u)^{n-m-1} \, du \\ &= |\mathcal{A}_m(K_V)| \cdot |T| \cdot B(m+1, n-m). \end{aligned}$$

Finally, by Lemma 3, the graphs in $\mathcal{A}_m(K_V)$ consist of m vertices, none of which are targets, unless the m vertices are chosen from $V \setminus (S \cup T)$. Hence

$$|\mathcal{A}_m(K_V)| = \binom{n - |T|}{m} - \binom{n - |S \cup T|}{m}$$

and (14) follows.

For the second part, we first show that

$$\mathbb{P}[\mathfrak{S}(K_V) = t] = \mathbb{P}[|K_{V,\mathfrak{S}}| = n - t] + \frac{g}{n - t + 1} \binom{n}{t - 1}^{-1} \tag{16}$$

for some explicit constant g . Observe that separation at time t must occur after the cutting of t vertices, either leaving behind some non-empty graph $K_{V,\mathfrak{S}}$ (in which case $|K_{V,\mathfrak{S}}| = n - t$, as the vertices are removed one at a time on K_V —this case is thus covered by the first summand on the right-hand side of (16)), or leaving behind an empty graph $K_{V,\mathfrak{S}}$. This second case occurs if and only if $K_{V,t-1}$ contains exactly one source (which will then have to be cut at time t) and at least one target. For the sake of this proof, denote the set of these subgraphs by \mathcal{G}_t . Hence

$$\mathbb{P}[\mathfrak{S}(K_V) = t, K_{V,\mathfrak{S}} = \emptyset] = \sum_{G_* \in \mathcal{G}_t} \frac{\mathbb{P}[K_{V,t-1} = G_*]}{|G_*|}.$$

Observe that all $G_* \in \mathcal{G}_t$ have exactly $n - t + 1$ vertices, and that if such a graph is obtained in the cutting procedure, it has to happen at time $t - 1$. Thus, invoking Equation (9) from Corollary 1, we obtain

$$\mathbb{P}[K_{V,t-1} = G_*] = \mathbb{P}[\exists i : K_{V,i} = G_*] = \binom{n}{t - 1}^{-1}$$

and therefore

$$\mathbb{P}[\mathfrak{S}(K_V) = t, K_{V,\mathfrak{S}} = \emptyset] = \frac{|\mathcal{G}_t|}{n - t + 1} \binom{n}{t - 1}^{-1}.$$

This accounts for the second summand in (16), with $g = |\mathcal{G}_t|$. To obtain Equation (15), we note that to build an element of \mathcal{G}_t , one needs to choose one source vertex and $n - t$ non-source vertices, unless neither of these vertices is a target. Hence,

$$|\mathcal{G}_t| = |S| \binom{n - |S|}{n - t} - |S \setminus T| \binom{n - |S \cup T|}{n - t}. \tag{17}$$

Now, plugging Equations (14) and (17) into Equation (16) yields (15) after some minor simplifications. □

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 $\mathbb{P}[|X_i| \geq M] < \varepsilon$ for all $i \in I$; cf. [7, p. 37]. We conclude this section by showing the following tightness result for the size of $G_{\mathfrak{S}}$.

Theorem 1. *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 $a_{m,n} = \sum_{A \in \mathcal{A}_m(G^{(n)})} |\partial_S A|$ and let $a_m := \lim_{n \rightarrow \infty} a_{m,n}$. Assume that there are constants $b > 0$ and $L \geq 0$ such that*

- (i) $|\partial_S A| \geq bm + 1$ for all $A \in \mathcal{A}_m(G^{(n)})$ 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

$$\int_0^1 f(x(1-x)^b) dx < \infty. \tag{18}$$

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

Observe that, since

$$c := \max_{x \in [0,1]} x(1-x)^b = \frac{b^b}{(b+1)^{b+1}},$$

the assumption (ii) requires that the radius of convergence, r , of the power series f is at least c . In the case $r > c$, this condition is trivially satisfied as the integrand is bounded. However, in the case of equality $r = c$, f has a singularity at c by virtue of Pringsheim’s theorem (cf. [13, Theorem IV.6]), and (18) is a non-trivial requirement. We also remark that the assumption (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 non-decreasing and eventually constant. Recall from Definition 2 that the sets $T^{(n)}$ consist of two parts, namely $T \cap V(G^{(n)})$ and vertices that are incident to edges leaving $G^{(n)}$. By copying the respective argument from the proof of Proposition 4, we can show again that $\text{dist}(S, T^{(n)} \setminus T) \rightarrow \infty$ as $n \rightarrow \infty$. 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 monotonically 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 non-decreasing in n . In particular, the limit a_m exists, is finite, and is an upper bound to $a_{m,n}$ for all n .

Let $M \geq L$. We now apply Proposition 5, where we set $p_{v_*}(x) := \mathbb{P}[\mathfrak{S}_c(H[v_*]) \geq -\ln x]$ (note that this still depends on $G_\mathfrak{S}$ as well!) for brevity. Summing over all $m \geq M$ and all $A \in \mathcal{A}_m(G^{(n)})$ yields

$$\begin{aligned} \mathbb{P}\left[|G_\mathfrak{S}^{(n)}| \geq M\right] &= \sum_{m=M}^\infty \sum_{A \in \mathcal{A}_m(G^{(n)})} \int_0^1 x^{m-1} (1-x)^{|\partial_S A|-1} \sum_{v_* \in \partial_S A} p_{v_*}(x) dx \\ &\leq \sum_{m=M}^\infty \sum_{A \in \mathcal{A}_m(G^{(n)})} |\partial_S A| \int_0^1 x^m (1-x)^{|\partial_S A|-1} dx, \end{aligned}$$

where we have applied the estimate (13). Using the assumption (i), we get

$$\begin{aligned} \mathbb{P}\left[|G_\mathfrak{S}^{(n)}| \geq M\right] &\leq \sum_{m=M}^\infty \sum_{A \in \mathcal{A}_m(G^{(n)})} |\partial_S A| \int_0^1 (x(1-x)^b)^m dx \\ &= \sum_{m=M}^\infty \int_0^1 a_{m,n} (x(1-x)^b)^m dx. \end{aligned}$$

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

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

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

$$\begin{aligned} \sum_{m=M}^{\infty} \int_0^1 a_m(x(1-x)^b)^m dx &= \int_0^1 \sum_{m=M}^{\infty} a_m(x(1-x)^b)^m dx \\ &\leq \int_0^1 \sum_{m=L}^{\infty} a_m(x(1-x)^b)^m dx = \int_0^1 f(x(1-x)^b) dx, \end{aligned}$$

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

Remark 3. Tightness of the sizes of the separation graphs $|G_{\mathfrak{S}}^{(n)}|$ is a helpful property for translating limit laws from the cutting times $\mathfrak{C}(G^{(n)})$ to the separation times $\mathfrak{S}(G^{(n)})$. Assume that there exist sequences α_n and $\beta_n > 0$ such that

$$\frac{\mathfrak{C}(G^{(n)}) - \alpha_n}{\beta_n} \xrightarrow{\mathcal{D}} X \quad \text{as } n \rightarrow \infty$$

for a random variable X with positive variance. If $\beta_n \rightarrow \infty$ and $|G_{\mathfrak{S}}^{(n)}|$ forms a tight family of random variables, then the deterministic estimate

$$\mathfrak{S}(G^{(n)}) \leq \mathfrak{C}(G^{(n)}) \leq \mathfrak{S}(G^{(n)}) + |G_{\mathfrak{S}}^{(n)}|$$

implies

$$\left| \frac{\mathfrak{C}(G^{(n)}) - \alpha_n}{\beta_n} - \frac{\mathfrak{S}(G^{(n)}) - \alpha_n}{\beta_n} \right| \leq \frac{|G_{\mathfrak{S}}^{(n)}|}{\beta_n} \rightarrow 0$$

in probability, and hence

$$\frac{\mathfrak{S}(G^{(n)}) - \alpha_n}{\beta_n} \xrightarrow{\mathcal{D}} X \quad \text{as } n \rightarrow \infty.$$

5. Separating trees

As we saw in Lemma 2, the expressions for probabilities of events occurring in Cut_c happen to be polynomial expressions in e^{-x} . This is not surprising: if the event E is dependent only on the subgraph at a time $x \geq 0$, then

$$\mathbb{P}[E] = \sum_{G_* \in E} \mathbb{P}[G_x = G_*],$$

where the summation ranges over all subgraphs $G_* \subseteq G$ in E , and from Equation (6), this is polynomial in e^{-x} . As it turns out, this is especially useful for rooted trees, where the recursive structure of the tree translates to a recursion for said polynomial expression.

Hence, 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, \dots, w_r for $r \geq 1$. Then define

$$p[w](x) := x \left(1 - \prod_{i=1}^r (1 - p[w_i](x)) \right). \tag{19}$$

Observe that in the case where w has only 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 6. *For the continuous-time cutting model on a rooted tree \mathcal{T} , we have*

$$\mathbb{P}[\mathfrak{S}_c(\mathcal{T}) \geq x] = p[\mathcal{T}](e^{-x}) \tag{20}$$

for all $x \geq 0$. Equivalently, one can interpret $p[\mathcal{T}](q)$ for $q \in [0, 1]$ as the probability that $\text{Ber}(q)$ -site percolation on \mathcal{T} contains a path from the root to a leaf.

Proof. For $\mathfrak{S}_c(\mathcal{T})$ to be larger than x , there must exist a path from the root node to a leaf which is still intact at time x . Now, use induction over the height of \mathcal{T} , as follows.

If \mathcal{T} consists of a single node, then $\mathbb{P}[\mathfrak{S}_c(\mathcal{T}) \geq x] = e^{-x} = p[\mathcal{T}](e^{-x})$. Now assume that (20) holds true for all trees up to a certain height $h \geq 1$, and let \mathcal{T} be a rooted tree of height $h + 1$. Denote the children of the root in \mathcal{T} by w_1, \dots, w_r for some $r \geq 1$. Removing the root node creates r trees $\mathcal{T}_1, \dots, \mathcal{T}_r$, which we endow with the new root nodes w_1, \dots, w_r , respectively. As the height of $\mathcal{T}_1, \dots, \mathcal{T}_r$ is at most h , applying the induction hypothesis yields

$$\mathbb{P}[\mathfrak{S}_c(\mathcal{T}_i) \geq x] = p[\mathcal{T}_i](e^{-x}), \quad \text{for } i = 1, \dots, r.$$

The existence of a desired path in \mathcal{T} is equivalent to the root node not yet having been cut, intersected with the event that not all of the subtrees \mathcal{T}_i have yet been separated. Hence,

$$\begin{aligned} \mathbb{P}[\mathfrak{S}_c(\mathcal{T}) \geq x] &= e^{-x} \left(1 - \prod_{i=1}^r \mathbb{P}[\mathfrak{S}_c(\mathcal{T}_i) < x] \right) \\ &= e^{-x} \left(1 - \prod_{i=1}^r (1 - p[\mathcal{T}_i](e^{-x})) \right) = p[\mathcal{T}](e^{-x}) \end{aligned}$$

by (19), which concludes the proof. □

A *transversal* in a rooted tree \mathcal{T} is defined to be a subset of vertices that intersects every path from the root to a leaf. It then follows from the proof of Proposition 6 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 [11, 12].

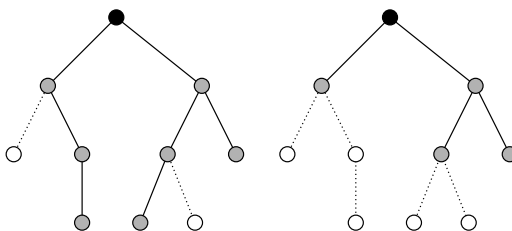


FIGURE 2. A faithful (left) and a trimmed (right) subtree of the same underlying rooted tree \mathcal{T} . The root is shown in black, with vertices belong to the subtrees being coloured grey. Dotted edges and white vertices belong to \mathcal{T} , but not to the subgraph.

Remark 4. 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 (19), $p[\mathcal{T}](x) \leq x$ for $x \in [0, 1]$.

Definition 4. Let \mathcal{T} be a rooted tree.

1. 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} (cf. Figure 2, left). We denote the set of all faithful subtrees of \mathcal{T} by $\mathcal{F}(\mathcal{T})$, and the set of all faithful subtrees on n vertices by $\mathcal{F}_n(\mathcal{T})$.
2. 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 $\mathcal{P}_\ell(\mathcal{T})$.
3. A subtree \mathcal{T}_* is called *trimmed* if it contains the root of \mathcal{T} as its own root and if $\deg_{\mathcal{T}_*}(v) = \deg_{\mathcal{T}}(v)$ for all non-leaves v of \mathcal{T}_* (cf. Figure 2, right).

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

Proposition 7. Write $p[\mathcal{T}](x) = \sum_{j=0}^n a_j x^j$ with $a_j \in \mathbb{Z}$ for $j = 0, 1, \dots, n$. Denote by $L(\mathcal{T}_*)$ the set of leaves of \mathcal{T}_* . Then

$$a_j = \sum_{\mathcal{T}_* \in \mathcal{F}_j(\mathcal{T})} (-1)^{|L(\mathcal{T}_*)|+1} \tag{21}$$

$$= \sum_{\ell \geq 1} (-1)^{\ell+1} |\mathcal{F}_\ell(\mathcal{T}) \cap \mathcal{P}_\ell(\mathcal{T})|. \tag{22}$$

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

Proof. First note that Equation (21) implies the other statements: every subtree of \mathcal{T} contains at most $|\mathcal{T}|$ vertices, and the only subtree with exactly $|\mathcal{T}|$ vertices is \mathcal{T} itself. Hence $a_j = 0$ for $j > |\mathcal{T}|$ and $a_{|\mathcal{T}|} = \pm 1$. On the other hand, any faithful subtree must contain at least one path from the root to $L(\mathcal{T})$, and such a path requires at least $\text{dist}(\text{root}, L(\mathcal{T})) + 1$ vertices.

To show (21), we employ $p[\mathcal{T}](e^{-x}) = \mathbb{P}[\mathfrak{S}_c(\mathcal{T}) \geq x]$ from Proposition 6 and express the probability on the right-hand side in a different manner. To this end, write \mathcal{T}_x for the tree

obtained after time x . As observed already, $\{\mathfrak{S}_c(\mathcal{T}) \geq x\}$ is the same as asking that \mathcal{T}_x contain at least one path from the root to the leaves. Therefore, applying the inclusion–exclusion principle yields

$$\mathbb{P}[\mathfrak{S}_c(\mathcal{T}) \geq x] = \mathbb{P}\left[\bigcup_{\mathcal{T}_* \in \mathcal{P}_1(\mathcal{T})} \{\mathcal{T}_* \subseteq \mathcal{T}_x\}\right] = \sum_{\ell \geq 1} (-1)^{\ell+1} \sum_{\mathcal{T}_* \in \mathcal{P}_\ell(\mathcal{T})} \mathbb{P}[\mathcal{T}_* \subseteq \mathcal{T}_x].$$

Moreover, $\mathbb{P}[\mathcal{T}_* \subseteq \mathcal{T}_x] = (e^{-x})^{|\mathcal{V}(\mathcal{T}_*)|}$; hence

$$p[\mathcal{T}](e^{-x}) = \mathbb{P}[\mathfrak{S}_c(\mathcal{T}) \geq x] = \sum_{\ell \geq 1} (-1)^{\ell+1} \sum_{\mathcal{T}_* \in \mathcal{P}_\ell(\mathcal{T})} (e^{-x})^{|\mathcal{V}(\mathcal{T}_*)|}.$$

Substituting u for e^{-x} and rearranging the sums on the right-hand side, we obtain the desired expressions for the coefficients of $p[\mathcal{T}]$. □

Corollary 3. *We have*

$$\mathbb{P}[G_{\mathfrak{S}} = \emptyset] = \int_0^1 \frac{p[\mathcal{T}](u)}{u} du. \tag{23}$$

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

$$\mathbb{P}[G_{\mathfrak{S}} = \mathcal{T}_*] = \sum_{v_* \in \partial \mathcal{T}_*} \int_0^1 u^{|\mathcal{T}_*|-1} (1-u)^{|\partial \mathcal{T}_*|-1} p[v_*](u) du. \tag{24}$$

Proof. Both statements follow immediately from Propositions 5 and 6. □

Lemma 4. *Let \mathcal{T} be a rooted tree.*

1. *If $\mathcal{T}_* \subseteq \mathcal{T}$ is a faithful subtree then $p[\mathcal{T}_*](x) \leq p[\mathcal{T}](x)$ for all $x \in [0, 1]$.*
2. *If $\mathcal{T}_* \subseteq \mathcal{T}$ is a trimmed subtree then $p[\mathcal{T}_*](x) \geq p[\mathcal{T}](x)$ for all $x \in [0, 1]$.*

Proof. Assume $\mathcal{T}_* \subseteq \mathcal{T}$ is faithful. If, at time $-\ln x$, there exists an uncut path in \mathcal{T}_* connecting the root to $L(\mathcal{T}_*)$, then this same path also exists in \mathcal{T} and connects the root to $L(\mathcal{T})$. Hence we obtain from Proposition 6

$$p[\mathcal{T}_*](x) = \mathbb{P}[\mathfrak{S}_c(\mathcal{T}_*) \geq -\ln x] \leq \mathbb{P}[\mathfrak{S}_c(\mathcal{T}) \geq -\ln x] = p[\mathcal{T}](x).$$

Assume now that $\mathcal{T}_* \subseteq \mathcal{T}$ is trimmed. If there is a path in \mathcal{T} connecting the root to $L(\mathcal{T})$ at time $-\ln x$, then this path also connects the root to $L(\mathcal{T}_*)$. Thus,

$$p[\mathcal{T}_*](x) = \mathbb{P}[\mathfrak{S}_c(\mathcal{T}_*) \geq -\ln x] \geq \mathbb{P}[\mathfrak{S}_c(\mathcal{T}) \geq -\ln x] = p[\mathcal{T}](x),$$

analogously to the first case. □

We conclude this section by looking at a version of the tree polynomial for two variables that turns out to be useful for modifying trees at a fringe node. For a fixed leaf v_* , initialise $p_{v_*}[v_*](x, y) = y$ and all other leaves by x as before. Then recursively define $p_{v_*}[w]$ as in (19), replacing all terms $p[\cdot](x)$ by $p_{v_*}[\cdot](x, y)$. Whereas in the original construction we introduced one variable x for every vertex, we now introduce one variable x for every vertex bar v_* , which gets a y -variable. If the choice of v_* is fixed or irrelevant for the purpose of the argument, we

will drop the index and simply write $p[w](x, y)$. Analogously to before, we extend the notation to allow polynomials to be assigned to fringe subtrees.

Lemma 5. *In the setting just introduced, $p[\mathcal{T}](x, y) = a_1(x)y + a_2(x)$ for $a_1, a_2 \in \mathbb{Z}[x]$. Moreover, the map $A: C[0, 1] \rightarrow C[0, 1]$ sending f to $a_1(x)f(x) + a_2(x)$ is a (strict) contraction with respect to the maximum norm, unless \mathcal{T} is a path with the root positioned on an end vertex.*

Proof. From the construction of the polynomial, $p[\mathcal{T}](x, x) = p[\mathcal{T}](x)$. Consider the maximal faithful subtree of \mathcal{T} not containing the leaf v_* to which y was assigned, and denote it by $\mathcal{T}_{\setminus y}$. It follows from Proposition 7 that $p[\mathcal{T}_{\setminus y}](x)$ is the part of $p[\mathcal{T}](x)$ that accounts for all those y -avoiding subtrees; thus $a_2(x) = p[\mathcal{T}_{\setminus y}](x)$. (We set $p[\emptyset](x) = 0$ for the empty tree, to cover the case where \mathcal{T} had only one leaf to begin with.) Consequently, $p[\mathcal{T}] - p[\mathcal{T}_{\setminus y}]$ is the polynomial corresponding to all faithful subtrees containing the y -leaf. However, we still need to exchange one x -variable for a y -variable; thus $a_1(x) = \frac{1}{x} (p[\mathcal{T}](x) - p[\mathcal{T}_{\setminus y}](x))$. Observe that this is indeed a polynomial, as every one-variate tree polynomial is divisible by x .

To show that the operator A is a contraction, it suffices to verify that $\|a_1\|_{C[0,1]} < 1$, since

$$\|Af - Ag\|_{C[0,1]} = \|a_1(f - g)\|_{C[0,1]} \leq \|a_1\|_{C[0,1]} \cdot \|f - g\|_{C[0,1]}, \tag{25}$$

as desired. To this end, recall that by Proposition 6, the polynomial $p[\mathcal{T}_{\setminus y}](x)$ can be interpreted as the probability that $\mathcal{T}_{\setminus y}$ contains any path from the root to a leaf of \mathcal{T} different from v_* , when nodes are deleted independently with probability $1 - x \in [0, 1]$, and kept otherwise. Since the analogous statement is also true for $p[\mathcal{T}](x)$, this gives the following probabilistic interpretation of $a_1(x) = \frac{1}{x} (p[\mathcal{T}](x) - p[\mathcal{T}_{\setminus y}](x))$: the event that the only path from the root to a leaf in \mathcal{T} is the path to v_* , conditioned on the node v_* being present, has probability $a_1(x)$. From this, it follows immediately that $0 \leq a_1(x) < 1$ for $x \in [0, 1]$ unless v_* is the unique leaf of \mathcal{T} . The inequality $\|a_1\|_{C[0,1]} < 1$ now follows from the fact that a_1 is a polynomial and therefore continuous. \square

Let $(\mathcal{T}_n)_{n \in \mathbb{N}}$ be a family of finite sets of rooted trees, say

$$\mathcal{T}_n = \{\mathcal{T}_{n,1}, \dots, \mathcal{T}_{n,k_n}\}$$

for integers k_1, k_2, \dots . Given any rooted tree \mathcal{T} and any positive integer n , we denote by $M_n(\mathcal{T})$ the rooted tree which, upon removal of its root node, decomposes into one copy of $\mathcal{T}, \mathcal{T}_{n,1}, \dots, \mathcal{T}_{n,k_n}$ each, with the roots of these trees being the children of the root in $M_n(\mathcal{T})$. We can think of $M_n(\cdot)$ for fixed n as being a map from the set of rooted trees to itself. Accordingly, those maps can be composed, and we use the shorthand notation

$$M_{n_\ell, \dots, n_1}(\cdot) = M_{n_\ell} \circ \dots \circ M_{n_1}(\cdot)$$

for any finite sequence n_1, \dots, n_ℓ of positive integers.

Theorem 2. *Let \bullet be the rooted tree consisting of one vertex, and let \mathcal{T}_* be a fixed rooted tree. Consider a family $(\mathcal{T}_n)_{n \in \mathbb{N}}$ of finite sets of rooted trees, as above. If there is a constant $c < 1$ such that*

$$\left\| x \prod_{j=1}^{k_n} (1 - p[\mathcal{T}_{n,j}](x)) \right\|_{C[0,1]} \leq c \text{ for all } n \in \mathbb{N}, \tag{26}$$

then

$$\|p[M_{n, \dots, 1}(\mathcal{T}_*)](x) - p[M_{n, \dots, 1}(\bullet)](x)\|_{C[0,1]} \rightarrow 0$$

as $n \rightarrow \infty$.

Proof. Let $m \in \mathbb{N}$, and let \mathcal{T} be any rooted tree. By virtue of (19), we have

$$p[M_m(\mathcal{T})](x) = x - x \prod_{j=1}^{k_m} (1 - p[\mathcal{T}_{m,j}](x)) + xp[\mathcal{T}](x) \prod_{j=1}^{k_m} (1 - p[\mathcal{T}_{m,j}](x)).$$

Simultaneously, fixing the leaf \bullet in $M_k(\bullet)$ and assigning to it the variable y , we obtain the bivariate polynomial

$$p_\bullet[M_m(\bullet)](x, y) = x - x \prod_{j=1}^{k_m} (1 - p[\mathcal{T}_{m,j}](x)) + xy \prod_{j=1}^{k_m} (1 - p[\mathcal{T}_{m,j}](x)).$$

In particular,

$$p[M_m(\mathcal{T})](x) = p_\bullet[M_m(\bullet)](x, p[\mathcal{T}](x)).$$

Hence the operator $A_m: p[\mathcal{T}](x) \mapsto p[M_m(\mathcal{T})](x)$ has the form described in Lemma 5 and is therefore a strict contraction with an operator norm bounded by

$$\left\| x \prod_{j=1}^{k_m} (1 - p[\mathcal{T}_{m,j}](x)) \right\|_{C[0,1]},$$

according to the estimate (25). Therefore, iterating (25) yields

$$\begin{aligned} & \|p[M_{n,\dots,1}(\mathcal{T}_*)](x) - p[M_{n,\dots,1}(\bullet)](x)\|_{C[0,1]} \\ & \leq \prod_{i=1}^n \left\| x \prod_{j=1}^{k_i} (1 - p[\mathcal{T}_{i,j}](x)) \right\|_{C[0,1]} \cdot \|p[\mathcal{T}_*](x) - p[\bullet](x)\|_{C[0,1]} \\ & \leq c^n \|p[\mathcal{T}_*](x) - p[\bullet](x)\|_{C[0,1]} \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$. □

The general message of Theorem 2 is thus that two trees that differ only 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 the condition (26) holds. We will apply this idea in Proposition 8 below.

6. Examples: caterpillar trees, stars, and complete binary trees

Example 1. Let CP_n denote the binary caterpillar tree on $2n + 1$ nodes, that is, a rooted binary tree where every node either is a leaf, or has at least one child which is a leaf (cf. Figure 3, left). The associated polynomials then satisfy the recurrence relation

$$\begin{aligned} p[CP_{n+1}](x) &= x(1 - (1 - x)(1 - p[CP_n](x))) \\ &= x^2 + xp[CP_n](x) - x^2p[CP_n](x) \end{aligned}$$

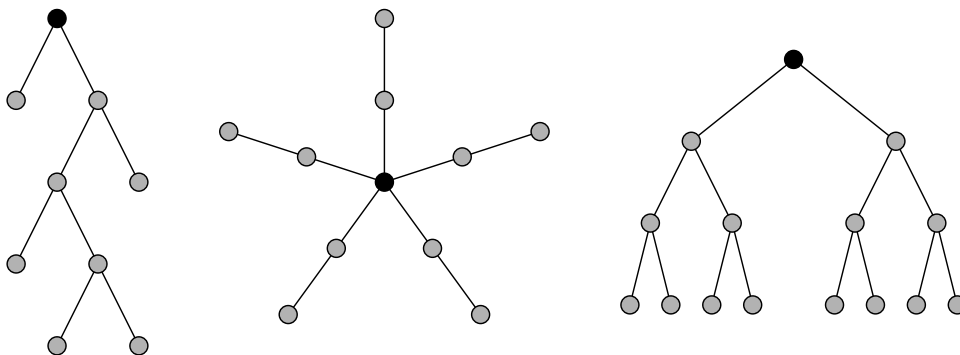


FIGURE 3. The trees CP_4 , $ST_{5,2}$, and CBT_4 , from left to right, with the root shown in black.

with $p[CP_0](x) = x$. By the Banach fixed point theorem in $C[0,1]$, this recurrence converges to the unique solution of the associated fixed point equation $f = x^2 + xf - x^2f$, which is $f(x) = \frac{x^2}{x^2-x+1}$. Finally, since all functions involved are bounded on the interval $[0,1]$, we obtain

$$\mathbb{P}[\mathfrak{S}(CP_n) = \mathfrak{C}(CP_n)] \longrightarrow \int_0^1 \frac{x}{1-x+x^2} dx = \frac{\pi}{3\sqrt{3}} \approx 0.6046$$

as $n \rightarrow \infty$ by applying the dominated convergence theorem to Corollary 3.

Moreover, we can use these very same tools to compute the limiting distribution of $|CP_{n,\mathfrak{S}}|$. Indeed, every admissible graph must be a path of some length, say $m \geq 0$, which then has m leaves as boundary plus one further node that induces a subtree isomorphic to CP_{n-m} . Hence, by Corollary 3, we have for all $m \geq 0$ as $n \rightarrow \infty$

$$\begin{aligned} \mathbb{P}[|CP_{n,\mathfrak{S}}| = m] &= \int_0^1 x^{m-1}(1-x)^m(mx + p[CP_{n-m}](x)) dx \\ &\longrightarrow m B(m+1, m+1) + \int_0^1 x^m(1-x)^m \frac{x}{1-x+x^2} dx. \end{aligned} \tag{27}$$

To evaluate this limit, we note first that

$$\frac{x^{m+1}(1-x)^m}{x^2-x+1} = \frac{x}{x^2-x+1} - \sum_{j=0}^{m-1} x^{j+1}(1-x)^j,$$

so that the integral in (27) equals

$$\begin{aligned} \int_0^1 \frac{x^{m+1}(1-x)^m}{1-x+x^2} dx &= \int_0^1 \frac{x}{x^2-x+1} dx - \sum_{j=0}^{m-1} B(j+2, j+1) \\ &= \frac{\pi}{3\sqrt{3}} - \sum_{j=0}^{m-1} \frac{1}{j+1} \binom{2j+2}{j+1}^{-1}. \end{aligned}$$

Using these results, we obtain for the limit in (27)

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\text{CP}_{n,\mathfrak{S}}| = m] = \int_0^1 \left(m(x - x^2)^m + \frac{x(x - x^2)^m}{1 - x + x^2} \right) dx \tag{28}$$

$$= \frac{\pi}{3\sqrt{3}} + \frac{m}{(m + 1)\binom{2m+1}{m}} - \sum_{j=0}^{m-1} \frac{1}{(j + 1)\binom{2j+2}{j+1}}. \tag{29}$$

These limiting expressions indeed form a probability distribution on the non-negative integers, as can be comfortably shown using Theorem 1: in the notation of this theorem, we have $|\partial_S A| = m + 1$ for all admissible graphs A on m vertices, so $b = 1$, and the radius of convergence of

$$\sum_{m=0}^{\infty} a_m x^m = \sum_{m=0}^{\infty} (m + 1)x^m = \frac{1}{(1 - x)^2}$$

equals $1 > 1/4$. Hence, the random variables $|\text{CP}_{n,\mathfrak{S}}|$ form a tight, vaguely converging sequence, which must thus also converge in distribution.

Remark 5. It is also possible to verify directly that (28) defines a random variable. To see this, observe that the integrand in (28) is non-negative. We thus obtain from the monotone convergence theorem that

$$\begin{aligned} \sum_{m=0}^{\infty} \lim_{n \rightarrow \infty} \mathbb{P}[|\text{CP}_{n,\mathfrak{S}}| = m] &= \int_0^1 \left(\sum_{m=0}^{\infty} m(x - x^2)^m + \frac{x}{1 - x + x^2} \sum_{m=0}^{\infty} (x - x^2)^m \right) dx \\ &= \int_0^1 \frac{2x - x^2}{(1 - x + x^2)^2} dx = 1, \end{aligned} \tag{30}$$

as required.

Furthermore, the estimate

$$m(x - x^2)^m + \frac{x(x - x^2)^m}{1 - x + x^2} \leq \frac{m + 1}{4^m}, \quad x \in [0, 1],$$

implies that a random variable X having the probability distribution defined by (28) has exponential tails. In particular, all its moments exist, and they can be computed in a similar fashion to (30). Omitting the details, we obtain

$$\begin{aligned} \mathbb{E}[X] &= \frac{1}{54}(9 + 4\sqrt{3}\pi) \approx 0.5697, \\ \mathbb{V}[X] &= \frac{1}{972}(783 + 16\sqrt{3}\pi - 16\pi^2) \approx 0.7327. \end{aligned}$$

Example 2. Let $\text{ST}_{r,n}$ denote the rooted tree on $rn + 1$ vertices which, upon removal of the root node, splits into r copies of a path on n vertices (cf. Figure 3, middle). Hence, the tree polynomial of $\text{ST}_{r,n}$ is given by

$$p[\text{ST}_{r,n}](x) = x(1 - (1 - x^n)^r), \tag{31}$$

and the probability of separating the tree by cutting the root node is therefore given by

$$\begin{aligned} \mathbb{P}[\mathfrak{S}(\text{ST}_{r,n}) = \mathfrak{C}(\text{ST}_{r,n})] &= \int_0^1 1 - (1 - x^n)^r \, dx \\ &= 1 - \frac{1}{n} \int_0^1 y^{\frac{1}{n}-1} (1 - y)^r \, dy = 1 - \frac{1}{n} \text{B} \left(\frac{1}{n}, r + 1 \right) \end{aligned} \tag{32}$$

after substituting $y = x^n$, according to Corollary 3. Letting the size of the tree grow to infinity, we obtain different asymptotic behaviours of this probability, depending on the relationship between r and n . To see this, let $\rho \in (0, 1)$ and set $r = r(n) = (1 - \rho)^{-n} - 1$. In the interest of brevity, write $z := (1 - \rho)^{-n} = r(n) + 1$. Then, using Stirling’s approximation for the gamma function, we obtain

$$\begin{aligned} \frac{1}{n} \text{B} \left(\frac{1}{n}, r(n) + 1 \right) &= \Gamma \left(1 + \frac{1}{n} \right) \frac{\Gamma(z)}{\Gamma \left(z - \frac{\ln(1-\rho)}{\ln z} \right)} \\ &\sim \sqrt{\frac{z - \frac{\ln(1-\rho)}{\ln z}}{z}} \cdot \frac{e^{z - \frac{\ln(1-\rho)}{\ln z}}}{e^z} \cdot \frac{z^z}{\left(z - \frac{\ln(1-\rho)}{\ln z} \right)^{z - \frac{\ln(1-\rho)}{\ln z}}} \longrightarrow 1 - \rho \end{aligned}$$

as n (and therefore z) tends to infinity. Hence, in light of (32), and by employing monotonicity properties with respect to r and n , we get the following results:

1. If the number r of rays grows subexponentially with n , then we have $\mathbb{P}[\mathfrak{S}(\text{ST}_{r,n}) = \mathfrak{C}(\text{ST}_{r,n})] \rightarrow 0$. In particular, this covers the intuitively obvious case when the number of rays is bounded.
2. If $r \sim (1 - \rho)^{-n}$, then by the above, $\mathbb{P}[\mathfrak{S}(\text{ST}_{r,n}) = \mathfrak{C}(\text{ST}_{r,n})] \rightarrow \rho$.
3. If r grows superexponentially with n (in particular, if n is bounded), then $\mathbb{P}[\mathfrak{S}(\text{ST}_{r,n}) = \mathfrak{C}(\text{ST}_{r,n})] \rightarrow 1$.

Example 3. 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 $h = 0, \dots, n - 1$ (cf. Figure 3, right). Observe that thus CBT_1 is the tree consisting of only the root node, and that CBT_{n+1} splits into two copies of CBT_n upon removal of the root node. Thus, the associated polynomials satisfy the recurrence equation

$$\begin{aligned} p[\text{CBT}_{n+1}](x) &= x(1 - (1 - p[\text{CBT}_n](x))^2) \\ &= 2xp[\text{CBT}_n](x) - xp[\text{CBT}_n](x)^2 \end{aligned}$$

with $p[\text{CBT}_1](x) = x$. The fixed point equation $f = 2xf - xf^2$ admits two solutions, $f \equiv 0$ and $f(x) = 2 - \frac{1}{x}$. For convenience, let $\varphi(x) := \max \left\{ 0, 2 - \frac{1}{x} \right\}$. Direct verification reveals that if a function g satisfies $g(x) \geq \varphi(x)$ for all x , then $2xg(x) - xg(x)^2 \geq \varphi(x)$ as well. Furthermore, the sequence of polynomials $p[\text{CBT}_n]$ decreases monotonically pointwise by Lemma 4, and hence converges pointwise to $\varphi(x)$.

In light of Proposition 4, 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 $\text{Ber}(x)$ -site percolation, as can be verified independently. For example, in [14, p. 256], a similar recursive argument is used to determine that the probability of the root being contained in an infinite cluster of $\text{Ber}(p)$ -bond percolation is $\varphi(p)/p$. While they are not the same, one can define a bijection between the edges and the non-root vertices in a rooted tree by mapping any edge to the endpoint farther away from the root. This allows us to translate between site and bond percolation and explains Grimmett’s additional factor of p^{-1} .

Continuing with our analysis, the probability of the remaining tree at separation being empty now follows handily from Corollary 3:

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\text{CBT}_{n, \mathfrak{S}}| = 0] = \int_0^1 \frac{\varphi(x)}{x} dx = \ln 4 - 1 \approx 0.3863.$$

In a similar fashion, we can continue to determine the limiting probability of separation graphs of any size $m \geq 0$: since there are $C_m = \frac{1}{m+1} \binom{2m}{m}$ subtrees (the C_m are, of course, the Catalan numbers, [24, A000108]) of the infinite rooted binary tree on m vertices, and each of those has $m + 1$ boundary vertices, we claim that

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\text{CBT}_{n, \mathfrak{S}}| = m] = \binom{2m}{m} \int_0^1 x^{m-1} (1-x)^m \varphi(x) dx. \tag{33}$$

This can be derived analogously to Equation (27), where the exchange of the limit and integral can be motivated by recalling the estimate $p[\text{CBT}_n](x) \leq x$, so that the integrand is bounded by $x^m (1-x)^m$. Hence, by the dominated convergence theorem, we arrive at (33).

Observe that, in the notation of Theorem 1, the sequence $a_m = \binom{2m}{m}$ has generating function $\frac{1}{\sqrt{1-4x}}$ (cf. [24, A000984]), thus violating the integrability condition (18) for the proper constant $b = 1$. However, we can check by hand that (33) defines a probability distribution with the following computation, again relying on the generating function of $\binom{2m}{m}$:

$$\begin{aligned} & \sum_{m=0}^{\infty} \binom{2m}{m} \int_0^1 x^{m-1} (1-x)^m \varphi(x) dx \\ &= \int_{1/2}^1 \left(\sum_{m=0}^{\infty} \binom{2m}{m} (x-x^2)^m \right) \cdot \frac{2x-1}{x^2} dx \\ &= \int_{1/2}^1 \frac{1}{\sqrt{1-4x+4x^2}} \cdot \frac{2x-1}{x^2} dx \\ &= \int_{1/2}^1 \frac{1}{x^2} dx = 1. \end{aligned} \tag{34}$$

Note also that a random variable X having the probability distribution defined by (33) does not have a finite first moment: imitating the approach of (34) leads to

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

where the integral on the right-hand side diverges.

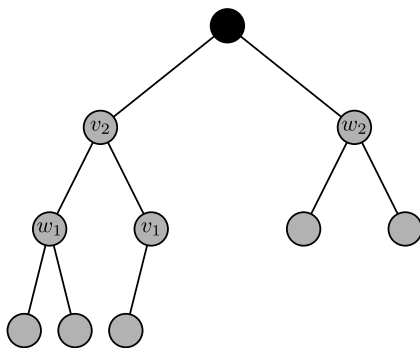


FIGURE 4. The vertices v_i and w_i from the proof of Proposition 8 in \mathcal{T}_{10} .

Finally, we can use Lemmas 4 and 5 to verify that the same limiting distribution 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 $0 \leq k < \lfloor \lg n \rfloor =: m$, with the remaining $n - 2^m + 1$ vertices at height m in their leftmost positions.

Proposition 8. *With $\varphi(x) = \max \left\{ 0, 2 - \frac{1}{x} \right\}$ as above, we have $p[\mathcal{T}_n](x) \rightarrow \varphi(x)$ uniformly over $[0, 1]$, as $n \rightarrow \infty$.*

Consequently, the limiting distribution of $|\mathcal{T}_{n,\mathfrak{S}}|$ coincides with that of $|\text{CBT}_{n,\mathfrak{S}}|$ and is given by Equation (33).

Proof. If n is odd, then the number of vertices at height $m + 1$ is even, so we have $\text{CBT}_{m-1} \subseteq \mathcal{T}_n \subseteq \text{CBT}_m$, where, in both inclusions, the smaller tree is a trimmed subtree of the larger according to Definition 4. Accordingly, we obtain from Lemma 4 that

$$p[\text{CBT}_{m-1}](x) \geq p[\mathcal{T}_n](x) \geq p[\text{CBT}_m](x)$$

for all $x \in [0, 1]$. Hence $p[\mathcal{T}_n](x) \rightarrow \varphi(x)$ pointwise (and uniformly) for odd $n \rightarrow \infty$.

For even n , observe that \mathcal{T}_n and \mathcal{T}_{n-1} differ only in a fringe subtree of height 2. Indeed, let root, v_{m-1}, \dots, v_1 be the path from the root to the parent of the unique leaf ℓ without a sibling vertex in \mathcal{T}_n . Each of the vertices v_i is also in \mathcal{T}_{n-1} , and in both trees they have a unique sibling, w_i , for $1 \leq i < m$ (see Figure 4). Writing $\mathcal{T}_n^{(v)}$ for the fringe subtree of \mathcal{T}_n rooted at a vertex v , and setting $\mathcal{T}_j := \left\{ \mathcal{T}_n^{(w_j)} \right\}$, we observe that the condition (26) is satisfied: since every $\mathcal{T}_n^{(w_j)}$ is a non-empty full complete binary tree, we have

$$\varphi(x) \leq p \left[\mathcal{T}_n^{(w_j)} \right] (x) \leq x$$

for $x \in [0, 1]$ by the results of Example 3. Accordingly,

$$\left\| x \left(1 - p \left[\mathcal{T}_n^{(w_j)} \right] (x) \right) \right\|_{C[0,1]} \leq \|x(1 - \varphi(x))\|_{C[0,1]} = \frac{1}{2}$$

for all j . Therefore,

$$\|p[\mathcal{T}_n] - p[\mathcal{T}_{n-1}]\|_{C[0,1]} \rightarrow 0$$

as $m \rightarrow \infty$ for even n by Theorem 2, yielding the desired convergence. (In fact, repeating the estimates from the proof of Theorem 2 reveals the more concise estimate $\|p[\mathcal{T}_n] - p[\mathcal{T}_{n-1}]\| \leq 2^{1-m}$).

The second claim concerning the limiting distribution of the size of the separation graph follows in the same way as when it was first derived in Example 3. □

Observe that by this proposition and by (34), the random variables $|\mathcal{T}_{n,\mathfrak{S}}|$ converge in distribution and are therefore tight. By Remark 3 this means that the limit law obtained by Janson in [17, 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 4. *Let $n \rightarrow \infty$ such that $\{\lg n - \lg \lg n\} \rightarrow \gamma \in [0, 1]$. Write $f(\gamma) := 2^\gamma - 1 - \gamma$. Let W_γ be a random variable with an infinitely divisible distribution with characteristic function*

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

with the Lévy measure v_γ being supported on $[0, \infty)$ with $dv_\gamma = 2^{\{\lg x + \gamma\}} x^{-2} dx$. Then

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

7. Further questions

We use this final section to present several (deliberately broad) questions and 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, in continuation of the wide variety of work done on the cutting number of 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? This promises to be non-trivial already for rooted trees and their leaves. Moreover, 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 can one evaluate the asymptotic distribution of \mathfrak{S} directly, without relying on previous knowledge of \mathfrak{C} as in Corollary 4?
4. On trees, the difference between edge-cutting and vertex-cutting is usually negligible because there is a bijection between edges and non-root vertices, assigning the endpoint farther away from the root to each edge. For general graphs, no such bijection exists. However, 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 following 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 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 $j = 0$ specialises to the separation number as we defined it. However, since the notion of a boundary that we used to prove the results in Section 4 is no longer relevant if $j > 0$, the question is what kind of statements can be obtained for the more general case.

Acknowledgements

The author wishes to thank his academic advisors, Cecilia Holmgren and Svante Janson, for their generous support and many helpful remarks and discussions. He is also grateful to Jonas Sjöstrand, who pointed out a mistake in Corollary 2 and suggested a simplification for the formulas there.

Funding information

This work was partially supported by grants from the Knut and Alice Wallenberg Foundation, the Ragnar Söderberg Foundation, and the Swedish Research Council.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

- [1] ABRAMOWITZ, M. AND STEGUN, I. A. (1972). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. Dover, New York.
- [2] ADDARIO-BERRY, L., BROUTIN, N. AND HOLMGREN, C. (2014). Cutting down trees with a Markov chainsaw. *Ann. Appl. Prob.* **24**, 2297–2339.
- [3] ALON, N., BENJAMINI, I. AND STACEY, A. (2004). Percolation on finite graphs and isoperimetric inequalities. *Ann. Prob.* **32**, 1727–1745.
- [4] BERTOIN, J. (2012). Fires on trees. *Ann. Inst. H. Poincaré Prob. Statist.* **48**, 909–921.
- [5] BERZUNZA, G. (2017). The cut-tree of large trees with small heights. *Random Structures Algorithms* **51**, 404–427.
- [6] BERZUNZA, G., CAI, X. S. AND HOLMGREN, C. (2019). The k -cut model in deterministic and random trees. Preprint. Available at <https://arxiv.org/abs/1907.02770>.
- [7] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*, 1st edn. John Wiley, New York.
- [8] BROUTIN, N. AND WANG, M. (2017). Cutting down \mathbf{p} -trees and inhomogeneous continuum random trees. *Bernoulli* **23**, 2380–2433.
- [9] CAI, X. S. AND HOLMGREN, C. (2019). Cutting resilient networks—complete binary trees. *Electron. J. Combinatorics* **26**, article no. P4.43.
- [10] CAI, X. S., HOLMGREN, C., DEVROYE, L. AND SKERMAN, F. (2019). k -cut on paths and some trees. *Electron. J. Prob.* **24**, article no. 53.
- [11] CAMPOS, V., CHVATAL, V., DEVROYE, L. AND TASLAKIAN, P. (2013). Transversals in trees. *J. Graph Theory* **73**, 32–43.
- [12] DEVROYE, L. (2011). A note on the probability of cutting a Galton-Watson tree. *Electron. J. Prob.* **16**, 2001–2019.
- [13] FLAJOLET, P. AND SEDGEWICK, R. (2009). *Analytic Combinatorics*. Cambridge University Press.
- [14] GRIMMETT, G. R. (1999). *Percolation*, 2nd edn. Springer, Berlin, Heidelberg.
- [15] HOLMGREN, C. (2010). Random records and cuttings in binary search trees. *Combinatorics Prob. Comput.* **19**, 391–424.
- [16] HOLMGREN, C. (2011). A weakly 1-stable distribution for the number of random records and cuttings in split trees. *Adv. Appl. Prob.* **43**, 151–177.

- [17] JANSON, S. (2004). Random records and cuttings in complete binary trees. In *Mathematics and Computer Science III: Algorithms, Trees, Combinatorics and Probabilities*, Birkhäuser, Basel, pp. 241–253.
- [18] JANSON, S. (2006). Random cutting and records in deterministic and random trees. *Random Structures Algorithms* **29**, 139–179.
- [19] KUBA, M. AND PANHOLZER, A. (2008). Isolating a leaf in rooted trees via random cuttings. *Ann. Combinatorics* **12**, 81–99.
- [20] KUBA, M. AND PANHOLZER, A. (2008). Isolating nodes in recursive trees. *Aequat. Math.* **76**, 258–280.
- [21] KUBA, M. AND PANHOLZER, A. (2014). Multiple isolation of nodes in recursive trees. *Online J. Anal. Combinatorics* **9**, 14 pp.
- [22] LYONS, R. AND PERES, Y. (2017). *Probability on Trees and Networks*. Cambridge University Press.
- [23] MEIR, A. AND MOON, J. W. (1970). Cutting down random trees. *J. Austral. Math. Soc.* **11**, 313–324.
- [24] OEIS Foundation Inc. (2021). The On-Line Encyclopedia of Integer Sequences. Available at <http://oeis.org>.
- [25] PANHOLZER, A. (2003). Non-crossing trees revisited: cutting down and spanning subtrees. In *Discrete Random Walks (Discrete Mathematics and Theoretical Computer Science Proceedings AC)*, eds C. Banderier and C. Krattenthaler, DMTCS, Nancy, pp. 265–276.
- [26] PANHOLZER, A. (2006). Cutting down very simple trees. *Quaest. Math.* **29**, 211–227.