# The Uniform Spanning Tree and related models

Antal A. Járai

December 18, 2009

## 1 Introduction

The following type of question is characteristic of discrete probability: given a large finite set of objects, defined by some combinatorial property, what does a typical element "look like", that is, can we describe some of its statistical features. Simple as it is to state such questions, they can lead to surprisingly deep and difficult mathematical problems.

In this course we are going to look at a very special example: random spanning trees of graphs (see precise definitions later in this introduction). Spanning trees have been well-known in combinatorics for a long time; see [6]. However, the study of random spanning trees is relatively recent. The topic proves to be very interesting from several points of view. There is a surprising connection between spanning trees, random walk and electric circuits. The connection with random walk yields efficient algorithms to generate a spanning tree of a graph uniformly at random. We will see two such beautiful algorithms. Uniform spanning trees are also interesting from the point of view of statistical physics, as they are a special case of so-called random cluster measures; see Exercise 1.1. The notion of *amenability*, that comes from group theory, will enter some of our discussions. Another interesting aspect of random spanning trees is that they are an example of a *determinantal process* (see its definition later in this course).

We now introduce some basic notions.

### 1.1 Spanning Trees

A graph is called a *tree*, if it is connected, and contains no cycles. A graph that contains no cycles, and is not necessarily connected is called a *forest* (its connected components are trees). Let $G = (V, E)$ be a connected graph.

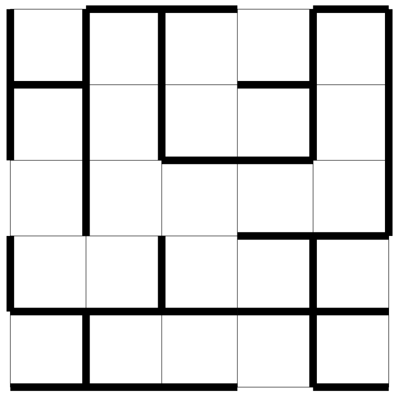

Figure 1: A subgraph of the square grid and one of its spanning trees

A *spanning subgraph* of $G$ is one that contains every vertex of $V$ (some of these may be isolated vertices not incident on any edge of the subgraph). A *spanning tree* of $G$ is a spanning subgraph of $G$ that is a tree. Similarly we define *spanning forest*. See Figure 1. We can identify a spanning subgraph of $G$ with an element of the space $\Omega = \{0,1\}^E$, by writing a 1 if an edge is present, and 0 if it is not. That is, $\omega = (\omega_e)_{e \in E} \in \Omega$ represents the spanning subgraph which contains precisely those edges $e \in E$ for which $\omega_e = 1$. Assume now that $G$ is a finite graph. By the *Uniform Spanning Tree* on $G$, we mean the probability distribution $\mathsf{UST}$ on $\Omega$ that assigns equal mass to each spanning tree of $G$, and no mass to subgraphs that are not spanning trees:

$$\mathsf{UST}[\omega] = \begin{cases} \frac{1}{N(G)} & \text{if } \omega \text{ is a spanning tree of } G; \\ 0 & \text{otherwise,} \end{cases},$$

where $N(G)$ is the number of spanning trees of $G$. By drawing an example or looking at Figure 1, you can convince yourself that a general connected graph usually has many spanning trees, so this definition meaningful.

Having introduced the space $\Omega$, we can view the Uniform Spanning Tree as a 0-1 valued random process $(\omega_e)$ indexed by the edges of $G$. It should be clear that this is very far from an i.i.d. process, since cycles are forbidden, and the subgraph is constrained to be connected. We will write $T = \{e \in E : \omega_e = 1\}$. Some questions we will consider: what is the probability that a given edge $e \in E$ belongs to $T$? More generally, given $K \subset E$, what is $\mathsf{UST}[K \subset T]$? A different type of question that proves to be very interesting is as follows. Suppose now that $G$ is the $d$-dimensional integer lattice: $G = (\mathbb{Z}^d, \mathbb{E}^d)$, $d \geq 2$, with $x, y \in \mathbb{Z}^d$ connected by an edge, if $|x - y| = 1$, that is, $\mathbb{E}^d = \{\{x, y\} : x, y \in \mathbb{Z}^d, |x-y| = 1\}$. This $G$ has infinitely many spanning trees, hence it is not clear what should be meant by picking one uniformly at random. It turns out, as will be discussed later in the course, that there is a very natural way this can be done. Consider a sequence $V_1 \subset V_2 \subset \dots V_n \subset \cdots \subset \mathbb{Z}^d$, of finite connected subsets of $\mathbb{Z}^d$, such that $\cup_n V_n = \mathbb{Z}^d$. Let $G_n = (V_n, E_n)$ be the subgraph induced by $V_n$, that is the graph with vertex set $V_n$ and edge set the set of all edges of $G$ that connect vertices in $V_n$ (Figure 1 shows an example $V$ when $d = 2$.) On each $G_n$, we can construct the measure $\mathsf{UST}_n$. Do these approximate, in a suitable sense, a measure on $\Omega = \{0,1\}^{\mathbb{E}^d}$? Making this precise: is it true that for a fixed $e \in \mathbb{E}^d$, the probabilities $\mathsf{UST}_n[e \in T]$ converge to some limit (noting that the probability is indeed well defined as soon as $e \in E_n$)? More generally, for $K \subset \mathbb{E}^d$ finite, does $\mathsf{UST}_n[K \subset T]$ converge to some limit? The reader is invited to give some thought to what it would involve to prove such convergence, before moving on. Once we prove that the limit exists, we will examine in detail what type of graph is the limiting $T$.

**Exercise 1.1** (Random Cluster Model)**.** Let $G = (V, E)$ be a finite graph, and consider the following probability measure on $\Omega = \{0,1\}^E$.

$$\mu_{p,q}(\omega) = \frac{1}{Z_{p,q}} p^{n(\omega)}(1-p)^{|E|-n(\omega)} q^{K(\omega)}, \quad \omega \in \Omega,$$

where $0 < p < 1$, and $q > 0$ are parameters, $n(\omega) = \sum_{e \in E} \omega_e$, $K(\omega) =$ number of connected components of the graph $\omega$, and $Z_{p,q}$ is a constant normalizing $\mu_{p,q}$ to be a probability distribution. This is called the Random Cluster Measure on $G$.
(a) Check that when $q = 1$, the edges are i.i.d. This is called the Percolation Model on $G$ with edge density $p$.
(b) Prove that as $p \to 0$ and $q/p \to 0$, $\mu_{p,q}[\omega] \to \mathsf{UST}[\omega]$.

## 2 The Aldous-Broder algorithm

A large graph usually has many spanning trees; often exponentially many in the number of edges of the graph. (The famous Matrix-Tree Theorem in combinatorics [6, Corollary II.13] gives the exact number of them.) Hence for a general graph, it is not obvious how to simulate a uniformly random spanning tree in reasonable time; that is, polynomial in the number of edges. The algorithm below that achieves this, was discovered independently by Aldous [1] and Broder [7].

Let $G = (V, E)$ be a connected finite graph. We will write $\deg(u)$ for the number of edges incident on $u$. Let $(X_n)_{n\geq 0}$ be the simple random walk on $G$, that is, the Markov chain with state space $V$ and transition probability

$$p_{uv} = \begin{cases} \frac{1}{\deg(u)} & \text{if } \{u, v\} \in E; \\ 0 & \text{otherwise.} \end{cases}$$

Since $G$ is connected, $(X_n)$ is irreducible. Let $T_v$ be the first hitting time of $v$:

$$T_v := \inf\{n \geq 0 : X_n = v\}.$$

Let $C$ be the *cover time* of the random walk:

$$C := \max_v T_v = \inf\{n \geq 0 : \{X_0, X_1, \ldots, X_n\} = V\}.$$

By irreducibility, $C$ is finite with probability 1.

**Algorithm 2.1** (Aldous-Broder Algorithm)**.** Choose $X_0$ arbitrarily. Run simple random walk on $G$ up to the cover time $C$. Consider the set of edges:

$$T := \{\{X_{T_w-1}, X_{T_w}\} : w \neq X_0\}.$$

That is, every time you visit a vertex that you have not seen before, record which edge was used to enter that vertex. Then $T$ is a spanning tree, and it is uniformly distributed over all spanning trees of $G$.

Let us note that it is easy to see that $T$ is a spanning tree: we always draw edges to vertices that have not been visited before, so no cycles are formed during the process. By induction, $T$ is connected, since every edge has an endvertex that has been visited already.

The starting point for the proof that the algorithm works is a beautiful and surprising theorem that shows a connection between spanning trees and Markov chains.

## 2.1 The Markov Chain-Tree Theorem

Let us put aside simple random walk for a moment, and let $(X_n)_{n\geq 0}$ be an arbitrary irreducible finite state Markov chain on a state space $V$, with transition matrix $P = (p_{xy})$. We associate to it a directed graph $G = (V, E)$, where $[x, y] \in E$, if $p_{xy} > 0$. That is, we draw an edge from $x$ to $y$, if it is possible to move from $x$ to $y$ in one step. By a *rooted tree* $(t, r)$, we mean a tree $t$ with a distinguished vertex $r$, called the root. From every vertex of the tree, there is a unique path to the root that involves no backtracking. Thus for every vertex $v$ of a rooted tree, there is a unique edge $\{v, w\}$ of the tree that leads one step closer to the root, and we orient this edge towards the root, that is as $[v, w]$. It is easy to see that this way each edge of the rooted tree gets an orientation.

Consider a rooted spanning tree $(t, r)$ of $G$. We assign to this the weight

$$q(t, r) = \prod_{[v,w]\in E(t,r)} p_{vw}.$$

That is, the weight of a rooted spanning tree is the product of the transition probabilities along the edges of the tree. Let

$$p(v) := \sum_{(t,v)} q(t, v), \quad v \in V,$$

where the sum is over all rooted spanning trees of $G$ with root $v$.

**Theorem 2.1** (Markov Chain-Tree Theorem)**.** *The stationary distribution of $X$ is proportional to $p(v)$.*

*Proof.* The key to the proof is that one can build a Markov chain on rooted trees on top of $(X_n)$. Consider the following operation. Given a rooted spanning tree $(t', v')$ of $G$, and an edge $[v', v] \in E$, we define a new rooted spanning tree $(t, v)$. We first add the edge $[v', v]$ to $t'$. Since $t'$ already contains a directed path from $v$ to $v'$, this together with the edge $[v', v]$ forms a directed cycle. Note that due to the orientation of edges in $t'$, all other edges are oriented towards this cycle. Hence if we now remove the outgoing edge from $v$, $[v, w]$ say, then we obtain a rooted spanning tree $t$ rooted at the new vertex $v$. See Figure 2. Note that it may happen that $w = v'$, in which case the cycle consists of the edges $[v', v], [v, v']$. It is also no problem if $p_{v'v'} > 0$, and hence the loop-edge $[v', v']$ is present. In this case we add $[v', v']$ and then remove it. Let us note here what the backwards operation is. If $(t, v)$ and $w$ are specified, then adding the edge $[v, w]$ we get back the directed cycle, and then $v'$ is uniquely determined as the vertex preceeding

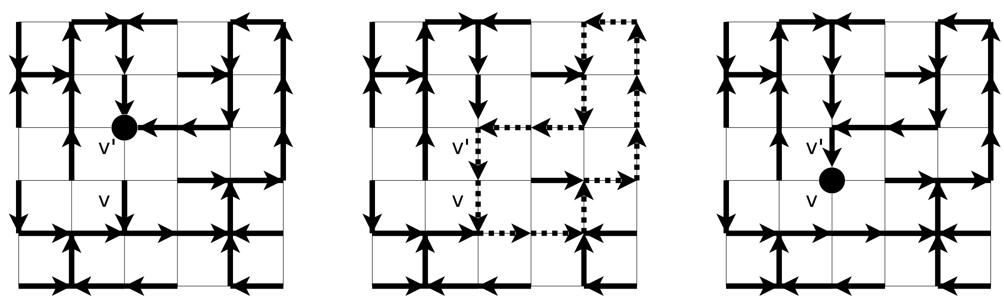


Figure 2: The operation that changes a spanning tree rooted at $v'$ into a spanning tree rooted at $v$. The dashed lines indicate the unique oriented cycle that is created by adding the edge $[v', v]$.

$v$ in this cycle, and hence also $t'$ is uniquely determined. Also note that for fixed $(t, v)$, to any $w$ with $[v, w] \in E$ there corresponds a "precursor tree" $(t', v')$. Again, since adding $[v, w]$ creates a unique directed cycle, etc. When $(t, v)$ is fixed, let us write $(t', v') = \mathrm{prec}((t, v), w)$, to indictae the relationship between $(t, v)$, $w$ and $(t', v')$.

We turn the above operation into a Markov chain $(Y_n)$, by letting the root perform the underlying Markov chain $(X_n)$. That is,

$$\mathsf{P}[Y_{n+1} = (t, v) \,|\, Y_n = (t', v')] := \mathsf{P}[X_{n+1} = v \,|\, X_n = v'] = p_{v'v}.$$

We will see in Exercise 2.1, that $(Y_n)$ is irreducible.

We claim that the stationary distribution of $(Y_n)$ is proportional to $q(t, v)$. We can write

$$q(t, v) = \sum_{w:[v,w] \in E} p_{vw} q(t, v). \tag{1}$$

Let $(t', v')$ denote the rooted tree uniquely determined by $(t, v), w$. Then

$$p_{vw} q(t, v) = p_{v'v} q(t', v').$$

Hence the right-hand side of (1) can be rewritten

$$\sum_{(t',v')} p_{v'v} q(t', v'). \tag{2}$$

Since $p_{v'v}$ is the transition probability for moving from $(t', v')$ to $(t, v)$, this proves the claim (assuming Exercise 2.1).

The Theorem now follows, since $p(v)$ is proportional to the long run fraction of time that the root equals $v$, which is the long run fraction of time $X_n = v$. $\square$

**Exercise 2.1.** Show that given a rooted tree $(t, v)$, and for any $X_0$ and $Y_0$, there exists $N$ and a sequence $X_1, \ldots, X_N$ of allowed steps such that $Y_N = (t, v)$. Conclude that the chain $(Y_n)$ is irreducible. (Hint: traverse the edges of $t$ in a suitable order.)

## 2.2 The backwards tree construction

Let $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ be the stationary tree-chain run from time $-\infty$. We will write $\mathsf{P}$ for the probability measure governing this chain. Then $X_n$, the root of $Y_n$ is the stationary $X$ chain.

We define the *backward tree* at time $n$ as follows. For $w \neq X_n$, let

$$L_w^n = \sup\{k : k < n,\ X_k = w\}.$$

Due to irreducibility, $L_w^n > -\infty$ with probability 1. Then we have

$$Y_n = \{[X_{L_w^n}, X_{L_w^n+1}] : w \neq X_n\}.$$

That is, if for every $w \neq X_n$, we mark the edge that was used on the last exit from $w$ prior to time $n$, we get $Y_n$. This is easily verified from the definition of the $Y$-chain.

## 2.3 Verification of the Aldous-Broder algorithm

We now return to the task of generating a uniform spanning tree of a connected graph $G = (V, E)$. This is done by reversing the time in the backwards tree construction.

Let $X_n$ be the simple random walk on $G$. Then the directed graph constructed from $X$ is the same as $G$, with each edge being present with both orientations. It is easy to check that the stationary distribution is proportional to $\deg(u)$. We still write $\mathsf{P}$ for the probability measure corresponding to the stationary chain.

The simple random walk is reversible, that is, $\mathsf{P}[X_0 = x_0, \ldots, X_k = x_k] = \mathsf{P}[X_k = x_0, \ldots, X_0 = x_k]$. In particular,

$$\mathsf{P}[X_1 = x_1, \ldots, X_k = x_k \mid X_0 = v] = \mathsf{P}[X_{-k} = x_k, \ldots, X_{-1} = x_1 \mid X_0 = v]. \tag{3}$$

*Proof.* [Aldous-Broder Algorithm works] Orient each edge of $T$ towards the root. Let us prove that $T$ is uniformly distributed. For simple random walk, the weight of a rooted tree $(t, v)$ is

$$q(t, v) = \prod_{w \neq v} \frac{1}{\deg(w)} = \text{const} \cdot \deg(v). \tag{4}$$

It is clear from the definitions of the backwards tree and $T$, that the forward tree constructed from the sequence $v, x_1, x_2, \ldots$ will be $(t, v)$ if and only if the backwards tree constructed from $\ldots, x_2, x_1, v$ is $(t, v)$. Hence (3) gives

$$\begin{aligned} \mathsf{P}[T = (t, v) \mid X_0 = v] = \mathsf{P}[Y_0 = (t, v) \mid X_0 = v] &= \frac{\mathsf{P}[Y_0 = (t, v)]}{\mathsf{P}[X_0 = v]} \\ &= \frac{\text{const} \cdot q(t, v)}{\text{const} \cdot \deg(v)}. \end{aligned}$$

Due to (4), the right hand side is a constant independent of $(t, v)$. Hence each tree rooted at $v$ is equally likely to result from the algorithm. Forgetting the root then yields a uniformly distributed spanning tree. □

# 3 Wilson's algorithm

In this section, we present a second algorithm to generate random spanning trees, due to Wilson [34]. It turns out to be extremely useful not only as a simulation tool, but also for theoretical analysis. Before presenting the algorithm, we generalize the notion of random spanning trees from the uniform case to a weighted case.

## 3.1 Some terminology

Let $G = (V, E)$ be an unoriented graph. We allow multiple edges or loops in the graph. For many purposes we will be able to disregard loops, as they can never be contained in a tree. Often we will use oriented edges, and in this case, each edge will be present in $E$ with both orientation. For an oriented edge $e$ we write $e = [\underline{e}, \overline{e}]$, and call $\underline{e}$ the *tail* of $e$, and $\overline{e}$ the *head* of $e$. We define $\check{e} = [\overline{e}, \underline{e}]$ the reversal of $e$.

By a *network*, we mean a pair $(G, C)$, where $C : E \to (0, \infty)$, defined on unoriented edges (or, equivalently, $C$ is symmetric: $C(\check{e}) = C(e)$). We call $C(e)$ the *conductance* of $e$. For any vertex $v \in V$, we call $C_v := \sum_{e:\underline{e}=v} C(e)$, the *conductance* of $v$. For infinite networks, we assume that $C_v < \infty$ for all $v$. The *resistance* of $e$ is defined as $R(e) := 1/C(e)$. Every graph carries the default network $(G, \mathbf{1})$, where every edge has conductance 1.

For every $v \in V$, the *network random walk* started at $v$ is the Markov chain with $\mathsf{P}_v[X(0) = v] = 1$ and transition probabilities $\mathsf{P}_v[X(n+1) = w \,|\, X(n) = u] = C(u, w)/C_u$. Here, $C(u, w) = \sum_{e:\underline{e}=u,\overline{e}=w} C(e)$, allowing multiple edges.

On a finite connected network, we define the weight of a spanning tree by $\text{weight}(t) = \prod_{e\in t} C(e)$. In the sequel, by a *random spanning tree*, we mean one chosen with probability proportional to weight. The distribution $\mathsf{UST}$ is obtained as a special case when the conductances are 1.

## 3.2 Loop-erased random walk

Wilson's method is based on Loop-Erased Random Walk, which we now define. If $\mathcal{P} = [u_0, u_1, \ldots, u_n]$ is a path in $G$, we define its *loop-erasure* by chronologically removing loops from $\mathcal{P}$. Formally, this is defined as follows. We first set $\gamma_0 := u_0$. Let $s_1 := \max\{m : u_m = \gamma_0\}$. If $s_1 = n$, then $\text{LE}(\mathcal{P}) = [\gamma_0]$. Otherwise, let $\gamma_1 := u(s_1 + 1)$. Assuming $\gamma_0, \ldots, \gamma_k$ have been defined, we let $s_{k+1} := \max\{m : u_m = \gamma_k\}$. If $s_{k+1} = n$, we stop and $\text{LE}(\mathcal{P}) = [\gamma_0, \ldots, \gamma_k]$. Otherwise, we let $\gamma_{k+1} := u(s_{k+1} + 1)$. Note that the order in which we remove loops *does matter*, and in the above definition, we remove loops, as they are created, following the path. When loop-erasure is applied to a random walk path, we talk about Loop-Erased Random Walk (LERW).

## 3.3 Wilson's method

Let $(G, C)$ be a finite connected network. Pick any $r \in V$, called the "root". We define a growing sequence of subtrees $T(i)$, $i \geq 0$ of $G$. We let $T(0) := \{r\}$. Let $\langle v_1, \ldots, v_{n-1} \rangle$ be an enumeration of $V \setminus \{r\}$. Suppose $T(i)$ has been generated. Start a network random walk at $v_{i+1}$, and stop when it hits $T(i)$. Let

$$T(i+1) := T(i) \cup \mathrm{LE}(\text{path from } v_{i+1} \text{ to } T(i)).$$

The output of the algorithm is $T = T(n-1)$.

**Theorem 3.1.** *$T$ is distributed proportional to weight.*

*Proof.* To each $v \in V$, $v \neq r$, we assign an infinite stack of random oriented edges (arrows) $e_j^v$, $j = 1, 2, \ldots$, such that $\underline{e_j^v} = v$, and $\mathsf{P}[e_j^v = e] = C(e)/C_v$. All elements in all the stacks are independent. We colour the $i$-th element of each stack with colour $i$.

The arrows on top of the stacks define a random oriented graph with vertex set $V$. This may contain oriented cycles. If there are no oriented cycles, then, since every vertex $v \neq r$ has a unique outgoing edge, the graph is a tree directed towards the root $r$.

Consider the notion of *cycle popping*. If $C = \langle e_1, \ldots, e_k \rangle$ is an oriented cycle on top of the stacks, by popping $C$, we mean removing all the edges $e_j$, $j = 1, \ldots, k$ from their stacks, and lifting those stacks up by one level, so that now new edges sit of top of those stacks.

The collection of stacks provides a probability space on which the algorithm can be defined. Since the arrows at each vertex were chosen with the random walk transition probabilities, the random walk started at $v_1$ can follow the oriented edges. When a vertex is revisited for the first time, an oriented cycle has been found. If we pop this cycle from the stacks, then the revisited vertex receives a fresh, independent arrow, which can be followed to continue the random walk. It is easy to see that if every completed cycle is popped, then following the arrows on top of the current stacks is precisely the network random walk. Moreover, when $r$ is hit, on top of the stacks we see the LERW from $v_1$ to $r$, as well as "fresh" arrows (independent of previously visited ones) on top of the stacks of $v \notin T(1)$. Similarly, the LERW from $v_2$ correpsonds to following the arrows from $v_2$ and popping cycles as they are encountered. When the algorithm is complete, the algorithm outputs the tree on top of the stacks.

The above shows that the algorithm defines a particular way of popping cycles until a tree is uncovered. Note that each popped cycle is coloured (each of its oriented edges has a colour), and a particular coloured cycle can be popped at most once. We now prove that we may in fact pop cycles in any order we wish, and regardless of what we do, the same coloured cycles will be popped, and the same tree uncovered.

Let $C_1, \ldots, C_m$ be a sequence of oriented coloured cycles that can be popped (in this order), with a tree $t$ resulting. We know from the algorithm, that such a finite sequence exists with probability 1. Let $D_1, D_2, \ldots$ be any other possible sequence of coloured oriented cycles that can be popped. We show that the $D$-sequence consists of the same coloured cycles as the $C$-sequence. We prove this by induction on $m$. If $m = 0$, then there are no cycles to be popped, hence the $D$-sequence is also empty. Assume now that $m \geq 1$, and that the statement is true for $C$-sequences of length less than $m$. Let $D_i$ be the first cycle in the $D$-sequence that is not disjoint from $C_1$. Hence $D_i$ and $C_1$ share a vertex $w_1$. Since $D_1, \ldots, D_{i-1}$ are disjoint from $C_1$, they do not contain $w_1$, and hence the colour of $w_1$ is the same in $C_1$ and $D_i$. It follows that $w_1$ has the same successor $w_2$ in $C_1$ and $D_i$. Using again that $w_2$ does not occur in $D_1, \ldots, D_{i-1}$, we see that $w_2$ has the same colour in $C_1$ and $D_i$, and so on. We get that $C_1$ and $D_i$ are the same coloured cycle. Popping $D_i = C_1$ commutes with popping $D_2, \ldots, D_{i-1}$, hence popping the $D$-sequence has the same effect as popping $D_i = C_1, D_2, D_3, \ldots, D_{i-1}, D_{i+1}, \ldots$. Now pop $C_1$ from the top of the stacks. By the induction hypothesis, $C_2, \ldots, C_m$ consists of the same coloured cycles as $D_2, \ldots, D_{i-1}, D_{i+1}, \ldots$, and it uncoveres the same tree $t$. The claim follows.

We now show that the tree uncovered by cycle popping has probability proportional to weight. Let $C_1, \ldots, C_m$ be any sequence of coloured oriented cycles that can be popped in this order, and $t$ any tree rooted at $r$. The probability that $C_1$ can be popped is the probability that the arrows in $C_1$ all point "the right way", that is $\prod_{e \in C_1} C(e)/C_{\underline{e}}$. Using independence of the elements in the stacks, the conditional probability that $C_i$ can be popped, given that $C_1, \ldots, C_{i-1}$ can be popped is $\prod_{e \in C_i} C(e)/C_{\underline{e}}$. Finally, using again independence of the elements in the stacks, conditioned on all the cycles popped, the probability that the tree $t$ is uncovered, is $\prod_{e \in t} C(e)/C_{\underline{e}}$. Hence we have

$$\begin{aligned} &\mathsf{P}[C_1, \ldots, C_m \text{ can be popped and as a result } t \text{ is uncovered}] \\ &\quad = \prod_{i=1}^{m} \prod_{e \in C_i} \frac{C(e)}{C_{\underline{e}}} \times \prod_{e \in t} \frac{C(e)}{C_{\underline{e}}}. \end{aligned}$$

The last factor is const $\cdot$ weight$(t)$. Note that $t$ varies independently of the cycles: given any sequence of cycles popped, the tree uncovered may be any tree rooted at $r$. This shows more than we need: even if we condition on the sequence of cycles popped, the tree has the claimed distribution. $\square$

*Remark.* The proof showed not only that the result of the algorithm is independent of the sequence $v_1, \dots, v_n$, but also that we may, if we wish, pick $v_{i+1}$ depending on what happened in the algorithm up to that point. The proof also showed that the distribution of the number of random walk steps needed is independent of how $v_1, \dots, v_n$ was chosen. The proof in fact establishes a probability space on which any choice results in the same number of steps. This is called a *coupling* of the different instances of the algorithm.

*Remark.* The same proof works for any Markov chain in place of a network random walk. That is, if $G$ is the directed graph constructed from a Markov chain as in Section 2, the algorithm outputs a rooted tree $(t, r)$ proportional to $q(t, r) = \prod_{[v,w] \in t} p_{vw}$.

*Remark.* The expected number of Markov chain steps used by the algorithm can be found as follows. To find the expected number of times that a step out of $u$ is used, we may assume, by the proof, that $v_1 = u$. Then we need to find $\mathsf{E}[\#$ visits to $u$ before $\tau_r]$, where $\tau_r = \inf\{n \geq 0 : X(n) = r\}$. For any irreducible Markov chain, this expectation equals $\pi(u)[\mathsf{E}_u \tau_r + \mathsf{E}_r \tau_u]$, where $\pi$ is the stationary distribution. [2, Chapter 2]. Hence the expected running time of the algorithm (measured by the number of Markov chain steps) is

$$\sum_{u \neq r} \pi(u)[\mathsf{E}_u \tau_r + \mathsf{E}_r \tau_u].$$

Note that this is always at most $2\mathsf{E}C$, where $C$ is the cover time, and may be smaller.

*Remark.* Later we will see variants of Wilson's algorithm on infinite graphs give useful information. As a first example, consider the square lattice $G = (\mathbb{Z}^2, \mathbb{E}^2)$. Simple random walk is recurrent, that is, for all $v, w \in \mathbb{Z}^2$,

$$\mathsf{P}_v[X(n) = w \text{ for some } n \geq 1] = 1.$$

Pick $r \in \mathbb{Z}^2$, and let $v_1, v_2, \dots$ list all the vertices of $\mathbb{Z}^2$. Since $T(i)$ is hit with probability 1, the algorithm can be run, and results in a spanning tree of $(\mathbb{Z}^2, \mathbb{E}^2)$.

# 4 Electric networks and spanning trees

In this section we describe the relation between electrical networks, spanning trees and random walk. To a great extent, we follow the exposition in [5]. The goal is to define the notion of *current*, and use it to describe the joint probability that edges $e_1, \dots, e_k$ belong to a random spanning tree. Although the physical interpretations of "resistance", "current", etc. will not be needed, we note that a nice introduction to these and their connection with random walk can be found in [11].

## 4.1 The gradient and divergence operators

Let $(G, C)$ be a finite network. We denote by $\ell^2(V)$ the space of real-valued functions on $V$, with the inner product:

$$(f, g)_C := \sum_{v \in V} C_v f(v) g(v), \tag{5}$$

and norm $\|f\|_C$. We denote by $\ell^2_-(E)$ the space of *antisymmetric* functions on oriented edges, that is, functions $\theta : E \to \mathbb{R}$ satisfying $\theta(\check{e}) = -\theta(e)$ for all $e \in E$. We equip this space with the inner product:

$$(\theta, \theta')_R := \frac{1}{2} \sum_{e \in E} R(e)\theta(e)\theta'(e) = \sum_{e \in E_{1/2}} R(e)\theta(e)\theta'(e), \tag{6}$$

where $E_{1/2}$ contains each edge of $G$ with exactly one orientation. The *energy* of $\theta$ is

$$\mathcal{E}(\theta) := (\theta, \theta)_R = \|\theta\|_R^2.$$

The *gradient operator* is $\nabla : \ell^2(V) \to ell^2_-(E)$, defined by

$$(\nabla F)(e) := C(e)(F(\overline{e}) - F(\underline{e})). \tag{7}$$

The *divergence operator* is $\operatorname{div} : \ell^2_-(E) \to \ell^2(V)$, defined by

$$(\operatorname{div} \theta)(v) := \frac{1}{C_v} \sum_{e : \underline{e} = v} \theta(e). \tag{8}$$

$-\nabla$ and div are adjoints of each other: $(\theta, -\nabla F)_R = (\operatorname{div} \theta, F)_C$. To see this, write

$$(\theta, -\nabla F)_R = \frac{1}{2} \sum_{e \in E} R(e)\theta(e)C(e)(F(\underline{e}) - F(\overline{e})) = \frac{1}{2} \sum_{e \in E} \theta(e)(F(\underline{e}) - F(\overline{e})).$$

Since $\theta(e)(-F(\overline{e})) = \theta(\check{e})F(\underline{\check{e}})$, this can be written as

$$\sum_{e\in E}\theta(e)F(\underline{e}) = \sum_{e\in E}\sum_{v\in V} I[\underline{e}=v]F(v)\theta(e) = \sum_{v\in V} C_v F(v)\frac{1}{C_v}\sum_{e\in E} I[\underline{e}=v]\theta(e)$$
$$= \sum_{v\in V} C_v F(v)\operatorname{div}\theta(v).$$

To motivate what follows, imagine that the network $(G, C)$ is an electric network, where edge $e$ is a resitor with resistance $R(e)$. Suppose that we hook up a battery between the two endpoints of the edge $e$, and suppose that a unit of current flows through the battery. Let $I^e(f)$ be the amount of current that flows along the edge $f$. How to determine $I^e(f)$? We know that current is conserved at each vertex $v \neq \underline{e}, \overline{e}$. Hence, $\operatorname{div} I^e(v) = 0$. A unit of current comes in at $\underline{e}$, and a unit of current is taken out at $\overline{e}$, so

$$\operatorname{div} I^e = \frac{1}{C_{\underline{e}}}\mathbf{1}_{\underline{e}} - \frac{1}{C_{\overline{e}}}\mathbf{1}_{\overline{e}}. \tag{9}$$

This does not uniquely specify $I^e$, however, there is a physical principle called Thompson's principle that states that $I^e$ has minimal energy among all flows that have divergence equal to (9). Here we will adopt this characterization as the mathematical defintion of the current in (10) below.

## 4.2 The gradient and cycle spaces

We define the unit flow along the edge $e$: $\chi^e \in \ell^2_-(E)$ defined by $\chi^e := \mathbf{1}_e - \mathbf{1}_{\check{e}}$, where $\mathbf{1}_f$ denotes the function that is 1 on the edge $f$ and zero elsewhere. Note that this has the required divergence (9). Let Gr denote the space of gradients:

$$\mathrm{Gr} := \nabla\ell^2(V).$$

The flow $-\nabla\mathbf{1}_v = \sum_{e;\underline{e}=v}\chi^e$ is called the *star* at $v$. The space Gr is spanned by the stars. If $e_1, \ldots, e_k$ is an oriented cycle in $G$, then the flow $\sum_{i=1}^k \chi^{e_i} \in \ell^2_-(E)$ is called a *cycle*. We call

$$\mathrm{Cyc} := \text{linear span of all cycles} \subset \ell^2_-(E).$$

**Lemma 4.1.** *We have* $\ell^2_-(E) = \mathrm{Gr} \oplus \mathrm{Cyc}$.

*Proof.* Stars are orthogonal to cycles. For this note that if $v$ is not on the cycle $e_1, \ldots, e_k$, then the star at $v$ and this cycle are supported on disjoint edges, and hence are orthogonal. When $v = \overline{e_i} = \underline{e_{i+1}}$, then the inner product is $R(e_i)(-C(e_i)) + R(e_{i+1})C(e_{i+1}) = 0$ (note that if $v$ occurs multiple times in the cycle, a similar calculation remains valid). Therefore, what we need to show is that if $\theta$ is orthogonal to Cyc, then it is a gradient. This goes by a well-known argument. Fix $o \in V$, and for any $v \in V$ let $e_1, \dots, e_k$ be any path in $G$ from $o$ to $v$. Define $F(v) := \sum_j R(e_j)\theta(e_j)$. This definition is independent of the path chosen, since if $e'_1, \dots, e'_l$ is another path, then $\theta$ is orthogonal to the cycle $e_1, \dots, e_k, \check{e}'_l, \dots, \check{e}'_1$. It is easy to see that $\nabla F = \theta$, and this completes the proof. □

For a subspace $Z$ of $\ell^2_-(E)$, let $P_Z$ be the orthogonal projection onto $Z$, and let $P_Z^\perp$ be the orthogonal projection onto the orthogonal complement of $Z$.

We are ready to define

$$I^e := P_{\mathrm{Gr}}\chi^e. \tag{10}$$

Since $I^e - \chi^e \perp \mathrm{Gr}$, we have, for any $F$ $0 = (I^e - \chi^e, \nabla F)_R = -(\operatorname{div} I^e - \operatorname{div} \chi^e, F)_C$, which is equivalent to $\operatorname{div} I^e = \operatorname{div} \chi^e$. Hence $I^e$ indeed has smallest energy among flows with divergence (9).

## 4.3 Connection with spanning trees

Let $e$ and $f$ be oriented edges of $G$. We define

$$\beta(e, f) = \mathsf{P}[\text{path from } \underline{e} \text{ to } \overline{e} \text{ in random spanning tree uses } f].$$

Start a network random walk at $\underline{e}$, and stop when it hits $\overline{e}$. Let $J^e(f)$ be the net number of times edge $f$ is used, that is

$$J^e(f) := \mathsf{E}[\#\text{ times } f \text{ is used} - \#\text{ times } \check{f} \text{ is used}].$$

**Theorem 4.1.** *We have*

$$\beta(e, f) - \beta(e, \check{f}) = J^e(f) = I^e(f).$$

*In particular,*

$$\mathsf{P}[e \in T] = \mathsf{P}_{\underline{e}}[\textit{first hit } \overline{e} \textit{ via the edge } e] = I^e(e). \tag{11}$$

*Remark.* The equality of the spanning tree quantity and the current is due to Kirchhoff [19], and the equality of the random walk quantity and the current is due to Doyle and Snell [11].

*Proof.* By Wilson's algorithm, that path in the random spanning tree is a LERW from $\underline{e}$ to $\overline{e}$. Hence

$$\beta(e, f) - \beta(e, \check{f}) = \mathsf{E}[\#\text{ times LERW uses } f - \#\text{ times LERW uses } \check{f}]. \tag{12}$$

A cycle is traversed an equal number of times in expectation in both directions, since if $e_1, \dots, e_k$ is a cycle with $v = \underline{e_1}$, then

$$\begin{aligned} &\mathsf{P}[e_1, \dots, e_k \text{ are traversed} \,|\, X(n) = v] \\ &\quad = \frac{C(e_1)}{C_{\underline{e_1}}} \frac{C(e_2)}{C_{\underline{e_2}}} \cdots \frac{C(e_k)}{C_{\underline{e_k}}} \\ &\quad = \mathsf{P}[\check{e_k}, \dots, \check{e_1} \text{ are traversed} \,|\, X(n) = v]. \end{aligned} \tag{13}$$

This imples that adding the loops of the random walk in the right hand side of (12), does not change the expected net number of times $f$ is used, hence the $\beta(e, f) - \beta(e, \check{f}) = J^e(f)$.

To prove the second equality, let

$$F(v) := \mathsf{E}_{\underline{e}}[\# \text{ visits to } v \text{ up to } \tau_{\overline{e}}].$$

Since for every $v \neq \underline{e}, \overline{e}$, any incoming step to $v$ is balanced by an outgoing step from $v$, we have $\operatorname{div} J^e(v) = 0$. At $v = \underline{e}$ there is one more outgoing step than incoming step, and at $v = \overline{e}$, there is one incoming step and no outgoing step. Hence $\operatorname{div} J^e = (1/C_{\underline{e}})\mathbf{1}_{\underline{e}} - (1/C_{\overline{e}})\mathbf{1}_{\overline{e}} = \operatorname{div} I^e$. Hence, $J^e - I^e \perp \mathrm{Gr}$, and it is enough to show that $J^e \in \mathrm{Gr}$. Let

$$\begin{aligned} \theta_v &= \mathsf{P}_v[\text{first step of r.w. uses } f] - \mathsf{P}_v[\text{first step of r.w. uses } \check{f}] \\ &= -\frac{1}{C_v}\mathbf{1}_v \in \mathrm{Gr}. \end{aligned}$$

Then it is easy to check that $J^e = \sum_v F(v)\theta_v \in \mathrm{Gr}$. It follows that $J^e = I^e$, and the proof is complete. ☐

**Exercise 4.1.** Complete the argument regarding loops being traversed an equal number of times on average. Condition on the loop-erasure of the random walk path from $\underline{e}$ to $\overline{e}$, and decompose the random walk path into its loop-erasure $v_0 = \underline{e}, v_1, \dots, v_N = \overline{e}$, and the oriented cycles $C_0, C_1, \dots, C_{N-1}$, where $C_i$ is a cycle based at $v_i$. Further condition on the cycles without their orientation, and deduce from (13) that the two orientations are equally likely. Deduce that the expectation of the quantity defining $J^e(f)$ is equal to $\beta(e, f) - \beta(e, \check{f})$.

The matrix $Y(e, f) = I^e(f) = \frac{1}{R(e)}(I^e, \chi^f)_R = C(e)(P_{\mathrm{Gr}}\chi^e, \chi^f)_R$ is called the *transfer current matrix*.

## 4.4 Contracting edges in a network

**Definition 4.1.** Let $G = (V, E)$ be a graph, $F \subset E$. We denote by $G/F$ the graph obtained by identifying, for every $f \in F$, the endpoints of $f$. We identify the edges of $G$ with those of $G/F$, where some edges in $G$ have become loops in $G/F$. Hence we also have an identification of the space $\ell^2_(E)$ for the two graphs.

Let us write $T_G$ for the random spanning tree of $G$.

**Proposition 4.1.** *Assume that $G$ is a finite connected network, and that no cycle can be formed of the edges in $F$. Then $T_G$ conditioned on $F \subset T_G$ has the same distribution as $T_{G/F} \cup F$.*

*Proof.* For any $C \subset E$, $C \cap F = \emptyset$, we have that $C \cup F$ contains a cycle of $G$ if and only if $C$ contains a cycle of $G/F$. Hence $C \cup F$ is a spanning tree of $G$ if and only if $C$ is a spanning tree of $G/F$. Since weight$(C \cup F) = \prod_{e \in C} C(e) \prod_{f \in F} C(f)$, we get the statement. □

The above proposition shows that if we condition on a certain set of edges $F$ to be present in $T_G$, the conditional probabilities are given by the contracted network $G/F$. Hence we would like to see how the current $I^e$ changes when we contract edges in a network. This will turn out to be given by applying an orthogonal projection to $I^e$.

Recall that we identify edges in $G$ and edges in $G/F$ (where the edges in $F$ became loops in $G/F$). This gives an identification of the spaces $\ell^2_-(E)$ for the two graphs.

Let $\widehat{\mathrm{Gr}}$ denote the gradients in $G/F$, and let $\widehat{\mathrm{Cyc}}$ denote the linear span of cycles in $G/F$.

**Lemma 4.2.**

$$\widehat{\mathrm{Cyc}} = \mathrm{Cyc} + \langle \chi^F \rangle,$$

*where $\langle \chi^F \rangle$ is short for the linear span of $\{\chi^f : f \in F\}$.*

*Proof.* It is clear that cycles in $G$ remain cycles in $G/F$, so Cyc is a subspace of $\widehat{\mathrm{Cyc}}$. Since edges in $F$ become loops in $G/F$, it is also clear that $\langle \chi^F \rangle$ is a subspace of $\widehat{\mathrm{Cyc}}$. This shows that the right hand side is contained in the left hand side. Suppose now that $e_1, \dots, e_k$ form a cycle in $G/F$. We show that we can insert edges from $F$ into this sequence to get a cycle in $G$, which will show that $\widehat{\mathrm{Cyc}}$ is contained in the right hand side. If $\overline{e_{i-1}} = \underline{e_i}$ in the graph $G$, then there is no need to insert an edge between $e_{i-1}$ and $e_i$. If $\overline{e_{i-1}} = v_1 \neq v_2 = \underline{e_i}$ in $G$, then $v_1$ and $v_2$ belong to the same connected component of $F$ (since they are contracted to the same vertex in $G/F$). Hence we can insert edges $f_1, \dots f_{k(i)} \in F$ to get a path from $v_1$ to $v_2$. □

Now write

$$\ell^2_-(E) = \mathrm{Gr} \oplus \mathrm{Cyc} = \widehat{\mathrm{Gr}} \oplus \widehat{\mathrm{Cyc}},$$

where $\widehat{\mathrm{Cyc}} \supset \mathrm{Cyc}$, and consequently $\widehat{\mathrm{Gr}} \subset \mathrm{Gr}$. We can write $\mathrm{Gr} = \widehat{\mathrm{Gr}} \oplus (\mathrm{Gr} \cap \widehat{\mathrm{Cyc}})$ and $\widehat{\mathrm{Cyc}} = (\mathrm{Gr} \cap \widehat{\mathrm{Cyc}}) \oplus \mathrm{Cyc}$. Hence

$$\ell^2_-(E) = \widehat{\mathrm{Gr}} \oplus (\mathrm{Gr} \cap \widehat{\mathrm{Cyc}}) \oplus \mathrm{Cyc}.$$

We show that the middle subspace equals $Z := P_{\mathrm{Gr}}\langle \chi^F \rangle$. For this note that

$$\mathrm{Gr} \cap \widehat{\mathrm{Cyc}} = P_{\mathrm{Gr}}\widehat{\mathrm{Cyc}} = P_{\mathrm{Gr}}\mathrm{Cyc} + P_{\mathrm{Gr}}\langle \chi^F \rangle = Z.$$

Hence

$$\ell^2_-(E) = \widehat{\mathrm{Gr}} \oplus Z \oplus \mathrm{Cyc}.$$

Let now $e$ be an edge that does not form a cycle together with edges from $F$ (so that it is not a loop in $G/F$). Then writing $\widehat{I}^e$ for the current in $G/F$, we have

$$\widehat{I}^e = P_{\widehat{\mathrm{Gr}}}\chi^e = P_Z^\perp P_{\mathrm{Gr}}\chi^e = P_Z^\perp I^e. \tag{14}$$

## 4.5 The transfer current theorem

The expression (14) allows us to prove the following beautiful theorem on the joint probability that edges $e_1, \ldots, e_k$ belong to the random spanning tree. The theorem is due to [8], the proof we present is from [5].

**Theorem 4.2** (Transfer current theorem)**.** *[Burton, Pemantle; 1993] Let $G$ be a finite connected network. For distinct edges $e_1, \ldots, e_k \in E$, we have*

$$\mathsf{P}[e_1, \ldots, e_k \in T_G] = \det[Y(e_i, e_j)]_{1 \le i,j \le k},$$

*where $Y(e, f) = I^e(f)$.*

*Proof.* The left hand side is 0, if a cycle can be formed from the edges $e_1, \ldots, e_k$, and we need to show that in this case the determinant is also 0. Write the cycle as $\sum_j \alpha_j \chi^{e_j} \in \mathrm{Cyc}$ (where $\alpha_j \in \{-1, 0, 1\}$). Then the following linear combination of columns of the matrix $Y$ vanishes:

$$\begin{aligned} \sum_j \alpha_j R(e_j) Y(e_i, e_j) &= \sum_j \alpha_j R(e_j) I^{e_i}(e_j) \chi^{e_j}(e_j) = \sum_j \alpha_j (I^{e_i}, \chi^{e_j})_R \\ &= (I^{e_i}, \sum_j \alpha_j \chi^{e_j})_R = (P_{\mathrm{Gr}}\chi^{e_i}, \sum_j \alpha_j \chi^{e_j})_R = 0. \end{aligned}$$

The last equality follows, since gradients are orthogonal to cycles. Therefore, the determinant is also 0.

Assume now that no cycle can be formed from the edges $e_1, \dots, e_k$. We write, using that $P_{\mathrm{Gr}} = P_{\mathrm{Gr}}^2$, and that $P_{\mathrm{Gr}}$ is self-adjoint:

$$\begin{aligned} Y(e,f) = I^e(f) &= \frac{1}{R(f)}(I^e, \chi^f)_R = C(f)(I^e, \chi^f)_R \\ &= C(f)(P_{\mathrm{Gr}}\chi^e, \chi^f)_R = C(f)(P_{\mathrm{Gr}}\chi^e, P_{\mathrm{Gr}}\chi^f)_R \\ &= C(f)(I^e, I^f)_R. \end{aligned} \tag{15}$$

Hence

$$\det[Y(e_i, e_j)] = \prod_{i=1}^{k} C(e_i) \det Y_k, \tag{16}$$

where $Y_k(e_i, e_j) = (I^{e_i}, I^{e_j})_R$. The matrix $Y_k$ has the following structure: we have $k$ vectors, $I^{e_1}, \dots, I^{e_k}$, and the $(i,j)$ element is the inner product of the $i$-th and $j$-th vectors. Such a matrix is called a *Gram matrix*. The determinant of a gram matrix equals the squared volume of the paralelepiped spanned by its determining vectors. This is easy to see if the determining vectors are pairwise orthogonal, as then the matrix is diagonal with diagonal entries equal to the squared length of the determining vectors. When the vectors are not orthogonal, we can applied the Gram-Schmidt orthogonalization process to them, and it is not hard to check that the determinant does not change, and neither the squared volume of the paralelepiped spanned by the vectors. Therefore, we can write (16) as

$$\prod_{i=1}^{k} C(e_i) \| P_{Z_i}^{\perp} I^{e_i} \|_R^2, \tag{17}$$

where

$$Z_i = \mathrm{span}\{I^{e_1}, \dots, I^{e_{i-1}}\} = P_{\mathrm{Gr}}(\mathrm{span}\{\chi^{e_1}, \dots, \chi^{e_{i-1}}\}).$$

Now we have

$$\begin{aligned} \mathsf{P}[e_1, \dots, e_k \in T_G] &= \prod_{i=1}^{k} \mathsf{P}[e_i \in T_G \mid e_1, \dots, e_{i-1} \in T_G] \\ &\overset{\text{Prop. 4.1}}{=} \prod_{i=1}^{k} \mathsf{P}[e_i \in T_{G/\{e_1,\dots,e_{i-1}\}}] \overset{(11)}{=} \prod_{i=1}^{k} \widehat{I}^{e_i}(e_i) \\ &\overset{(15)}{=} \prod_{i=1}^{k} C(e_i)(\widehat{I}^{e_i}, \widehat{I}^{e_i})_R \overset{(14)}{=} \prod_{i=1}^{k} C(e_i) \| P_{Z_i}^{\perp} I^{e_i} \|_R^2. \end{aligned}$$

The last expression is the same as (17), which completes the proof. ∎

*Remark.* A 0–1 valued process $\{X_\alpha\}_{\alpha\in I}$ is called *determinantal* if

$$\mathsf{P}[X_{\alpha_1},\dots,X_{\alpha_k}=1]=\det[K(\alpha_i,\alpha_j)]_{1\le i,j,\le k}$$

for some kernel $K$.

## 4.6 Monotonicity properties of currents

We need some monotnicity properties, that will be very important when we look at limits of spanning trees on infinite graphs.

Let $(G,C)$ be a finite network and $e\in E$. When we need to indicate that a current is computed in a network $H$, we write $I^e_H$.

**Proposition 4.2.** *(a) If $G'$ is a subgraph of $G$ containing $e$, then $I^e_{G'}(e)\ge I^e_G(e)$.*
*(b) If $F\subset E$ and $F\cup\{e\}$ has no cycles containing $e$, then $I^e_{G/F}(e)\le I^e_G(e)$.*

*Proof.* In both cases, the inequality results from the fact that one of the currents can be written as an orthogonal projection of the other, that reduces the norm.

(a) Write $G'=(V',E')$. Since $E'\subset E$, there is a natural embedding of $\ell^2_-(E')$ into $\ell^2_-(E)$, by setting $\theta(e)=0$ for all $e\in E\setminus E'$. We have $\ell^2_-(E')=\mathrm{Gr}'\oplus\mathrm{Cyc}'$, where $\mathrm{Cyc}'\subset\mathrm{Cyc}$ (any cycle in $G'$ is also a cycle in $G$). We have $\chi^e\in\ell^2_-(E')\subset\ell^2_-(E)$. Since $I^e_{G'}$ is the projection of $\chi^e$ onto $\mathrm{Gr}'$, we can write $\chi^e=I^e_{G'}+f'$, with $f'\in\mathrm{Cyc}'$. Then

$$I^e_G=P_{\mathrm{Gr}}\chi^e=P_{\mathrm{Gr}}I^e_{G'}+P_{\mathrm{Gr}}f'=P_{\mathrm{Gr}}I^e_{G'}.$$

Hence

$$\begin{aligned}I^e_G(e)&\overset{(15)}{=}C(e)(I^e_G,I^e_G)_R=C(e)\|I^e_G\|^2_R=C(e)\|P_{\mathrm{Gr}}I^e_{G'}\|^2_R\le C(e)\|I^e_{G'}\|^2_R\\&=I^e_{G'}(e).\end{aligned}$$

(b) We have $I^e_{G/F}=P^\perp_Z I^e_G$. Hence, similarly to (a):

$$I^e_{G/F}(e)=C(e)\|I^e_{G/F}\|^2_R=C(e)\|P^\perp_Z I^e_G\|^2_R\le C(e)\|I^e_G\|^2_R=I^e_G(e).$$

□

Proposition 4.2 has the following consequence for random spanning trees.

**Proposition 4.3.** *Let $G = (V, E)$ be a finite connected network, $F \subset E$.*
*(a) If $G'$ is a subgraph of $G$ containing $F$, then*

$$\mathsf{P}[F \subset T_{G'}] \geq \mathsf{P}[F \subset T_G].$$

*(b) If $e \neq f$ then*

$$\mathsf{P}[f \in T_G \,|\, e \in T_G] \leq \mathsf{P}[f \in T_G].$$

*More generally, if $F \cap F' = \emptyset$, then*

$$\mathsf{P}[F \subset T_{G/F'}] \leq \mathsf{P}[F \subset T_G].$$

*Proof.* (a) We may assume that $F$ contains no cycles. We induct on $|F|$. If $F = \{e\}$, then by (11) and Proposition 4.2(a),

$$\mathsf{P}[e \in T_{G'}] = I^e_{G'}(e) \geq I^e_G(e) = \mathsf{P}[e \in T_G].$$

Induction step: $F = F_1 \cup \{e\}$. Then $G'/F_1$ is a subgraph of $G/F_1$. Therefore,

$$\begin{aligned}\mathsf{P}[F \subset T_{G'}] &= \mathsf{P}[e \in T_{G'/F_1}]\mathsf{P}[F_1 \subset T_{G'}] \geq \mathsf{P}[e \in T_{G/F_1}]\mathsf{P}[F_1 \subset T_G] \\ &= \mathsf{P}[F \subset T_G].\end{aligned}$$

(b) We have

$$\mathsf{P}[f \in T_G | e \in T_G] = \mathsf{P}[f \in T_{G/e}] = I^f_{G/e}(f) \leq I^f_G(f) = \mathsf{P}[f \in T_G].$$

General case follows by induction on $|F|$. If $F = \{e\}$, then

$$\mathsf{P}[e \in T_{G/F'}] = I^e_{G/F'}(e) \leq I^e_G(e) = \mathsf{P}[e \in T_G].$$

For the induction step, write $F = F_1 \cup \{e\}$. Note that $G/F'/F_1$ is a contraction of $G/F_1$. Therefore,

$$\begin{aligned}\mathsf{P}[F \subset T_{G/F'}] &= \mathsf{P}[e \in T_{G/F'/F_1}]\mathsf{P}[F_1 \subset T_{G/F'}] \\ &\leq \mathsf{P}[e \in T_{G/F_1}]\mathsf{P}[F_1 \subset T_G] = \mathsf{P}[F \subset T_G].\end{aligned}$$

□

# 5 Random spanning forests of infinite graphs

Some properties of random spanning trees become especially transparent when we consider the limit of an infinite graph. As an analogy, think about simple random walk, and how the phenomenon of recurrence/transience is most naturally expressed as a property of an infinite walk (although it could be formulated as a limiting property of finite walks).

It is not obvious, how to define a "uniform" spanning tree on an infinite graph, and we will encounter some surprises. The first is that there will be more than one natural way to define the limit, and the results can be different. The second is that in the context of infinite graphs, the natural objects will be spanning forests, rather than spanning trees.

To construct the limiting objects on infinite graphs, we exploit the monotnicity properties proved in the previous section. For this section, $G = (V, E)$ is an infinite connected network.

## 5.1 Measurable space

We will work on the space $\Omega = \{0, 1\}^E$. This is a compact metric space in the product topology. The Borel $\sigma$-algebra is generated by the *elementary cylinders*, that is, sets of the form

$$A_{B,K} = \{F \subset E : F \cap K = B\},$$

where $B \subset K$ are finite sets of edges. The event $A_{B,K}$ expresses that the edges in $B$ are present and the edges in $K \setminus B$ are not.

## 5.2 Exhaustion

Consider $V_1 \subset V_2 \subset \cdots \subset V$, such that $\cup_n V_n = V$, $V_n$ is finite. Let $G_n = (V_n, E_n)$ be the subgraph induced by $V_n$. The sequence $\langle G_n \rangle$ is called an *exhaustion* of $G$ by finite subgraphs.

## 5.3 Free spanning forest

Let $\mu_n^F$ be the random spanning tree measure on $G_n$ (this measure can be realized on $\Omega$).

Since $G_n$ is a subgraph of $G_{n+1}$, Proposition 4.3 implies that for any fixed $B \subset E$ finite we have

$$\mu_n^F[B \subset T] \geq \mu_{n+1}^F[B \subset T].$$

Note that this makes sense for large enough $n$, as then $B \subset E_n$. Hence we can define

$$\mu^F[B \subset T] := \lim_{n\to\infty} \mu_n^F[B \subset T], \tag{18}$$

where the limit exists by monotonicity.

**Exercise 5.1.** Show that if $B \subset K$ are fixed finite sets of edges, then

$$\mu^F[T \cap K = B] := \lim_{n\to\infty} \mu_n^F[T \cap K = B],$$

exists. *Hint:* use the inclusion-exclusion principle to reduce the statement to (18).

By the result of the exercise, $\mu^F$ is defined on all elementary cylinders. By Kolmogorov's extension theorem, $\mu^F$ has a unique extension to a measure on $\Omega$.

**Exercise 5.2.** Show that $\mu_n^F \Rightarrow \mathrm{FSF}$, in the sense of weak convergence of probability measures on $\Omega$.

The limit does not depend on the exhaustion: if $\langle G'_n \rangle$ is another exhaustion, we can find

$$V_{n_1} \subset V'_{n'_1} \subset V_{n_2} \subset V'_{n'_2} \subset \dots$$

and the limit along this third exhaustion has to coincide with both the limit along $\langle G_n \rangle$ and the limit along $\langle G'_n \rangle$.

The measure $\mu^F$ is called the *(weighted) free spanning forest* measure on $G$, we will denote it FSF. We will write $\mathcal{F}$ for the set of edges present in an element of $\Omega$. Then $\mathrm{FSF}[\mathcal{F} \text{ contains a cylce}] = 0$, since any specific cycle being present is a cylinder event that has 0 probability under any $\mu_n^F$, and hence also under FSF. It is obvious that $\mathrm{FSF}[\mathcal{F} \text{ is spanning}] = 1$. Hence $\mathrm{FSF}[\mathcal{F} \text{ is a spanning forest}] = 1$.

## 5.4 Wired spanning forest

The word "free" refers to the fact how in defining $\mu_n^F$, we disconnected $G_n$ from the rest of the graph $G$. There is another natural way of taking the limit, that is to force all connections outside of $G_n$ to occur.

Define the graph $G_n^W$ as the result of identifying all vertices in $V \setminus V_n$ to a single vertex $z_n$. This will result in infitely many loop edges at $z_n$, which we omit, as they cannot be part of a spanning tree. Hence another description of the graph $G_n^W$ is that we add to the graph $G_n$ a new vertex $z_n$, and for each edge of $G$ connecting a vertex $v \in V_n$ with a vertex in $V \setminus V_n$, we place an edge between $v$ and $z_n$.

Let $\mu_n^W$ be the random spanning tree measure on $G_n^W$ (also realized on $\Omega$). Now $G_n^W$ is obtained by contracting some edges in $G_{n+1}^W$. Therefore, by Proposition 4.3, we have

$$\mu_n^W[B \subset T] \leq \mu_{n+1}^W[B \subset T],$$

where this makes sense as soon as $B \subset E_n$. We again conclude that

$$\mu^W[B \subset T] := \lim_{n \to \infty}[B \subset T]$$

exists, and similalry for $\mu^W[T \cap K = B]$. The limit again does not depend on the exhaustion. The measure defined this way is called the *(weighted) wired spanning forest* measure, and we call it WSF. It is again concentrated on spanning forests of $G$.

The name "wired" refers to how the complement of $G_n$ has been short-circuited in the graph $G_n^W$.

## 5.5 Examples

Let us pause a little to consider some simple examples of the constructions above.

Consider the $d$-dimensional integer lattice, that is the graph with vertex set $\mathbb{Z}^d$, where $\{x, y\}$ is an edge if and only if $|x - y| = 1$. We set $C(e) = 1$ for all edges. Let $V_n = [-n, n]^d \cap \mathbb{Z}^d$, and let $G_n = (V_n, E_n)$ be the subgraph induced by $V_n$. We considered the random spanning tree measure (in this case this is the uniform spanning tree measure) $\mu_n^F$, and showed that it converges weakly to a measure FSF. This was first proved by Pemantle [30]. The word "free" refers to boundary condition we used by disconnecting $G_n$ from the rest of the lattice. We also considered a different boundary condition, called "wired", where we obtained the graph $G_n^W$ by adding a new vertex $z_n$ to $G_n$ and for every pair $v \in V_n$, $w \in V \setminus V_n$, $v \sim w$, we place an edge between $v$ and $z_n$. The random spanning tree measure on $G_n^W$ was $\mu_n^W$, and we showed it has a weak limit WSF. We will see later that on $\mathbb{Z}^d$, and on many other graphs, we have FSF = WSF. That is, in the case of $\mathbb{Z}^d$, the difference in boundary conditions washes away as $n \to \infty$.

When are the two measures different? Here is a simple example where they are. Let $G = (V, E)$ be a 3-regular tree, $o$ a fixed vertex in $V$, and $V_n = \{x \in V : \text{dist}(o, x) \leq n\}$. Since $G_n$ is a tree, its only spanning tree is itself, and hence $\mu_n^F$ concentrates on the single point $\{G_n\}$. It follows that FSF concentrates on the single point $\{G\}$. The limit is more interesting for the wired boundary condition. Let $a$ be a neighbour of $o$, and use Wilson's method to generate the random spanning tree on $G_n^W$, starting with the vertices $o$ and $a$. Since a random walk started at $o$ has probability 2/3 to step further from $o$ and probability 1/3 to step closer to $o$, whenever not at $o$, we have

$$\mathsf{P}_o[X_k \text{ hits } z_n \text{ before } a] \geq \alpha > 0$$

uniformly in $n$, and similarly,

$$\mathsf{P}_a[X_k \text{ hits } z_n \text{ before } o] \geq \alpha > 0.$$

It follows that

$$\mathrm{WSF}[\overline{oa} \notin \mathcal{F}] = \lim_n \mu_n^W[\overline{oa} \notin \mathcal{F}] \geq \alpha^2 > 0.$$

Therefore, WSF $\neq$ FSF. Later we will see that under WSF, there are infinitely many trees a.s.

We will also see later that independent random walks on $\mathbb{Z}^d$ with $d \geq 5$ have a positive probability of never intersecting, and we will be able to conclude that in this case there are also infinitely many trees a.s.

## 5.6 Wilson's method on transient networks

Assume that $(G, C)$ is an infinite network on which the network random walk is transient. The following variant of Wilson's method was introduced in [5].

We construct a random spanning forest of $G$. Let $\mathcal{F}_0 := \emptyset$, and let $v_1, v_2, \dots$ be an enumeration of $V$. We inductively define $\mathcal{F}_n$ as follows. Start a network random walk at $v_n$, and stop it if it hits $\mathcal{F}_{n-1}$, otherwise run it indefinitely. Let $\mathcal{P}_n$ be the path of this walk. Due to transience, $\mathcal{P}_n$ visits any vertex finitely often. Hence $\mathrm{LE}(\mathcal{P}_n)$ is well-defined. Set $\mathcal{F}_n := \mathcal{F}_{n-1} \cup \mathrm{LE}(\mathcal{P}_n)$. Finally, let $\mathcal{F} := \cup_n \mathcal{F}_n$. It is clear that $\mathcal{F}$ contains no cycles, and it spans $G$, so $\mathcal{F}$ is a random spanning forest of $G$. We call this process *Wilson's method rooted at infinity.*

**Lemma 5.1.** *[5] The distribution of $\mathcal{F}$ does not depend on the chosen enumeration of $V$.*

This lemma can be proved similarly to the finite case. Its statement also follows from Theorem 5.1 below, however, the proof gives more: it shows that one can realize the method with any choice of enumeration on the same probability space in such a way that the same random forest is generated.

*Proof.* Consider stacks and cycle popping as in the finite case. Fix all the stacks. Let us call a sequence $D_1, D_2, \dots$ of coloured cycles *legal*, if (i) the cycles can be popped in this sequence; (ii) if at any stage an (uncoloured) cycle is present then it is popped at some later stage. We say that a legal sequence *terminates*, if any finite $F \subset E$, becomes void of cycles eventually. For a legal, terminating sequence, the end result of popping all the cycles $D_1, D_2, \dots$ is well-defined, and is a spanning forest.

A fixed choice of $v_1, v_2, \dots$ yields, almost surely, a legal, terminating sequence $C_1, C_2, \dots$ of coloured cycles. Fix the stacks so that this holds. Then we show that any other other legal sequence $D_1, D_2, \dots$ is necessarily also terminating, and results in the same spanning forest. Let $D_{i_1}$ be the first cycle in the $D$-sequence that is not disjoint from $C_1$. As in the finite case, we see that $C_1 = D_{i_1}$ as coloured cycles. Popping $D_{i_1}$ commutes with popping $D_1, \dots, D_{i_1-1}$, hence $D' = \langle D_{i_1}, D_1, D_2, \dots, D_{i_1-1}, D_{i_1+1}, \dots \rangle$ is also legal. Pop now $C_1 = D_{i_1}$ from the stacks. Repeating the argument, we see that there exists $i_2 \in \{1, 2, \dots\} \setminus \{i_1\}$, such that $C_2 = D_{i_2}$ as coloured cycles, and that moving $D_{i_2}$ to the beginning, we get a legal sequence. Inductively, we find $i_k \in \{1, 2, \dots\} \setminus \{i_1, \dots, i_{k-1}\}$, such that $C_k = D_{i_k}$ as coloured cycles. We show that $\{i_1, i_2, \dots\} = \{1, 2, \dots\}$. Consider $D_k$. The $C$-sequence guarantees that the uncoloured cycle $D_k$ is not present after popping $C_1, \dots, C_N$ for some $N$. Hence, $k \in \{i_1, \dots, i_N\}$. We have thus proved that the $C$- and $D$-sequences consist of the same coloured cycles. It follows that $D_1, D_2, \dots$ is terminating, and results in the same spanning forest. The result follows. □

We now identify the distribution of the resulting spanning forest.

**Theorem 5.1.** *[5] On any transient network, Wilson's method rooted at infinity yields a spanning forest with distribution* WSF.

*Proof.* If $\mathcal{P} = \langle x_k : k \geq 0 \rangle$ is a deterministic path that visits every vertex finitely often, then

$$\mathrm{LE}(\langle x_k : k \leq K \rangle) \xrightarrow{K \to \infty} \mathrm{LE}(\mathcal{P}).$$

The meaning of convergence here is that for any $i$, the $i$-th step of the path on the left hand side is the same for all large enough $K$, and coincides with the $i$-th step on the right hand side. If $\mathcal{P}$ is a random walk path, almost surely,

$$\mathrm{LE}(\langle X_k : k \leq K \rangle) \xrightarrow{K \to \infty} \mathrm{LE}(\mathcal{P}),$$

by transience.

Recall that $G_n^W$ is obtained from $G$ by identifying $V \setminus V_n$ to a single vertex $z_n$. Consider the network random walk on $G$ up to the hitting time of $V \setminus V_n$. This is identical in law to the network random walk on $G_n^W$ up to the hitting time of $z_n$, hence we can use the network random walk on $G$ for Wilson's algorithm on $G_n^W$ with root $z_n$.

We write $T(n)$ for the random spanning tree on $G_n^W$, and we write $\mathcal{G}$ for its limit, the wired spanning forest on $G$. Fix $e_1, \dots, e_M \in E$. Let $u_1, \dots, u_L$ contain all vertices that are endpoints of some $e_j$. Let $\langle X_k(u_j) \rangle$ be the random walk started at $u_j$. Wilson's method on $G_n^W$, rooted at $z_n$, uses the stopping times $\tau_j^n$ that are the first time when the earlier part of the spanning tree is hit. Wilson's method on $G$, rooted at infinity, uses the stopping times $\tau_j$. Let

$$\begin{aligned} \mathcal{P}_j^n &:= \langle X_k(u_j) : k \le \tau_j^n \rangle \\ \mathcal{P} &:= \langle X_k(u_j) : k \le \tau_j \rangle \\ \gamma_j^n &:= \mathrm{LE}(\mathcal{P}_j^n) \\ \gamma_j &:= \mathrm{LE}(\mathcal{P}_j) \\ \mathcal{F}_j^n &:= \mathcal{F}_{j-1}^n \cup \gamma_j^n . \end{aligned}$$

We have

$$\mathsf{P}[e_1, \dots, e_M \in T(n)] = \mathsf{P}\left[e_1, \dots, e_M \in \cup_{j=1}^L \gamma_j^n\right]. \tag{19}$$

We claim that almost surely, $\tau_j^n \to \tau_j$, $\gamma_j^n \to \gamma_j$, $j = 1, \dots, L$. This can be proved by induction on $j$.

When $j = 1$, $\tau_1 = \infty$. Here $\tau_1^n =$ hitting time of $V \setminus V_n$, that goes to infinity as $n \to \infty$.

Assume now $j \ge 2$. Assume first the event $\{\tau_j = \infty\}$. Fix a large integer $N$. By the induction hypothesis, we have $\gamma_i^n \cap G_N = \gamma_i \cap G_N$ for all large enough $n$, $i < j$. (The statement is clear if $\gamma_i$ is a finite path. When $\gamma_i$ is infinite, the statement holds if $\mathcal{P}_i$ does not return to $G_N$ after time $\tau_i^n$.) Therefore, we have

$$\mathcal{F}_{j-1}^n \cap G_N = \cup_{i<j} \gamma_i^n \cap G_N = \cap_{i<j} \gamma_i \cap G_N = \mathcal{F}_{j-1} \cap G_N.$$

Since $\mathcal{P}_j$ does not hit $\mathcal{F}_{j-1}$, it follows that $\tau_j^n$ is greater than the exit time from $G_N$. Letting $N \to \infty$, we get $\tau_j^n \to \infty$.

Assume now the event $\{\tau_j < \infty\}$. For large $N$, we have $\mathcal{P}_j \subset G_N$. Again, from the induction hypothesis, $\mathcal{F}_{j-1}^n \cap G_N = \mathcal{F}_{j-1} \cap G_N$ for all large enough $n$. For such $n$, we have $\tau_j^n = \tau_j$, $\mathcal{P}_j^n = \mathcal{P}_j$ and $\gamma_j^n = \gamma_j$. The left hand side of (19) approaches $\mathsf{P}[e_1, \dots, e_M \in \mathcal{G}]$. Due to the convergence of the $\tau_j^n$'s and $\gamma_j^n$'s, the right hand side of (19) converges to

$$\mathsf{P}[e_1, \dots, e_M \in \cup_{j=1}^L \gamma_j] = \mathsf{P}[e_1, \dots, e_M \in \mathcal{F}].$$

This completes the proof. $\square$

## 5.7 Automorphisms

By a *graph automorphism* of $G$ we mean a pair of bijections $\phi : V \to V$, $\phi : E \to E$ (for convenience, we denote them by the same letter), such that $e$ is an edge between $v_1$ and $v_2$, if and only if $\phi(e)$ is an edge between $\phi(v_1)$ and $\phi(v_2)$. By a *network automorphism* of $(G, C)$, we mean a graph automorphism of $G$ that in addition satisfies $C(\phi(e)) = C(e)$ for all $e \in E$.

A graph automorphism induces a map $\phi : \Omega \to \Omega$ by letting $e \in \phi(\mathcal{F})$ if and only if $\phi^{-1}(e) \in \mathcal{F}$.

**Proposition 5.1.** FSF *and* WSF *are invariant under any network automorphisms.*

*Proof.* If $\langle G_n \rangle$ is any exhaustion, then $\langle \phi^{-1}(G_n) \rangle$ is also an exhaustion. For any finite $B \subset E$, we have

$$\begin{aligned}\mathrm{WSF}[B \subset \phi(\mathcal{F})] &= \mathrm{WSF}[\phi^{-1}(B) \subset \mathcal{F}] = \lim_n \mu^W_{\phi^{-1}(G_n)}[\phi^{-1}(B) \subset T] \\ &= \lim_n \mu^W_{G_n}[B \subset T] = \mathrm{WSF}[B \subset \mathcal{F}].\end{aligned}$$

□

## 5.8 Trees are infinite

**Proposition 5.2.** *On any infinite network $(G, C)$, all components are infinite* WSF*-a.s. and* FSF*-a.s.*

*Proof.* For any finite tree $t \subset G$, the event $\{t \text{ is a component}\}$ is a cylinder event, that has probability 0 with respect to $\mu^F_n$ for all large $n$ (since under $\mu^F_n$, $T$ is connected). Hence this event has probability 0 in the limit, and there are countably many such events. The proof for WSF is the same. □

## 5.9 Wilson's method on recurrent networks / equality

Let $(G, C)$ be an infinite recurrent network. We can make sense of Wilson's method on $G$, if we place the root at a fixed vertex $r \in V$. Indeed, if $v_1, v_2, \dots$ is any enumeration of $V \setminus \{r\}$, then let $\mathcal{F} := \{r\}$, let $\mathcal{P}_n$ be the path of the network random walk started at $v_n$, stopped when it hits $\mathcal{F}_{n-1}$, which is finite almost surely, by recurrence. Then let $\mathcal{F}_n := \mathcal{F}_{n-1} \cup \mathrm{LE}(\mathcal{P}_n)$, and $\mathcal{F} := \cup_n \mathcal{F}_n$. It is clear that $\mathcal{F}$ is a spanning tree of $G$.

**Proposition 5.3.** *[5] On any recurrent network, and for any enumeration, Wilson's method rooted at $r$ yields a tree with distribution* FSF $=$ WSF*.*

*Proof.* Let $B$ be a cylinder event, $\langle G_n \rangle$ an exhaustion. Let $K_0$ be the endpoints of edges on which $B$ depends, and $K := \{v_j : \exists i \geq j, v_i \in K_0\}$. Write

$$\partial_{\text{int}} G_n = \{x \in V_n : \exists y \in V \setminus V_n, x \sim y\}.$$

Run Wilson's method on $G_n^W$, rooted at $r$ to generate $T_{G_n^W}$ (note here that we use $r$, rather than $z_n$, as the root, which is the same, due to Theorem 3.1). Run Wilson's method, rooted at $r$, to generate $\mathcal{F}$. The random walks in the two constructions are indistinguishable until one of $\partial_{\text{int}} G_n$ is hit. Hence we can put the two constructions on the same probability space in such a way that they use the same random walks until $\partial_{\text{int}} G_n$ is first hit. Let $C_1 := \{\mathcal{F} \in B\}$, $C_2 := \{T_{G_n^W} \in B\}$. then

$$\begin{aligned}
\left|\mathsf{P}[\mathcal{F} \in B] - \mu_n^W(B)\right| &= |\mathsf{P}[C_1] - \mathsf{P}[C_2]| \\
&\leq \mathsf{P}[C_1 \Delta C_2] \\
&\leq \mathsf{P}[\text{some random walk started in } K \text{ hits } \partial_{\text{int}} G_n] \\
&\leq \sum_{v \in K} \mathsf{P}_v[\tau_{\partial_{\text{int}} G_n} < \tau_r] \xrightarrow{n \to \infty} 0,
\end{aligned}$$

by recurrence. This shows that $\mathcal{F}$ has the distribution of WSF. In exactly the same way, we obtain that $\mathcal{F}$ has the distribution of FSF. $\square$

It follows from the above theorem that on $\mathbb{Z}^2$, FSF = WSF.

## 5.10 Stochastic domination

In this section, we prove that WSF is always *stochastically smaller* than FSF; see (20) below for what this means precisely. This is a very powerful comparison of the two measures, that sometimes allows us to conclude that the two measures are in fact the same.

Let $\langle G_n \rangle$ be an exhaustion of $G$ by finite subnetworks. Since $G_n$ is a subgraph of $G_n^W$, we have, by Proposition 4.3(a),

$$\mu_n^F[e \in T] \geq \mu_n^W[e \in T]$$

for any edge $e \in E_n$. Letting $n \to \infty$, this implies that

$$\text{FSF}[e \in \mathcal{F}] \geq \text{WSF}[e \in \mathcal{F}], \quad e \in E.$$

We now state a more general inequality.

**Definition 5.1.** An event $A \subset \Omega$ is called *increasing* (or *upwardly closed*) if $F_1 \in A$, $F_2 \supset F_1$ imply that $F_2 \in A$.

That is, an event is increasing, if whenever it occurs, adding more edges to the configuration preserves the event. A simple example is:

$$\{\text{at least one of } e_1, e_2, e_3 \text{ is present}\}.$$

**Proposition 5.4.** *[5, Section 5] For any increasing event $A$ depending on the edges in $E_n$, we have $\mu_n^F[A] \geq \mu_n^W[A]$.*

**Corollary 5.1.** *For any increasing cylinder event $A$,*

$$\mathrm{FSF}[A] \geq \mathrm{WSF}[A]. \tag{20}$$

It can be deduced from a theorem of Strassen [32] that (20) implies that there exists a measure $\nu$ on $\Omega \times \Omega$ with marginals WSF and FSF, such that the first coordinate is $\nu$-almost surely contained in the second coordinate. That is,

$$\begin{aligned} &\nu[(\mathcal{F}_1, \mathcal{F}_2) : \mathcal{F}_1 \subset \mathcal{F}_2] = 1 \\ &\nu[(\mathcal{F}_1, \mathcal{F}_2) : \mathcal{F}_1 \in B_1] = \mathrm{WSF}[B_1] \\ &\nu[(\mathcal{F}_1, \mathcal{F}_2) : \mathcal{F}_2 \in B_2] = \mathrm{FSF}[B_2]. \end{aligned} \tag{21}$$

Such a $\nu$ is called a *monotone coupling* of the measures WSF and FSF, and we say that WSF *stochastically dominates* FSF. For a proof of Strassen's theorem and the existence of the coupling $\nu$, see [26, Section 10.2].

It is easy to see that (20) is necessary for (21). Indeed, if $\nu$ with properties (21) exists, then for any increasing event $A$

$$\begin{aligned} \mathrm{WSF}[A] &= \nu[(\mathcal{F}_1, \mathcal{F}_2) : \mathcal{F}_1 \in A] = \nu[(\mathcal{F}_1, \mathcal{F}_2) : \mathcal{F}_1 \in A, \, \mathcal{F}_1 \subset \mathcal{F}_2] \\ &\leq \nu[(\mathcal{F}_1, \mathcal{F}_2) : \mathcal{F}_2 \in A] = \mathrm{FSF}[A]. \end{aligned}$$

Hence (20) and the existence of $\nu$ satisfying (21) are equivalent. The point of Strassen's theorem is to prove the non-trivial statement that (20) implies (21).

## 5.11 Sufficient conditions for $\mathrm{FSF} = \mathrm{WSF}$

We can use the monotone coupling to give sufficient conditions for the equality $\mathrm{FSF} = \mathrm{WSF}$. These will be used later to show that the measures agree on $\mathbb{Z}^d$ as well as on some other graphs.

We saw that $\mathrm{WSF}[e \in \mathcal{F}] \leq \mathrm{FSF}[e \in \mathcal{F}]$ for all edges $e \in E$.

**Proposition 5.5.** *If $\mathrm{WSF}[e \in \mathcal{F}] = \mathrm{FSF}[e \in \mathrm{FSF}]$ for all edges $e \in E$, then $\mathrm{FSF} = \mathrm{WSF}$.*

*Proof.* We have $\{e \in \mathcal{F}_1\} \subset \{e \in \mathcal{F}_2\}$ $\nu$-a.s. By assumption, we also have,

$$\nu[e \in \mathcal{F}_1] = \mathrm{WSF}[e \in \mathcal{F}] = FSF[e \in \mathcal{F}] = \nu[e \in \mathcal{F}_2].$$

Hence $\{e \in \mathcal{F}_1\} = \{e \in \mathcal{F}_2\}$ $\nu$-a.s., and $\mathcal{F}_1 = \mathcal{F}_2$ $\nu$-a.s. Note that without the monotone coupling, we could could gain little from the equality of the one-dimensional marginals. □

The following proposition is based very much on the same idea. It says that if the expected degree of vertex $v$ is the same in the free spanning forest as in the wired spanning forest, for each $v \in V$, then the free and wired spanning forests coincide.

**Proposition 5.6.** *[5] If* $\mathsf{E}[\deg_{\mathcal{F}}(v)]$ *is the same under the measures* FSF *and* WSF *for all* $v \in V$*, then* FSF = WSF.

*Proof.* Under the measure $\nu$, the set of edges incident on $v$ in $\mathcal{F}_1$ is a subset of the set of edges incident on $v$ in $\mathcal{F}_2$. Due to the assumption, the two sets of edges coincide $\nu$-a.s., and hence $\mathcal{F}_1 = \mathcal{F}_2$ $\nu$-a.s. □

# 6 The number of trees on $\mathbb{Z}^d$

In this section we look at the question: is the random spanning forest on $\mathbb{Z}^d$ connected (that is, a single tree) or not? We first state a theorem that we will prove in greater generality in the next section. The theorem is implicit in [30], and is proved explicitly in [15].

**Theorem 6.1.** *On $\mathbb{Z}^d$, we have* FSF = WSF *for all $d \geq 1$.*

In view of this theorem, we can speak simply of the "Uniform Spanning Forest" on $\mathbb{Z}^d$: there is no need to specify the boundary condition, and we call it uniform, since we use $C(e) \equiv 1$.

We will frequently use the notation: $f \asymp g$ to denote that the positive quantities $f$ and $g$ (depending on some argument), are of the same order, that, is, there exist constants $c_1, c_2 > 0$, such that $c_1 g \leq f \leq c_2 g$.

The main result of this section is the following theorem of Pemantle [30].

**Theorem 6.2.** *The uniform spanning forest on $\mathbb{Z}^d$ is a.s. a single tree, if $d \leq 4$, and it has infinitely many components a.s., if $d \geq 5$. When $d \geq 5$, $u \neq v \in \mathbb{Z}^d$, then*

$$\mathsf{P}[u,\ v \text{ in the same component of } \mathcal{F}] \asymp |u - v|^{4-d}. \tag{22}$$

In what follows, we will also use the notation $|||u||| := 1 + |u|$, so that we do not have to make explicit exceptions for a negative power of 0, when $u = 0$. Note that the right hand side of (22) is comparable to $|||u - v|||^{4-d}$.

*Proof for $d = 1, 2$.* The case $d = 1$ is easy to see. Since $G_n = [-n, n] \cap \mathbb{Z}^d$ is a tree, the FSF concentrates on the single point $\{\mathbb{Z}\}$.

For $d = 2$, we saw in Proposition 5.3 that the forest is connected. □

Henceforth we assume $d \geq 3$, so the random walk is transient, and due to Theorem 5.1 we can use Wilson's method rooted at infinity to generate the spanning forest. The proof we present is adapted from [27], [4] and [26], where substantially more general theorems are proved.

Fix $u \in \mathbb{Z}^d$, and let $\mathcal{P}^0 := \langle X_k : k \geq 0 \rangle$ and $\mathcal{P}^u := \langle Y_k : k \geq 0 \rangle$ be the paths of independent simple random walks started at $X_0 = 0$ and $Y_0 = u$, respectively. Using Wilson's method with an enumeration starting with $0, u$, we immediately obtain the following proposition.

**Proposition 6.1.** *Vertices 0 and $u$ belong to the same component a.s., if and only if* $\mathsf{P}[\mathrm{LE}(\mathcal{P}^0) \cap \mathcal{P}^u \neq \emptyset] = 1$.

We will first focus on the case $d \geq 5$.
*Preliminaries.* The *Green function* of the walk is defined by

$$\begin{aligned} G(x,y) &:= \sum_{n=0}^{\infty} p_n(x,y) = \mathsf{E}\Big[\sum_{n=0}^{\infty} I[X_n = y]\Big| X_0 = x\Big] \\ &= \mathsf{E}[\#\text{ visits of } X_n \text{ to } y | X_0 = x] = G(y,x). \end{aligned}$$

Here $p_n(x,y)$ is the probability for simple random walk started at $x$ to be at $y$ after $n$ steps. The last equality follows from symmetry of $p_n(x,y)$ (reversibility). Good estimates on $p_n(x,y)$ yield important information for $G(x,y)$. Such estimates are provided by the Local Central Limit Theorem. We state this below (although we will not need such a detailed estimate in what follows). We say that $x \in \mathbb{Z}^d$ is *even* if the sum of its coordinates are.

**Theorem** (Local CLT; [22, Section 1.2])**.** *If $x$ has the same parity as $n \geq 1$, then*

$$p_n(0,x) = 2\left(\frac{d}{2\pi n}\right)^{d/2} e^{-\frac{d|x|^2}{n}} + E(n,x),$$

*where*

$$E(n,x) = \begin{cases} O\left(n^{-\frac{d}{2}-1}\right) \\ O\left(\frac{n^{-d/2}}{|x|^2}\right). \end{cases}$$

Summing over $n$, one can deduce (see [22, Theorem 1.5.4]) that there exists a constant $a_d > 0$, such that

$$G(x,y) \sim \frac{a_d}{|x-y|^{d-2}}, \quad \text{as } |x-y| \to \infty.$$

In particular,

$$G(x,y) \asymp |||x-y|||^{2-d}. \tag{23}$$

We first consider how likely is it that $\mathcal{P}^0$ and $\mathcal{P}^u$ intersect. Consider

$$V := \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} I[X_n = Y_m] = \#\text{ intersections of } \mathcal{P}^0 \text{ and } \mathcal{P}^u.$$

We find

$$\begin{aligned} \mathsf{E}V &= \sum_z \sum_n \sum_m \mathsf{P}[X_n = z = Y_m] = \sum_z \left(\sum_n p_n(0,z)\right)\left(\sum_m p_m(u,z)\right) \\ &= \sum_z G(0,z)G(u,z) \asymp \sum_z |||z|||^{2-d}|||u-z|||^{2-d}. \end{aligned}$$

**Lemma 6.1.** *If $d \geq 5$, we have*

$$\sum_z |||z|||^{2-d}|||u-z|||^{2-d} \asymp |||u|||^{4-d}. \tag{24}$$

*Proof.* Suppose that $2^N \leq |||u||| < 2^{N+1}$. By symmetry, it is enough to consider the contribution from $|||z||| \leq |||u-z|||$. Then, separating the terms with $n \leq N+1$ and $n > N+1$, and using that there are order $(2^n)^d$ vertices with $2^n \leq |||z||| < 2^{n+1}$, we get

$$\begin{aligned} \text{LHS of (24)} &\asymp \sum_{n=0}^{\infty} \sum_{\substack{z: 2^n \leq |||z||| < 2^{n+1} \\ |||z||| \leq |||u-z|||}} |||z|||^{2-d}|||u-z|||^{2-d} \\ &\asymp \sum_{n=0}^{N+1} (2^n)^{2-d} \left(2^N\right)^{2-d} (2^n)^d + \sum_{n=N+2}^{\infty} (2^n)^{d-2} (2^n)^{d-2} (2^n)^d \\ &\asymp \left(2^N\right)^{4-d} \asymp |||u|||^{4-d}. \end{aligned}$$

□

This estimate is enough to yield the quantitative upper bound:

$$\begin{aligned} \mathsf{P}[0 \text{ and } u \text{ in the same component}] &= \mathsf{P}[\mathrm{LE}(\mathcal{P}^0) \cap \mathcal{P}^u \neq \emptyset] \\ \leq \mathsf{P}[\mathcal{P}^0 \cap \mathcal{P}^u \neq \emptyset] &= \mathsf{P}[V \geq 1] \leq \mathsf{E}[V] \leq c|||u|||^{4-d}. \end{aligned}$$

We can also conclude that there are infinitely many trees a.s. Fix $K \geq 1$, and pick vertices $u_1, \dots, u_K$ far away from each other, so that $|||u_i - u_j||| \geq \varepsilon^{-1}$, $i \neq j$. Then

$$\begin{aligned} \mathsf{P}[\exists \text{ at least } K \text{ components}] &\geq \mathsf{P}[u_1, \dots, u_K \text{ are in different components}] \\ &\geq 1 - \sum_{i \neq j} c|||u_i - u_j|||^{4-d} \\ &\geq 1 - C(K)\varepsilon^{d-4}. \end{aligned}$$

Letting $\varepsilon \downarrow 0$, we get that there exists at least $K$ components with probability 1. Since this holds for any $K \geq 1$, we get that there are infinitely many components a.s.

Proving the quantitative lower bound is a bit more involved. We compute the second moment of $V$.

$$\begin{aligned} \mathsf{E}[V^2] &= \sum_z \sum_{n,i} \sum_{m,j} \mathsf{P}[X_n = z = Y_m,\, X_i = w = Y_j] \\ &= \sum_{z,w} \sum_{n,i} \mathsf{P}[X_n = z,\, X_i = w] \sum_{m,j} \mathsf{P}[Y_m = z,\, Y_j = w]. \end{aligned} \tag{25}$$

By separating the cases $n \le i$ and $i < n$, we can write

$$\begin{aligned}\sum_{n,i} \mathsf{P}[X_n = z,\ X_i = w] &= \sum_{n=0}^{\infty}\sum_{i=n}^{\infty} \mathsf{P}[X_n = z]\mathsf{P}[X_i = w | X_n = z] \\ &\quad + \sum_{i=0}^{\infty}\sum_{n=i+1}^{\infty} \mathsf{P}[X_i = w]\mathsf{P}[X_n = z | X_i = w] \\ &\le G(0,z)G(z,w) + G(0,w)G(w,z).\end{aligned}$$

We similarly get

$$\sum_{m,j} \mathsf{P}[Y_m = z,\ Y_j = w] \le G(u,z)G(z,w) + G(u,w)G(w,z).$$

The product of the two sums gives four terms. Two and two of these are identical, when summed over $z, w$, by symmetry, so we get

$$\begin{aligned}\mathsf{E}[V^2] &\le 2\sum_{z,w} G(0,z)G(u,z)G(z,w)^2 + 2\sum_{z,w} G(0,z)G(z,w)^2 G(w,u) \\ &\asymp \sum_{z} |||z|||^{2-d}|||u-z|||^{2-d}\sum_{w} |||z-w|||^{4-2d} \\ &\quad + \sum_{z} |||z|||^{2-d}\sum_{w} |||z-w|||^{4-2d}|||w-u|||^{2-d}.\end{aligned} \tag{26}$$

Since $4-2d = -d+(4-d) < -d$, the sum over $w$ in the first term is bounded by a finite constant. The remaining sum over $z$, by Lemma 6.1, is bounded by $|||u|||^{4-d}$. In the second term, the sum over $w$ gives $|||z-u|||^{2-d}$; this can be seen by a decomposition similar to that used in the proof of Lemma 6.1, and we do not give the details. The remaing sum over $z$ can then be estimated using Lemma 6.1. Putting things together, we get $\mathsf{E}[V^2] \le C|||u|||^{4-d}$.

By the Cauchy-Schwarz inequality,

$$\mathsf{E}[V^2]\mathsf{P}[V \ge 1] \ge (\mathsf{E}[VI[V \ge 1]])^2 = (\mathsf{E}V)^2.$$

Hence

$$\mathsf{P}[\mathcal{P}^0 \cap \mathcal{P}^u \ne \emptyset] = \mathsf{P}[V \ge 1] \ge c|||u|||^{4-d}.$$

We need more, to estimate $\mathsf{P}[\mathrm{LE}(\mathcal{P}^0) \cap \mathcal{P}^u \ne \emptyset]$. Here is the basic idea, how we handle the loop-erasure. Suppose that the paths $\mathcal{P}^0$ and $\mathcal{P}^u$ intersect at $X_n = z = Y_m$. Loop-erase $\mathcal{P}^0$ up to the intersection at time $n$, and call the loop-erasure $\gamma$. If the earliest intersection of $\gamma$ (as measured along $\gamma$) with $\langle Y_{m+k} : k \ge 0\rangle$ comes no later than the earliest intersection with $\langle X_{n+k} : k \ge 0\rangle$, then the intersection of $\gamma$ with $\langle Y_{m+k} : k \ge 0\rangle$ stays after the rest of $X$ is loop-erased. Given the intersection $X_n = z = Y_m$, this event will happen with conditional probability at least $1/2$, which will yield the statement.

To write the argument precisely, we denote

$$\begin{aligned}\mathrm{LE}(\langle X_k : k \leq n\rangle) &=: \langle \gamma^n(i) : i \leq K(n)\rangle \\ \sigma(n) &:= \min\{0 \leq i \leq K(n) : \gamma^n(i) = X_k \text{ for some } k \geq n\} \\ \tau(m,n) &:= \min\{0 \leq i \leq K(n) : \gamma^n(i) = Y_k \text{ for some } k \geq m\}.\end{aligned}$$

Let $I_{m,n} := I[X_n = Y_m]I[\tau(m,n) \leq \sigma(n)]$. Given the event $\{X_n = z = Y_m\}$, the paths $\langle X_{n+k} : k \geq 0\rangle$ and $\langle Y_{m+k} : k \geq 0\rangle$ are exchangeable. Therefore,

$$\mathsf{P}[\tau(m,n) \leq \sigma(n) | X_n = z = Y_m] \geq 1/2. \tag{27}$$

Put

$$W := \sum_n \sum_m I_{m,n},$$

and note that $W \geq 1$ implies that $\mathrm{LE}(\mathcal{P}^0)$ and $\mathcal{P}^u$ intersect. Then

$$\begin{aligned}\mathsf{E}W &= \sum_n \sum_m \mathsf{E}I_{m,n} = \sum_z \sum_n \sum_m \mathsf{E}[I_{m,n} | X_n = z = Y_m]\mathsf{P}[X_n = z = Y_m] \\ &\geq \frac{1}{2}\mathsf{E}V.\end{aligned}$$

Since $W \leq V$, we also have $\mathsf{E}W^2 \leq \mathsf{E}V^2$. Hence

$$\mathsf{P}[\mathrm{LE}(\mathcal{P}^0) \cap \mathcal{P}^u \neq \emptyset] \geq \mathsf{P}[W \geq 1] \geq \frac{(\mathsf{E}W)^2}{\mathsf{E}[W^2]} \geq \frac{(\mathsf{E}V)^2}{4\mathsf{E}[V^2]} \geq c|||u|||^{4-d}.$$

This concludes the proof of Theorem 6.2 in the case $d \geq 5$.

The quantitative estimate shows that the trees in the uniform spanning forest are 4-dimensional, when $d \geq 5$: consider

$$\begin{aligned}\mathcal{V}_n &:= |B(n) \cap \text{component of } 0| \\ &= \sum_{z:|z|\leq n} I[z \text{ and } 0 \text{ are in the same component}].\end{aligned}$$

Then $\mathsf{E}[\mathcal{V}_n] \asymp \sum_{z:|z|\leq n} |||z|||^{4-d} \asymp n^4$. It is remarkable, that the dimension of the trees is stable, and does not depend on $d$ for large $d$. See [4] for further interesting geometric properties of the uniform spanning forest.

We now deal with the case $d = 3, 4$. It is easy to see, using (23) that $\mathsf{E}V = \sum_z G(0,z)G(u,z) = \infty$, when $d = 3, 4$. The difficulty is in handling the loop-erasure, and show that an intersection occurs with probability 1. It turns out to be helpful to prove more, and show that $\mathrm{LE}(\mathcal{P}^0)$ and $\mathcal{P}^u$ intersect infinitely often, with probability 1. To work with something finite, we define

$$V_N := \sum_{n=0}^{N} \sum_{m=0}^{N} I[X_n = Y_m]$$

$$G_N(x, y) = \sum_{n=0}^{N} p_n(x, y) = G_N(y, x).$$

**Lemma 6.2.** *Assume that $X_0 = 0 = Y_0$. Then*

$$\frac{(\mathsf{E}V_N)^2}{\mathsf{E}[V_N^2]} \geq \frac{1}{4}. \tag{28}$$

*Proof.* We have

$$\mathsf{E}V_N = \sum_z \sum_{n=0}^{N} \sum_{m=0}^{N} p_n(0, z) p_m(u, z) = \sum_z G_N(0, z)^2 =: b_N.$$

Similarly to the computation in (26), and using the Cauchy-Schwarz inequality for the second term arising, we get $\mathsf{E}V_N^2 \leq 4b_N^2$. □

We will need the fact that the lower bound (28) still holds asymptotically, when the walks start at arbitrary vertices $u, v \in \mathbb{Z}^d$. Write $\mathsf{P}_{u,v}$, $\mathsf{E}_{u,v}$ to denote that $X_0 = u$, $Y_0 = v$.

**Lemma 6.3.** *Assume now $X_0 = u$ and $Y_0 = v$. Then we have*

$$\liminf_{n \to \infty} \frac{(\mathsf{E}_{u,v} V_N)^2}{\mathsf{E}_{u,v}[V_N^2]} \geq \frac{1}{4}. \tag{29}$$

*Proof.* Similarly to the computations in (26), and using Cauchy-Schwarz, we get $\mathsf{E}_{u,v} V_N^2 \leq 4b_N^2$. For the first moment, we write:

$$\mathsf{E}_{u,v} V_N = \sum_z \sum_{n=0}^{N} \sum_{m=0}^{N} p_n(u, z) p_m(v, z) = \sum_{n,m=0}^{N} p_{n+m}(u, v).$$

Similarly,

$$\mathsf{E}_{0,0} V_N = \sum_{n,m=0}^{N} p_{n+m}(0, 0). \tag{30}$$

By the Local CLT, as $n + m \to \infty$ we have, $p_{n+m}(u, v) \sim p_{n+m}(0, 0)$. Since the sum in (30) diverges as $N \to \infty$, we conclude that $\mathsf{E}_{u,v} V_N \sim \mathsf{E}_{0,0} V_N$, as $N \to \infty$. The claim follows. □

The key to the proof is the following lemma, which says that infinitely many intersections will occur with a probability that is *uniform* over the starting points $u$ and $v$, and an arbitrary initial segment of the $X$-path.

**Lemma 6.4.** *Fix a path* $\langle x_j \rangle_{j=-\ell}^{-1}$*, and set* $X_j := x_j$ *for* $-\ell \leq j \leq -1$*. Let* $X_0 = u$*,* $Y_0 = v$*. Then*

$$\mathsf{P}\left[|\mathrm{LE}(\langle X_n : n \geq -\ell \rangle \cap \langle Y_m \rangle| = \infty\right] \geq \frac{1}{16}.$$

*Proof.* Denote

$$\begin{aligned}\langle \gamma^n(i) : 0 \leq i \leq K(n) \rangle &:= \mathrm{LE}(\langle X_k : -\ell \leq k \leq n \rangle) \\ \sigma(n) &:= \min\{0 \leq i \leq K(n) : \gamma^n(i) = X_k \text{ for some } k \geq n\} \\ \tau(m,n) &:= \min\{0 \leq i \leq K(n) : \gamma^n(i) = Y_k \text{ for some } k \geq m\}.\end{aligned}$$

Let $I_{m,n} = I[X_n = Y_m] I[\tau(m,n) \leq \sigma(n)]$, and

$$W_N := \sum_{n=0}^{N} \sum_{m=0}^{N} I_{m,n}.$$

As in (27), we see that $\mathsf{E}W_N \geq \frac{1}{2}\mathsf{E}_{u,v} V_N \to \infty$, as $N \to \infty$. Also, $\mathsf{E}W_N^2 \leq \mathsf{E}_{u,v} V_N^2$. Cauchy-Schwarz gives

$$\begin{aligned}\mathsf{E}[W_N^2]\mathsf{P}[W_N \geq \varepsilon \mathsf{E}W_N] &\geq \left(\mathsf{E}[W_N I[W_N \geq \varepsilon \mathsf{E}W_N]]\right)^2 \geq \left(\mathsf{E}W_N - \varepsilon \mathsf{E}W_N\right)^2 \\ &= (1-\varepsilon)^2 (\mathsf{E}W_N)^2.\end{aligned}$$

Hence, for large $N$,

$$\begin{aligned}\mathsf{P}[W_N \geq \varepsilon \mathsf{E}W_N] &\geq (1-\varepsilon)^2 \frac{(\mathsf{E}W_N)^2}{\mathsf{E}[W_N^2]} \geq \frac{(1-\varepsilon)^2}{4} \frac{(\mathsf{E}_{u,v}V_N)^2}{\mathsf{E}_{u,v}[V_N^2]} \\ &\geq \frac{(1-\varepsilon)^2}{4}\left(\frac{1}{4} - \varepsilon\right).\end{aligned}$$

Since $W_N$ is monotone, this implies that $W_N \to \infty$ with probability at least $\frac{1}{4}(1-\varepsilon)^2(\frac{1}{4} - \varepsilon)$. Letting $\varepsilon \downarrow 0$, we get that $\mathsf{P}[W_N \to \infty] \geq 1/16$.

Finally, notice that on the event $\{W_N \to \infty\}$, the intersection $\mathrm{LE}(\langle X_n : n \geq -\ell \rangle) \cap \langle Y_m : m \geq 0 \rangle$ is infinite. This is because every intersection counted in $W_N$ is counted at most a finite number of times, by transience. ∎

*Proof of Theorem 6.2 when* $d = 3, 4$. Let $E = \{|\mathrm{LE}(\mathcal{P}^0) \cap \mathcal{P}^u| = \infty\}$. By Lévy's 0–1 law [12, Section 4.5], we have

$$I[E] \overset{\text{a.s.}}{=} \lim_n \mathsf{P}_{0,u}[E | X_1, \ldots, X_n, Y_1, \ldots, Y_n]. \tag{31}$$

By the Markov property, the conditional law of the continuation $\langle Y_{n+k} : k \geq 0\rangle$ is a random walk started at $v = Y_n$, and the law of the continuation $\langle X_{n+k} : k \geq 0\rangle$ is a random walk started at $u = X_n$. Noting that for any $n \geq 0$, $E$ is the same event as $\{|\mathrm{LE}(\mathcal{P}^0) \cap \langle Y_{n+k} : k \geq 0\rangle| = \infty\}$, we are in the setting of Lemma 6.4, and can conclude that the right hand side of (31) equals

$$\lim_n \mathsf{P}_{X_n,Y_n}[E|X_1,\ldots,X_n] \geq \frac{1}{16}.$$

Therefore, $I[E] \geq 1/16$ a.s., and hence $\mathsf{P}[E] = 1$. $\square$

# 7 Average degrees and amenability

We start this section by proving Theorem 6.1, following [5]. It will turn out that the proof is applicable in much greater generality. The generalization is related to the notion of *amenability*, and we explore this concept in the rest of this section.

*Proof of Theorem 6.1.* Let $V_n = [-n, n]^d \cap \mathbb{Z}^d$, $G_n = (V_n, E_n)$ the induced subgraph. First consider a fixed (deterministic) spanning forest $\mathcal{F}$ of $\mathbb{Z}^d$, such that all components of $\mathcal{F}$ are infinite. For $K \subset V$, we define the *external edge boundary* of $K$ as the set of edges between $K$ and $V \setminus K$:

$$\partial_E K := \{e \in E : \underline{e} \in K, \overline{e} \in V \setminus K\}.$$

Let $k_n$ be the number of trees in $\mathcal{F} \cap E_n$. Then $k_n \leq |\partial_E V_n|$, since an infinite tree intersecting $E_n$ contains a boundary edge, and different tree components of $\mathcal{F} \cap E_n$ cannot use the same boundary edge. We claim that (uniformly in $\mathcal{F}$)

$$\lim_{n\to\infty} \frac{1}{|V_n|} \sum_{x \in V_n} \deg_{\mathcal{F}}(x) = 2. \tag{32}$$

To see this, we note that the sum of the degrees counts each edge in $\mathcal{F} \cap E_n$ twice, and it counts each edge in $\mathcal{F} \cap \partial_E V_n$ once. Therefore,

$$\begin{aligned}\sum_{x \in V_n} \deg_{\mathcal{F}}(x) &\leq 2|\mathcal{F} \cap E_n| + |\partial_E V_n| = 2(|V_n| - k_n) + |\partial_E V_n| \\ &\leq 2|V_n| + |\partial_E V_n|.\end{aligned}$$

At the equality sign, we used that a tree component of $\mathcal{F} \cap E_n$ with $n$ vertices has $n - 1$ edges, and that $\mathcal{F}$ is spanning. On the other hand, we have

$$\sum_{x \in V_n} \deg_{\mathcal{F}}(x) \geq 2|\mathcal{F} \cap E_n| = 2(|V_n| - k_n) \geq 2|V_n| - 2|\partial_E V_n|.$$

Hence

$$2 - 2\frac{|\partial_E V_n|}{|V_n|} \leq \sum_{x \in V_n} \deg_{\mathcal{F}}(x) \leq 2 + \frac{|\partial_E V_n|}{|V_n|}.$$

Since $|\partial_E V_n| = O(n^{d-1})$, we get (32).

Now take expectation in (32) with respect to either FSF or WSF. By the dominated convergence theorem, we have

$$\lim_{n\to\infty} \frac{1}{|V_n|} \sum_{x \in V_n} \mathsf{E}[\deg_{\mathcal{F}}(x)] = 2.$$

By translation invariance, Proposition 5.1, we see that $\mathsf{E}[\deg_{\mathcal{F}}(x)]$ does not depend on $x$, and in fact $\mathsf{E}[\deg_{\mathcal{F}}(x)] = 2$ for all $x \in \mathbb{Z}^d$. This shows that the expected degree of each vertex is the same under both FSF and WSF. By Proposition 5.6, we conclude that FSF = WSF. □

The above proof used properties of $\mathbb{Z}^d$ only in two places: we used translation invariance, and we used that the sets $V_n$ have small boundary relative to their size. We now generalize these properties.

**Definition 7.1.** A graph or network is called *transitive*, if given $x, y \in V$, there is an automorphism $\phi$ such that $\phi(x) = y$.

By Proposition 5.1, for a transitive network, $\mathsf{E}[\deg_{\mathcal{F}}(x)]$ does not depend on $x$ for either FSF or WSF.

A large class of examples of transitive graphs is provided by *Cayley graphs*. Let $\Gamma$ be a countable group. We say that a set $S \subset \Gamma$ generates $\Gamma$, if the smallest subgroup containing $S$ is $\Gamma$. In what follows we will assume that $\Gamma$ is *finitely generated*, that is, a finite generating set $S$ exists. We will also assume that $S$ is symmetric, that is, whenever $s \in S$, we also have $s^{-1} \in S$. Then any $x \in \Gamma$ can be written as $x = s_1 s_2 \dots s_n$, $s_j \in S$.

**Definition 7.2.** The (right-) Cayley graph of $(\Gamma, S)$ is the graph $G = (V, E)$ with $V = \Gamma$, and $E = \{[x, y] : x = ys \text{ for some } s \in S\}$.

Left multiplication by any $\gamma \in \Gamma$ is an automorphism: if $\phi_\gamma(x) = \gamma x$, we have $x = ys$ if and only if $\gamma x = \gamma ys$. It follows that any Cayley graph is transitive, since given $x, y \in V$, $\phi_{yx^{-1}}(x) = yx^{-1}x = y$.

*Examples of Cayley graphs:*

1. $\mathbb{Z}^d$ with $S = \{\pm e_i : i = 1, \dots, d\}$, where $e_i$ are the unit coordinate vectors.

2. The free group on two letters, with $S = \{a, b, a^{-1}, b^{-1}\}$. This group can be described as the set of all finite words formed of elements of $S$ with no occurrence of the strings $aa^{-1}, a^{-1}a, bb^{-1}, b^{-1}b$. Composition is by concatenation, and removal of any forbidden strings. Some thought reveals that the Cayley graph is a 4-regular tree.

3. The free product of three copies of $\mathbb{Z}_2$. This can be described as the set of all finite words in the letters $a, b, c$ such that no $aa, bb, cc$ occur. Multiplication is again by concatenation, and removal of forbidden strings. With $S = \{a, b, c\}$, the Cayley graph is a 3-regular tree.

4. The free product of $\mathbb{Z}_2$ and $\mathbb{Z}_3$. This can be described as the set of all words in the letters $a, b$, such that no $aa$ or $bbb$ occurs. With $S = \{a, b, bb\}$, the Cayley graph consists of triangles joined together with line segments in a tree-like fashion.

Now we come to the generalization of the property of small boundary to volume ratio.

**Definition 7.3.** The *edge-isoperimetric constant* of a graph $G = (V, E)$ is defined by

$$\iota_E(G) := \inf\left\{\frac{|\partial_E K|}{|K|} : K \subset V \text{ finite}\right\}.$$

We call the graph *amenable*, if $\iota_E(G) = 0$. We call the graph *non-amenable*, if $\iota_E(G) > 0$.

Hence amenability of a graph is equivalent to the existence of finite subsets $K_n$ such that $\lim_n |\partial_E K_n|/|K_n| = 0$. In a non-amenable graph, each finite subset has a "large" edge boundary, in the sense that the size of the boundary is at least a constant fraction of the volume. In an amenable graph, there exist sets with small boundary.

*Remark.* A more general notion of amenability is introduced in [26, Chapter 6].

**Exercise 7.1.** Show that a $d$-regular tree is non-amenable if $d \geq 3$.

With the above notions, the same proof we had for $\mathbb{Z}^d$, yields the following theorem.

**Theorem 7.1.** *On a transitive, amenable network,* FSF = WSF*.*

The notion of amenability was originally introduced in the context of groups, and in the rest of this section we will illustrate it from the point of view of Cayley graphs of groups. This will also explain the origin of the name.

We define the *external vertex boundary* of $K$ as

$$\partial_V K := \{x \in V \setminus K : \exists\, x \in K \text{ such that } x \sim y\}.$$

**Definition 7.4.** The *vertex-isoperimetric constant* of a graph $G = (V, E)$ is defined by

$$\iota_V(G) := \inf\left\{\frac{|\partial_V K|}{|K|} : K \subset V \text{ finite}\right\}.$$

**Proposition 7.1.** *For any transitive graph, $\iota_E(G) > 0$ if and only if $\iota_V(G) > 0$.*

*Proof.* Clear from the inequalities $|\partial_V K| \leq |\partial_E K| \leq D|\partial_V K|$, where $D$ is the degree of a vertex. □

In particular, for Cayley graphs, amenability is equivalent to

$$\exists\ K_n \subset V \text{ finite such that } \lim_n |\partial_V K_n|/|K_n| = 0. \tag{33}$$

The origin of the name "amenable" is explained by the concept of an *invariant mean* on a group. Let $\ell^\infty(\Gamma)$ be the space of all bounded real functions on $\Gamma$. A linear map $\mu : \ell^\infty(\Gamma) \to \mathbb{R}$ is called a *mean* if $\mu(\mathbf{1}) = 1$, and $\mu(f) \geq 0$ for $f \geq 0$. For $\gamma \in \Gamma$ and $f \in \ell^\infty(\Gamma)$, we define the function $R_\gamma f \in \ell^\infty(\Gamma)$ by $R_\gamma f(x) := f(x\gamma)$. The mean $\mu$ is called *invariant*, if $\mu(R_\gamma f) = \mu(f)$ for all $\gamma \in \Gamma$ and $f \in \ell^\infty(\Gamma)$. Hence an invariant mean is a way of averaging functions on $\Gamma$ in such a way that the average value is invariant under the transformations $R_\gamma$. As a play on words, we call $\Gamma$ *amenable*, if an invariant mean exists on $\Gamma$.

Let us see the connection of invariant means to the notion introduced for graphs. We remind that we restrict the discussion to finitely generated groups, although amenability of groups applies in greater generality. If the Cayley graph $G$ is amenable, as witnessed by the sequence $K_n$, then we can consider the means

$$\mu_n(f) := \frac{1}{|K_n|} \sum_{x \in K_n} f(x).$$

We now show that these means are "almost invariant". For $\gamma \in S$, we have

$$|\mu_n(f) - \mu_n(R_\gamma f)| = \frac{1}{|K_n|} \left| \sum_{x \in K_n} (f(x) - f(x\gamma)) \right|.$$

If $x_1 = x_2\gamma$, $x_1, x_2 \in K_n$, then the terms $f(x_1)$ and $f(x_2\gamma)$ cancel in the sum. The terms that do not cancel are: $x_1 \in K_n$, such that $x_2 = x_1\gamma^{-1} \notin K_n$, and $x_2 \in K_n$ such that $x_1 = x_2\gamma \notin K_n$. Hence the number of terms that do not cancel is bounded by $2|\partial_V K_n|$, and therefore,

$$|\mu_n(f) - \mu_n(R_\gamma f)| \leq 2\|f\|_\infty \frac{|\partial_V K_n|}{|K_n|} \xrightarrow{n \to \infty} 0. \tag{34}$$

Since any $\gamma \in \Gamma$ can be written as a product of elements of $S$, we conclude that (34) holds for all $\gamma \in \Gamma$.

In order to use (34) to extract an invariant mean, we need a few facts from functional analysis. The space $X = \ell^\infty(\Gamma)$ is a Banach space with the supremum norm, and means belong to the unit ball of the dual space $X^*$. The *weak** topology on $X^*$ is given by specifying the following base: with $\mu \in X^*$, $f_1, \dots, f_k \in X$, and $\varepsilon > 0$, the sets

$$U(\mu, f_1, \dots, f_k, \varepsilon) := \{\nu \in X^* : |\nu(f_j) - \mu(f_j)| < \varepsilon,\, j = 1, \dots, k\}$$

form a base. Since in our case $X$ is not separable, care should be taken as this topology is not metrizable. Let $B$ denote the unit ball in $X^*$. By a theorem of Alaoglu [25, Theorem 12.3], $B$ is compact in the weak* topology. The sets $A_N := \overline{\{\mu_n : n \geq N\}}$ have the finite intersection property, and hence the intersection $\cap_N A_N$ is non-empty. Let $\mu \in \cap_N A_N$. Then $\mu$ is a weak* cluster point of the sequence $\{\mu_n\}$, that is, given any neighbourhood $U$ of $\mu$, and any $N$, there exists $n \geq N$ such that $\mu_n \in U$. Now apply this to the neighbourhoods $U(\mu, f, R_\gamma f, \varepsilon)$ to conclude that that $\mu(f) - \mu(R_\gamma f) = 0$. We have shown that the condition (33) implies that an invariant mean $\mu$ exists on $\Gamma$.

A remarkable theorem of Følner [13] implies that the converse is also true:

**Theorem 7.2.** *If an invariant mean exists on* $\Gamma$*, then for any Cayley graph of* $\Gamma$*,* (33) *holds.*

We do not prove this here.

We conclude this section with a few results on the connection between algebraic properties of groups and their amenability. Write

$$B(n) := \{x \in \Gamma : x = s_1 \dots s_k,\, k \leq n,\, s_j \in S\};$$

this is the ball of radius $n$ centred at the identity in the Cayley graph.

**Proposition 7.2.** *If* $|B(n)|$ *does not grow exponentially, then* $\Gamma$ *is amenable.*

*Proof.* We have

$$|B(n+1)| = |B(n) \cup \partial_V B(n)| \geq (1 + \iota_V(G))|B(n)|.$$

Hence, if $\Gamma$ is non-amenable, then, $\iota_V(G) > 0$, and $|B(n)|$ grows exponentially. □

**Proposition 7.3.** *An Abelian group is amenable.*

*Proof.* We have $|B(n)| \leq (2n+1)^{|S|}$, hence by Proposition 7.2, the group is amenable. □

**Proposition 7.4.** *If* $\Gamma$ *is amenable,* $H$ *is a subgroup of* $\Gamma$*, then* $H$ *is also amenable.*

*Proof.* Let $\hat{\mu}$ be an invariant mean on $\Gamma$. We construct an invariant mean on $H$. Let $f \in \ell^\infty(H)$. The idea is to "lift" $f$ to a function on $\Gamma$, and use the inavriant mean on $\Gamma$. In each left coset of $H$, we fix an element $x_0$, so that the coset takes the form $x_0 H$. We define $\hat{f}(x_0 h) = f(h)$, $h \in H$. We put $\mu(f) := \hat{\mu}(\hat{f})$. The requirments $\mu(\mathbf{1}) = 1$ and positivity are immediate. To prove invariance, let $g \in H$. Chasing the defintions, we have

$$\widehat{R_g f}(x_0 h) = (R_g f)(h) = f(hg) = \hat{f}(x_0 hg) = (R_g \hat{f})(x_0 h),$$

hence $\widehat{R_g f} = R_g \hat{f}$. It follows that

$$\mu(R_g f) = \hat{\mu}(\widehat{R_g f}) = \hat{\mu}(R_g \hat{f}) = \hat{\mu}(\hat{f}) = \mu(f).$$

□

A consequence of Proposition 7.4 is that if $\Gamma$ contains a non-amenable subgroup, then it has to be non-amenable. For example, if $\Gamma$ contains a free group, then, by Exercise 7.1 and Theorem 7.2, $\Gamma$ is non-amenable.

**Proposition 7.5.** *If* $H$ *is a normal subgroup of* $\Gamma$*, and* $H$ *and* $\Gamma/H$ *are amenable, then* $\Gamma$ *is amenable.*

**Corollary 7.1.** *Solvable groups are amenable.*

The proof of Proposition 7.5 is based on somewhat similar ideas as the proof of Proposition 7.4, and we do not give it here. (We can average over cosets using the invariant mean on $H$, then average the averages using the invariant mean on $\Gamma/H$.)

# 8 Currents on infinite networks

In this section we extend the definition of current to infinite networks. There will be two natural ways to do this, that will correspond to the free and wired boundary conditions.

Let $(G, C)$ be an infinite network. The space $\ell^2_-(E)$ consists of antisymmetric functions $\theta$ on the edges such that $\mathcal{E}(\theta) = \|\theta\|_R^2 = \frac{1}{2}\sum_{e\in E} R(e)\theta(e)^2 < \infty$. This is a real Hilbert space with inner product defined by (6). The Hilbert space $\ell^2(V)$ has inner product (5). We define the operators $\nabla$ and div by the same formulas as in (7) and (8).

**Exercise 8.1.** Show that $\|\nabla F\|_R \le \sqrt{2}\|F\|_C$, and $\|\mathrm{div}\theta\|_C \le \sqrt{2}\|\theta\|_R$, and that $-\nabla$ and div are adjoints of each other.

Just as in the finite case, we see that each function $\nabla \mathbf{1}_v$ is orthogonal to each $\sum_{e\in C}\chi^e$, $C$ a cycle in $G$. It follows that if we put

$$\mathrm{Cyc}^0 := \overline{\mathrm{span}\langle \sum_{e\in C}\chi^e : C \text{ a cycle in } G\rangle},$$
$$\mathrm{Gr}^0 := \overline{\mathrm{span}\langle \nabla \mathbf{1}_v : v \in V\rangle},$$

then $\mathrm{Gr}^0 \perp \mathrm{Cyc}^0$. We do not know if $\mathrm{Gr}^0$ and $\mathrm{Cyc}^0$ together span $\ell^2_-(E)$ or not, and as we will see, this is equivalent to the question whether FSF = WSF.

We define the *free current* in $G$ as $I_F^e := P^{\perp}_{\mathrm{Cyc}^0}\chi^e$. We define the *wired current* in $G$ as $I_W^e := P_{\mathrm{Gr}^0}\chi^e$.

## 8.1 Free boundary condition

Let $\langle G_n\rangle$ be an exhaustion of $G$ by finite subnetworks, $G_n = (V_n, E_n)$. For $G_n$, Lemma 4.1 gives the decomposition:

$$\ell^2_-(E_n) = \mathrm{Gr}_n \oplus \mathrm{Cyc}_n \subset \ell^2_-(E),$$

where $\subset$ denotes the natural inclusion of $\ell^2_-(E_n)$ in $\ell^2_-(E)$. We earlier defined

$$I^e_{G_n} = P_{\mathrm{Gr}_n}\chi^e = P^{\perp}_{\mathrm{Cyc}_n}.$$

Here $P^{\perp}_{\mathrm{Cyc}_n}$ is a projection taking place in the space $\ell^2_-(E_n)$. By the natural inclusion, we may view it as taking place in $\ell^2_-(E)$.

**Proposition 8.1.** *We have $\|I^e_{G_n} - I^e_F\|_R \to 0$ as $n\to\infty$, and*

$$\mathcal{E}(I^e_F) = I^e_F(e)R(e). \tag{35}$$

*Proof.* As a cycle in $G_n$ is also a cycle in $G_{n+1}$, we have

$$\cdots \subset \mathrm{Cyc}_n \subset \mathrm{Cyc}_{n+1} \subset \cdots \subset \mathrm{Cyc}^0.$$

Since any cycle in $G$ is contained in some $G_n$, we have $\mathrm{Cyc}^0 = \overline{\cup_n \mathrm{Cyc}_n}$. The first claim now follows from the following exercise.

**Exercise 8.2.** For any $\theta \in \ell^2_-(E)$, we have $\lim_n \|P^\perp_{\mathrm{Cyc}_n}\theta - P^\perp_{\mathrm{Cyc}^0}\theta\|_R = 0$.

Proposition 8.1 indeed follows, since

$$I^e_{G_n} = P^\perp_{\mathrm{Cyc}_n}\chi^e \to P^\perp_{\mathrm{Cyc}^0}\chi^e = I^e_F.$$

For the second claim, write

$$R(e)I^e_F(e) = (P^\perp_{\mathrm{Cyc}^0}\chi^e, \chi^e)_R = (P^\perp_{\mathrm{Cyc}^0}\chi^e, P^\perp_{\mathrm{Cyc}^0}\chi^e)_R = \mathcal{E}(I^e_F). \tag{36}$$

□

## 8.2 Wired boundary condition

We have the "wired graph" $G^W_n = (V_n \cup \{z_n\}, E^W_n)$. Again we have, by Lemma 4.1, the decomposition

$$\ell^2_-(E^W_n) = \mathrm{Gr}^W_n \oplus \mathrm{Cyc}^W_n \subset \ell^2_-(E).$$

Letting $\nabla^{(n)}$ denote taking the gradient in the graph $G^W_n$, we have

$$\mathrm{Gr}^W_n = \mathrm{span}\langle \nabla^{(n)}\mathbf{1}_v : v \in V_n \cup \{z_n\}\rangle.$$

**Lemma 8.1.** *We have* $\mathrm{Gr}^W_n \subset \mathrm{Gr}^0$.

*Proof.* For $v \in V$, we have $\nabla^{(n)}\mathbf{1}_v = \nabla\mathbf{1}_v \in \mathrm{Gr}^0$, as functions in $\ell^2_-(E)$. For $z_n$, we can write $\mathbf{1}_{z_n} = 1 - \sum_{v\in V_n}\mathbf{1}_v$, and hence

$$\nabla^{(n)}\mathbf{1}_{z_n} = -\sum_{v\in V_n}\nabla^{(n)}\mathbf{1}_v = \sum_{v\in V_n}\nabla\mathbf{1}_v \in \mathrm{Gr}^0.$$

□

**Proposition 8.2.** *We have* $\|I^e_{G^W_n} - I^e_W\|_R \to 0$ *as* $n \to \infty$, *and*

$$\mathcal{E}(I^e_W) = I^e_W(e)R(e). \tag{37}$$

*Moreover,* $I^e_W$ *has minimal energy among all* $\theta \in \ell^2_-(E)$ *satisfying* $\mathrm{div}\theta = \mathrm{div}\chi^e$.

*Proof.* We have, using Lemma 8.1,

$$\cdots \subset \mathrm{Gr}_n^W \subset \mathrm{Gr}_{n+1}^W \subset \cdots \subset \mathrm{Gr}^0.$$

For any $v \in V$, $\nabla \mathbf{1}_v$ is in some $\mathrm{Gr}_n^W$, so we have $\mathrm{Gr}^0 = \overline{\cup_n \mathrm{Gr}_n^W}$. By a similar argument as in Exercise 8.2, we get

$$I^e_{G_n^W} = P_{\mathrm{Gr}_n^W}\chi^e \to P_{\mathrm{Gr}^0}\chi^e = I^e_W.$$

The statement on $\mathcal{E}(I^e_W)$ follows as in (36). For the second part of the proposition, observe that

$$\begin{aligned} \mathrm{div}\theta = \mathrm{div}\chi^e \quad &\Leftrightarrow \quad \forall v \in V \; 0 = (\mathrm{div}(\theta - \chi^e), \mathbf{1}_v)_C = (\theta - \chi^e, -\nabla \mathbf{1}_v)_R \\ &\Leftrightarrow \quad \theta - \chi^e \perp \mathrm{Gr}^0. \end{aligned}$$

Hence the projection $P_{\mathrm{Gr}^0}\chi^e$ has minimal norm. □

*Remark.* From the definition of the free and wired currents, it is clear that $\mathcal{E}(I^e_W) \le \mathcal{E}(I^e_F)$, and that equality holds if and only if $I^e_W = I^e_F$.

We now describe when the free and wired currents are the same. We call $F : V \to \mathbb{R}$ *harmonic* at $v$, if $\mathrm{div}\nabla F(v) = 0$. From the definitions we have $\mathrm{div}\nabla F(v) = \left[\frac{1}{C_v}\sum_w C(v,w)F(w)\right] - F(v)$, so this is a discrete version of $\Delta F = 0$. We say that $F : V \to \mathbb{R}$ is a *Dirichlet function*, if $\mathcal{E}(\nabla F) < \infty$. Let $\mathrm{HD}(G)$ denote the space of harmonic Dirichlet functions.

**Proposition 8.3.** $\ell^2_-(E) = \mathrm{Gr}^0 \oplus \mathrm{Cyc}^0 \oplus \nabla\mathrm{HD}$.

*Proof.* Suppose that $\theta \in (\mathrm{Gr}^0)^\perp \cap (\mathrm{Cyc}^0)^\perp$. Due to orthogonality to cycles, we see, just as in the finite case, that $\theta = \nabla F$ for some $F : V \to \mathbb{R}$. In particular, $F$ is a Dirichlet function. Due to the first orthogonality assumption, we have for all $v \in V$:

$$(\mathrm{div}\nabla F, \mathbf{1}_v)_C = (\nabla F, -\nabla \mathbf{1}_v)_R = 0.$$

Hence $F$ is harmonic, and we have shown that $(\mathrm{Gr}^0)^\perp \cap (\mathrm{Cyc}^0)^\perp \subset \nabla\mathrm{HD}$.

For the reverse inclusion, assume that $\nabla F \in \nabla\mathrm{HD}$. Due to harmonicity, for all $v \in V$ we have

$$(\nabla F, \nabla \mathbf{1}_v)_R = (\mathrm{div}\nabla F, \mathbf{1}_v)_C = 0,$$

so $\nabla F \perp \mathrm{Gr}^0$. Also, being a gradient, $\nabla F$ is orthogonal to cycles, hence $\nabla F \perp \mathrm{Cyc}^0$. The proposition follows. □

Constant functions are in $\mathrm{HD}(G)$, but their gradients do not contribute to the subspace $\nabla\mathrm{HD}$. Sometimes constants are the only harmonic Dirichlet functions. The following theorem was proved by Benjamini, Lyons, Peres and Schramm [5, Theorem 7.3].

**Theorem 8.1.** *The following are equivalent.*
*(i)* $\mathrm{FSF} = \mathrm{WSF}$*;*
*(ii)* $I_W^e = I_F^e$ *for all* $e \in E$*;*
*(iii)* $\ell^2_-(E) = \mathrm{Gr}^0 \oplus \mathrm{Cyc}^0$*;*
*(iv)* $\mathrm{HD}(G) \cong \mathbb{R}$*.*

*Proof.* We first show the equivalence of (i) and (ii). Due to Proposition 5.5, equality of FSF and WSF is equivalent to $\mathrm{WSF}[e \in \mathcal{F}] = \mathrm{FSF}[e \in \mathcal{F}]$ for all $e \in E$. This is the same as $I_W^e(e) = I_F^e(e)$ for all $e \in E$, which by (35) and (37) is equivalent to $\mathcal{E}(I_W^e) = \mathcal{E}(I_F^e)$ for all $e \in E$. Due to Remark 8.2, this is equivalent to (ii).

To see the equivalence of (ii) and (iii), note that (iii) is equivalent to $P^{\perp}_{\mathrm{Cyc}^0} = P_{\mathrm{Gr}^0}$. Since $\{\chi^e\}$ forms a basis, this is equivalent to $P^{\perp}_{\mathrm{Cyc}^0}\chi^e = P_{\mathrm{Gr}^0}\chi^e$ for all $e \in E$, which is (ii).

The equivalence of (iii) and (iv) is immediate from Proposition 8.3. □

# 9 Scaling limits

An important question is what can we say about our models when the graph $\mathbb{Z}^d$ is replaced by a very fine grid $\delta\mathbb{Z}^d$, and $\delta \to 0$. Does the model converge in a suitable sense to a continuum model, often called *scaling limit*? We will only look at this question for the basic building block of spanning trees, the loop-erased random walk.

## 9.1 LERW in dimensions $d \geq 5$

We give the proof of the result of Lawler [22] that suitably scaled loop-erased walk in dimensions $d \geq 5$ converges to Brownian motion. The idea of the proof is to show that a positive fraction of points are not erased from the random walk path generating the loop-erased walk, and hence the result will follow from the convergence of scaled simple random walk to Brownian motion.

Let $X$ and $Y$ be independent simple random walks, $X_0 = 0 = Y_0$. Let $V$ denote the number of intersections of the two paths. We have shown in Section 6, that $\mathsf{E}V = \sum_n \sum_m \mathsf{P}[X_n = Y_m] < \infty$, if $d \geq 5$. We show that this implies that there is positive probability that the intersection of the two paths only consists of the intersection at $n = 0 = m$. Write $(n, m) \preceq (n_1, m_1)$, if $n \leq n_1$ and $m \leq m_1$, and write $\prec$, when equality does not hold. Define

$$h := \mathsf{P}[X(n) \neq Y(m),\ (0,0) \prec (n,m)].$$

**Lemma 9.1.** *If $d \geq 5$, then $h > 0$.*

*Proof.* Call $(n, m)$ a **-last intersection*, if $X(n) = Y(m)$, and $X(n_1) \neq Y(m_1)$ for $(n, m) \prec (n_1, m_1)$. Note that in general, *-last intersections are not unique. Since $\mathsf{P}[V < \infty] = 1$, there exists at least one *-last intersection a.s. This implies that

$$\begin{aligned} 1 &\leq \sum_n \sum_m \mathsf{P}[(n,m) \text{ is a *-last intersection}] \\ &= \sum_n \sum_m \mathsf{P}[X(n) = Y(m),\ X(n_1) \neq Y(m_1) \text{ for } (n,m) \prec (n_1, m_1)] \\ &= \sum_n \sum_m \mathsf{P}[X(n) = Y(m)]h = h\mathsf{E}V. \end{aligned}$$

This shows that $h \geq (\mathsf{E}V)^{-1} > 0$. ☐

We introduce the notation $\langle \hat{X}(i) : i \geq 0 \rangle := \mathrm{LE}(\langle X(n) : n \geq 0 \rangle)$ Let

$$\sigma(i) = \text{time of last visit to } \hat{X}(i) = \sup\{n : X(n) = \hat{X}(i)\}.$$

Let

$$\rho(j) = i \quad \text{for } \sigma_i \le j < \sigma_{i+1}.$$

Then it is clear that $\rho(\sigma(i)) = i$, and $\hat{X}(i) = X(\sigma(i))$. We will use the shorthand $X[k,l] = \langle X(n) : k \le n \le l \rangle$, and similalry $X(k,l]$, etc. Let

$$Z_n := \begin{cases} 1 & \text{if } \sigma(i) = n \text{ for some } i \ge 0; \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\begin{aligned} \rho(n) &= \sum_{j=1}^{n} Z_j \\ &= \#\text{ points remaining of } X[0,n] \text{ after loop-erasure.} \end{aligned}$$

It will be useful to extend $X$ to a two-sided walk:

$$X_n := \begin{cases} X_n & 0 \le n < \infty; \\ Y_{-n} & -\infty < n \le 0. \end{cases}$$

Note that the increments $X(n+1) - X(n)$, $-\infty < n < \infty$ are i.i.d. We call $j$ *loop-free* if $X(-\infty, j] \cap X(j, \infty) = \emptyset$. Then

$$b := \mathsf{P}[j \text{ loop-free}] = \mathsf{P}[X(-\infty, j] \cap X(j, \infty] = \emptyset] \ge h > 0.$$

**Lemma 9.2.** *If $d \ge 5$, with probability 1, there are infinitely many positive and negative loop-free points.*

*Proof.* Let $U := \#$ positive loop-free points. We call $j$ *$n$-loop-free*, if $X[j-n, j] \cap X(j, j+n] = \emptyset$. Then

$$\mathsf{P}[j \text{ } n\text{-loop-free}] =: b_n \to b \quad \text{as } n \to \infty.$$

Let

$$\begin{aligned} V_{i,n} &:= \{(2i-1)n \text{ is loop-free}\} \\ W_{i,n} &:= \{(2i-1)n \text{ is } n\text{-loop-free}\}. \end{aligned}$$

Then $W_{i,n}$, $i = 1, 2, \dots$ are independent. We have

$$\begin{aligned} \mathsf{P}[U \ge k] &\ge \mathsf{P}\left[\sum_{i=1}^{m} I[V_{i,n}] \ge k\right] \\ &\ge \mathsf{P}\left[\sum_{i=1}^{m} I[W_{i,n}] \ge k\right] - \mathsf{P}\left[\sum_{i=1}^{m} I[W_{i,n} \setminus V_{i,n}] \ge 1\right] \end{aligned}$$

Given $\varepsilon > 0$, the first term can be made at least $1 - \varepsilon$, by choosing $m$ large (uniformly in $n$). The absolute value of the second term is at most $m(b_n - b)$, which can be made less than $\varepsilon$ by choosing $n$ large. Hence $\mathsf{P}[U \geq k] \geq 1 - 2\varepsilon$. Letting $\varepsilon \to 0$ and $k \to \infty$, we get $\mathsf{P}[U = \infty] = 1$. □

**Theorem 9.1.** *If $d \geq 5$, there exists $a = a(d) > 0$, such that*

$$\lim_{n\to\infty} \frac{\rho(n)}{n} = a \quad a.s.$$

*Proof.* Let $j_0 := \inf\{j \geq 0 : j \text{ loop-free}\}$. Let the sequence of loop-free points be

$$\cdots < j_{-2} < j_{-1} < j_0 < j_1 < j_2 < \dots$$

Erase loops on each piece $X[j_i, j_{i+1}]$ *separately.* Let

$$\begin{aligned} \tilde{Z}_n &:= I[n\text{-th point is not erased}] \\ &= I[\mathrm{LE}(X[j_i, n]) \cap X(n, j_{i+1}] = \emptyset], \end{aligned}$$

where $j_i \leq n \leq j_{i+1}$. The following observation is crucial: if $n \geq j_0$, then $Z_n = \tilde{Z}_n$. This is due to the following. Since $j_0$ is loop-free, $\mathrm{LE}(X[0, j_0])$ does not influence loop-erasure of the continuation $X(j_0, \infty]$. Similarly, since $j_1$ is loop-free, $\mathrm{LE}[j_0, j_1]$ does not influence loop-erasure of the continuation $X(j_1, \infty)$, etc. This implies the claim.

Shifting the path $X$ so that $X(n)$ becomes the origin, we see that the sequence $\langle \tilde{Z}_n : n \geq 0 \rangle$ is stationary. It is also ergodic, as a function of the i.i.d. process $X(n+1) - X(n)$, $-\infty < n < \infty$. Hence by the observation above, and the ergodic theorem [12, Section 6.2], we have

$$\lim_{n\to\infty} \frac{\rho(n)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} Z_j = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \tilde{Z}_j = \mathsf{E}[\tilde{Z}_0] =: a.$$

We have $a \geq \mathsf{P}[0 \text{ loop-free}] > 0$. □

Since $\sigma(n) \to \infty$ a.s., it follows from Theorem 9.1 that almost surely,

$$a = \lim_{n\to\infty} \frac{\rho(\sigma(n))}{\sigma(n)} = \lim_{n\to\infty} \frac{n}{\sigma(n)}. \tag{38}$$

This gives the time-rescaling necessary so that we can compare scaled loop-erased walk to scaled simple random walk. Let us write $\Rightarrow$ for weak convergence in the space $C[0,1]$ of continuous functions on $[0,1]$ with the supremum metric. Write $\{B(t)\}_{0\leq t\leq 1}$ for a standard $d$-dimensional Brownian motion.

**Theorem 9.2.** *Let $d \geq 5$. Put*

$$\hat{W}_n(t) := \frac{d\sqrt{a}\hat{X}(nt)}{\sqrt{n}},$$

*(with linear interpolation in place). Then $\hat{W}_n(t) \Rightarrow B(t)$.*

*Proof.* Put

$$W_n(t) := \frac{d\sqrt{a}X(\frac{nt}{a})}{\sqrt{n}}.$$

Then $W_n(t) \Rightarrow B(t)$. From (38), we get

$$\sup_{0\leq t\leq 1}\left|\frac{a\sigma(nt)}{n} - t\right| \to 0 \text{ in probability, as } n \to \infty.$$

We can rewrite $\hat{W}_n(t) = d\sqrt{a}X(\sigma(nt))/\sqrt{n}$. Given $\varepsilon > 0$, choose $K \subset C[0,1]$ compact so that $\mathsf{P}[\{W_n(t)\} \notin K] \leq \varepsilon$, $n = 1, 2, \ldots$ This can be done due to tightness of the sequence $\{W_n(t)\}$. By Arzelà's Theorem [20], compactness of $K$ implies that we can find $\delta > 0$, such that for all $f \in K$, $|f(t) - f(s)| \leq \varepsilon$, if $|t - s| \leq \delta$. Then

$$\begin{aligned}
&\mathsf{P}[\sup_{0\leq t\leq 1}|\hat{W}_n(t) - W_n(t)| \geq \varepsilon] \\
&\quad\leq \varepsilon + \mathsf{P}\left[\sup_{0\leq t\leq 1}\left|\frac{d\sqrt{a}X(\sigma(nt))}{\sqrt{n}} - \frac{d\sqrt{a}X(\frac{nt}{a})}{\sqrt{n}}\right| \geq \varepsilon,\ \{W_n(t)\} \in K\right] \\
&\quad\leq \varepsilon + \mathsf{P}\left[\sup_{0\leq t\leq 1}\left|\frac{a\sigma(nt)}{n} - t\right| \geq \delta\right] \\
&\quad\leq 2\varepsilon
\end{aligned}$$

for $n$ large. Hence $\{\hat{W}_n(t) - W_n(t)\} \to 0$ in probability in the space $C[0,1]$, and therefore they have the same weak limit. □

In $d = 4$, Lawler [22] proves the following. Let

$$a_n = \mathsf{P}[n\text{-th point is not erased}] = \mathsf{P}[\mathrm{LE}(X[0,n]) \cap X(n,\infty) = \emptyset].$$

Lawler shows that this sequence grows only logarithmically, and

$$\rho(n)(na_n)^{-1} \to 1 \text{ in probability.}$$

Then he proves that $\hat{W}_n(t) = d\sqrt{a_n}\hat{X}(nt)/\sqrt{n} \Rightarrow B(t)$.

Regarding $d = 3$, Kozma has shown that a scaling limit exists [21] and is invariant under dilations and rotations.

## 9.2 LERW in dimension $d = 2$

In two dimensions, LERW has a deep connection to complex analytic functions. We informally describe this connection in this section. We identify $\mathbb{R}^2 \cong \mathbb{C}$. Let $U, V \subset \mathbb{C}$ be domains. Recall that $f : U \to V$ is called *conformal*, if it is analytic and one-to-one. The Riemann mapping Theorem states that given any two simply connected domains $U, V \subset \mathbb{C}$, $U, V \neq \mathbb{C}$, there exists a conformal mapping between them.

Locally around a point $z$, an analytic function $f$ has the form $f(w) = f(z) + f'(z)(w - z) + o(|w - z|)$. Hence $f$ is approximately the composition of a translation, a rotation and a dilation. The traslation and rotation of a Brownian path is again a Brownian path, and the dilation of a Brownian path is again Brownian path (with time rescaled). It turns out that for an analytic function $f$ and Brownian motion $B(t)$, $f(B(t))$ is a time-changed Brownian motion. The proof of this relies on stochastic integrals and Itô's formula. Recall that Itô's formula in one dimension says that for a sufficiently smooth function $f : \mathbb{R} \to \mathbb{R}$

$$f(B(t)) - f(B(0)) = \int_0^t f'(B(t))dB(t) + \frac{1}{2}\int_0^t f''(B(s))ds,$$

where the first term is a stochastic integral. For a function $f : \mathbb{R}^2 \to \mathbb{R}$, the formula takes the form

$$f(B(t)) - f(B(0)) = \int_0^t \nabla f(B(s)) \cdot dB(t) + \frac{1}{2}\int_0^t \Delta f(B(s))ds. \tag{39}$$

Suppose now that $f : \mathbb{C} \to \mathbb{C}$ is analytic, and write $f = u + iv$. The Cauchy-Riemann equations $\partial_1 u = \partial_2 v$, $\partial_2 u = -\partial_1 v$ imply that $\Delta u = 0$ and $\Delta v = 0$, hence the terms with $\Delta u$ and $\Delta v$ vanish in (39). Write $B(t) = B^1(t) + iB^2(t)$ for a complex Brownian motion. Then formally applying Itô's formula we get:

$$f(B(t)) - f(B(0)) = \int_0^t f'(B(s)) \cdot dB(t), \tag{40}$$

where $\cdot$ represents complex multiplication, and $dB(s) = dB^1(s) + idB^2(s)$. The right hand side is continuous a.s., has independent increments, that are "infinitesimally" centred Gaussians. By properties of stochastic integrals:

$$\mathsf{E}\left|\int_0^t |f'(B(s))dB(s)\right|^2 = \int_0^t |f'(B(s))|^2 ds =: \zeta(t).$$

This suggests, that $f(B(t))$ is a Brownian motion looked at at time $\zeta(t)$.

For a proof of the following theorem, see [29].

**Theorem 9.3.** *Let $f : U \to \mathbb{C}$ be analytic, $z \in U$, and let $B(t)$ be a planar Brownian motion started at $z$. Let $\tau_U := \inf\{t \geq 0 : B(t) \notin U\}$. Then $f(B(t)) = \tilde{B}(\zeta(t))$, $0 \leq t \leq \tau_U$, for some planar Brownian motion $\tilde{B}$ started at $f(z)$.*

Heuristically, the above theorem suggests that the scaling limit of LERW in $d = 2$ should also be conformally invariant, if it exists, since a one-to-one transformation "does not affect the loop-structure". In fact, it is natural to ask, if one can erase loops from Brownian motion to define the scaling limit. This does not quite work: there are loops on all scales, so one cannot define chronological loop-erasure.

Let us see a bit more precisely what conformal invariance of LERW should mean. Suppose that $0 \in U \subset \mathbb{C}$ is a domain, and consider LERW on the lattice $U \cap \delta\mathbb{Z}^2$, from 0 to $\partial(U \cap \delta\mathbb{Z}^2)$. Saying that a scaling limit exists, means that the random path of the LERW converges in suitable sense to some random path in $U$ from 0 to $\partial U$. If now $f : U \to V$ is a conformal map, where for simlicity $f(0) = 0$, then conformal invariance means that the limit we get from LERW on $V \cap \delta\mathbb{Z}^2$ from 0 to $\partial(V \cap \delta\mathbb{Z}^2)$ has the same distribution as the image of the random curve in $U$ under $f$.

The scaling limit process was discovered by Schramm [24]. It is most conveniently described in the setting $U = \{z \in \mathbb{C} : |z| < 1\}$, as a curve growing from $\partial U$ towards 0 (rather than from 0 to $\partial U$). Hence we consider the time reversal of the LERW. By a result of Lawler [23], the time reversal of LERW has the same distribution as the *reverse loop-erasure* of the random walk path. That is, for a random walk path $X[0, \tau_{U^c}]$ from 0 to $\partial U$, $\mathrm{LE}(X[0, \tau_{U^c}])$ has the same distribution as the loop-erasure of $X(\tau_{U^c}), X(\tau_{U^c} - 1), \ldots, X(1), X(0)$. (Here $\tau_{U^c}$ is the exit time from $U$.) The following simple Markovian property is fundamental.

**Lemma 9.3.** *(Markovian property) Let $\gamma = \gamma[0, \ell]$ denote the reverse loop-erasure of $X[0, \tau_{U^c}]$. Conditioned on $\gamma[0, k] = [\omega_0, \ldots, \omega_k]$, $\gamma[k, \ell]$ has the distribution of LERW from $\gamma(k)$ to 0 in the domain $U \setminus \{\omega_0, \ldots, \omega_k\}$.*

The scaling limit curve has the following description, see [33].

Let $\gamma : [0, \infty) \to \bar{U}$ be a continuous, simple curve, with $\gamma(0, \infty) \subset U$. Assume that $\gamma(0) = 1$. Let $g_t : U \setminus \gamma[0, t] \to U$ be the unique conformal map that satisfies the normalization $g_t(0) = 0$ and $g_t'(0) > 0$. One can show that $g_t'(0)$ is strictly increasing and continuous. Hence one can choose the parametrization of $\gamma$ in such a way that $g_t'(0) = e^t$. Let

$$W(t) = \lim_{\substack{z \to \gamma(t) \\ z \in U \setminus \gamma[0,t]}} g_t(z),$$

which lies on the unit circle. Then $W : [0, \infty) \to \partial U$ is continuous, and satisfies the Loewner differential equation:

$$\frac{d}{dt} g_t(z) = -g_t(z) \frac{g_t(z) + W(t)}{g_t(z) - W(t)}$$
$$g_0(z) = z.$$

The path $\gamma$ can be uniquely recovered from $W$, using the Loewner equation.

Crucially, the Markovian property Lemma 9.3, implies that if the scaling limit exists and is conformally invariant, then $W(t)$ has stationary and independent increments. The scaling limit of LERW is obtained when $W(t) = \exp(i\sqrt{\kappa}B(t))$, with $\kappa = 2$, for a standard one-dimensional Brownian motion $B(t)$, and is called radial $\mathsf{SLE}_2$. Other values of $\kappa > 0$ give radial $\mathsf{SLE}_\kappa$, and they arise in the context of the random cluster measures of Exercise 1.1.

A precise theorem about the LERW is the following. Let $D$ be a domain in $\mathbb{Z}^2$ whose boundary consists of edges of $\mathbb{Z}^2$. Assume that $0 \in D$. The *inner radius* of $D$ is defined as $\inf\{|z| : z \in D^c\}$. Let $\gamma$ be the path of the reverse loop-erasure of a random walk in $D$ from 0 to $\partial D$. View $\gamma$ as a continuos path. For $t \geq 0$, let $f_t : D \setminus \gamma[0, t] \to U$ be the unique conformal mapping such that $f_t(0) = 0$ and $f_t'(0) > 0$. Assume that $\gamma$ has been parametrized so that $f_t'(0)/f_0'(0) = e^t$. Let $W(t) = f_t(\gamma(t))$, and write $W(t) = \exp(i\vartheta(t))$ with $\vartheta$ continuous.

**Theorem 9.4** (Lawler, Schramm, Werner; 2004 [24])**.** *For every $T > 0$ and $\varepsilon > 0$, there exists $r_1 = r_1(\varepsilon, T)$, such that for all domains $D$ as above with inner radius at least $r_1$, there is a coupling of $\gamma$ with a standard one-dimensional Brownian motion $\{B(t)\}$ starting at a uniform point in $[0, 2\pi)$, such that*

$$\mathsf{P}\left[\sup_{0 \leq t \leq T} |\vartheta(t) - B(2t)| > \varepsilon\right] < \varepsilon.$$

Thus the theorem says that the "driving function" of the LERW $\gamma$ converges to a $B(2t)$ (which has the same distribution as $\sqrt{2}B(t)$).

# 10 The Abelian sandpile / Chip-firing game

In this last section we look at the Abelian sandpile model, also called the chip-firing game. It has a surpising connection with spanning trees, that is not at all apparent from its definition. See [16] for some very interesting results that are not discussed here. See [9, 10] for the relevance of the model to the concept of *self-organized criticality*, and [31] for an overview.

Let $G = (V, E)$ be a finite, connected graph, with a distinguished vertex $\delta$, called the *sink*. Write $V_0 = V \setminus \{\delta\}$, and $n = |V|$. We will use the matrix indexed by $V_0 \times V_0$:

$$\Delta_{xy} := \begin{cases} \deg_G(x) & \text{if } x = y; \\ -1 & \text{if } x \sim y; \\ 0 & \text{otherwise,} \end{cases} \qquad x, y \in V_0.$$

By a *chip-configuration* we mean an assignment of finitely many chips to the non-sink vertices, where each vertex can hold any non-negative number of chips. Chip-configurations are collected in the set

$$X_G := \prod_{x \in V_0} \{0, 1, 2, \dots\}.$$

We say that $\eta \in X_G$ is *stable*, if $\eta_x < \deg_G(x)$, $x \in V_0$. If $\eta$ has an *unstable vertex*, that is $\eta_x \geq \deg_G(x)$, that vertex can *fire*, which means that one chip is sent along each edge from $x$, and any chips arriving at the sink are removed. This can be written as follows. When $x \in V_0$ fires, we replace $\eta$ by $T_x \eta$, where $(T_x \eta)_y = \eta_y - \Delta_{xy}$, $y \in V_0$. More concisely, $T_x \eta = \eta - \Delta_x$, where $\Delta_x$ is the row of $\Delta$ corresponding to $x$. Note that if the vertex fired is not a neighbour of the sink, then the number of chips is conserved, while if a neighbour of the sink is fired, then the number of chips is reduced by the number of edges to the sink. The following lemma shows that any chip-configuration can be stabilized, in a unique way.

**Lemma 10.1.** *(i) For any $\eta \in X_G$ there exists $k \geq 0$ and $x_1, \dots, x_k \in V_0$, such that $T_{x_k} \dots T_{x_1} \eta$ is defined and stable.*
*(ii) If $T_{y_\ell} \dots T_{y_1} \eta$ is another stabilization, then $\ell = k$ and $y_1, \dots, y_\ell$ is a permutation of $x_1, \dots, x_k$.*

*Proof.* (i) This property is ensured by the existence of the sink. Fix $x \in V_0$. Since $G$ is connected, there exists a path $x = y_0, \dots, y_r = \delta$. Since there are finitely many chips, $y_{r-1}$ can fire only finitely many times, because each time it fires, it sends a chip to the sink. It follows that $y_{r-2}$ can also fire only finitely many times, since each time it fires, it sends a chip to $y_{r-1}$. By induction, $x = y_0$ can only fire finitely many times.

(ii) Suppose there is a counterexample to the statement, and assume $k$ is minimal. Since $x_1$ is unstable in $\eta$, there exists a smallest $j$ such that $y_j = x_1$. Now note that if $z, w$ are both unstable in a configuration $\xi$, then $T_z T_w \xi = T_w T_z \xi$ (both are equal to $\xi - \Delta_z - \Delta_w$). It follows that $T_{y_j}$ can be commuted through $T_{y_{j-1}}, \dots, T_{y_1}$, so that

$$T_{y_\ell} \dots T_{y_1} \eta = T_{y_\ell} \dots T_{y_{j+1}} T_{y_{j-1}} \dots T_{y_1} T_{x_1} \eta.$$

This shows that $\eta' = T_{x_1} \eta$ provides a shorter counterexample, a contradiction. $\square$

Let us denote by

$$\Omega_G := \prod_{x \in V_0} \{0, 1, \dots, \deg_G(x) - 1\}$$

the set of stable configurations. Lemma 10.1 shows that there is a well-defined *stabilization map* $\mathcal{S} : X_G \to \Omega_G$, that is the result of carrying out all possible firings (in any order). Let $\delta_x$ denote the configuration with one chip at $x$ and no other chips. We define the *addition operators*

$$a_x : X_G \to \Omega_G, \qquad a_x \eta = \mathcal{S}(\eta + \delta_x), \qquad x \in V_0,$$

that is, we add one chip at $x$ and stabilize.

**Lemma 10.2** (Abelian porperty)**.** *$a_x a_y = a_y a_x$ for all $x, y \in V_0$.*

*Proof.* Consider a sequence of firings that stabilizes $\eta+\delta_x$. The same sequence can be legally applied to $\eta + \delta_x + \delta_y$, since the extra chip at $y$ only makes the configuration larger, and the same vertices can be fired. This sequence takes $\eta + \delta_x + \delta_y$ to $\mathcal{S}(\eta + \delta_x) + \delta_y$. Stabilizaing further we obtain, using Lemma 10.1(ii), that $\mathcal{S}(\eta + \delta_x + \delta_y) = \mathcal{S}(\mathcal{S}(\eta + \delta_x) + \delta_y) = a_y a_x \eta$. Exchanging the roles of $x$ and $y$, we have $\mathcal{S}(\eta + \delta_x + \delta_y) = a_x a_y \eta$. $\square$

**Corollary 10.1.** *Applying $a_{x_k} \dots a_{x_1}$ gives the same result as adding all the chips at the beginning, and stabilizing.*

Stabilization involves subtracting rows of the the matrix $\Delta$. Hence it is plausible that the process of stabilization is related to equivalence modulo the row span of $\Delta$, and this what we will show soon. It is not hard to see that if we keep adding chips and stabilize, not all stable configurations will be seen in the long run. For example, if we start with a configuration that has two zeroes next to each other, then after adding chips at one of those vertices, we never see the two zeroes again: if one of the vettices has just fired, the neighbour cannot be zero at the same time. This motivates the definition of recurrent configuration below.

**Definition 10.1.** The *sandpile group* of $G$ is defined as

$$K_G := \mathbb{Z}^{n-1}/\mathbb{Z}^{n-1}\Delta,$$

where $n-1 = |V_0|$, and $\mathbb{Z}^{n-1}\Delta$ is the integer row span of $\Delta$.

We adopt the definition of recurrent configuration and the proof of Proposition 10.1 below from [14]. See [16] for several other equivalent definitions.

**Definition 10.2.** A stable configuration $\eta$ is called *recurrent*, if there exist arbitrarily large $\xi \in X_G$ such that $\mathcal{S}(\xi) = \eta$. We write $\mathcal{R}_G$ for the set of recurrent configurations.

**Proposition 10.1.** *Every equivalence class mod $\Delta$ contains exactly one element of $\mathcal{R}_G$.*

*Proof.* (i) We first show that every equivalence class contains a recurrent configuration. Let $F$ be all elements of a fixed equivalence class that are larger componentwise than $\deg_G(x)$. This set is infinite, and since $\Omega_G$ is finite, some element of $\mathcal{S}(F)$ has infinitely many preimages, and that element is then necessarily recurrent.

(ii) We now show that if $\eta_1, \eta_2 \in \mathcal{R}_G$ and $\eta_1 \overset{\Delta}{\sim} \eta_2$, then $\eta_1 = \eta_2$. Take $\xi_1, \xi_2$ such that $\mathcal{S}(\xi_1) = \eta_1$ and $\mathcal{S}(\xi_2) = \eta_2$. We may assume that $(\xi_2)_x \geq \deg_G(x)$, $x \in V_0$. In particular, $(\xi_2)_x > (\eta_1)_x$, $x \in V_0$. Since $\xi_2 \overset{\Delta}{\sim} \eta_1$, we can write

$$\xi_2 - \eta_1 = \sum_{x \in V_0} n_x \Delta_x, \tag{41}$$

for some integers $\{n_x\}$. We claim that $n_x \geq 0$, $x \in V_0$. Put $W := \{x \in V_0 : n_x < 0\}$, and assume that $W$ is non-empty. The relation (41) says that if we allow the possibility of negative number of chips, and we "unfire" the vertices in $W$ and then fire the vertices in $V_0 \setminus W$, according to $\{n_x\}$, then this takes $\xi_2$ to $\eta_1$. So first "unfire" the each vertex $x \in W$, $-n_x$ times, and denote the result $\xi_2'$. The total number of chips in $W$ in $\xi_2'$ is at least what it was in $\xi_2$. Hence there exists $y \in W$ such that $(\xi_2')_y \geq (\xi_2)_y$. Now fire each $x \in V_0 \setminus W$ $n_x$ times to get $\eta_1$. This can only increase the number of chips at $y$, and we get:

$$(\eta_1)_y \geq (\xi_2')_y \geq (\xi_2)_y > (\eta_1)_y.$$

This is a contradiction, so $W = \emptyset$, and all coefficients in (41) are non-negative.

Consider now the configuration $\xi_3 = \xi_1 + \xi_2 - \eta_1$. We want to start with $\xi_3$ and fire each vertex $x \in V_0$, $n_x$ times, to take the configuration $\xi_3$ to $\xi_1$. This may be non-legal, as it may create negative number of chips, somewhere. But if $\xi_1$ is large enough this cannot happen, and we are allowed to assume this. Hence for $\xi_1$ large enough, we have $\mathcal{S}(\xi_3) = \mathcal{S}(\xi_1) = \eta_1$.

Now consider the sequence of firings that takes $\xi_1$ to $\eta_1$. Adding extra chips accoridng to $\xi_2 - \eta_1 > 0$ does not affect the legality of these firigns, and then this sequence of firings takes $\xi_3$ to $\eta_1 + \xi_2 - \eta_1 = \xi_2$. It follows that $\mathcal{S}(\xi_3) = \mathcal{S}(\xi_2) = \eta_2$. We have proved that $\eta_1 = \mathcal{S}(\xi_3) = \eta_2$. □

**Lemma 10.3.** *$a_x$ leaves $\mathcal{R}_G$ invariant for all $x \in V_0$.*

*Proof.* Take $\eta \in \mathcal{R}_G$, with $\eta = \mathcal{S}(\xi)$. Then $a_x\eta = \mathcal{S}(\eta + \delta_x) = \mathcal{S}(\xi + \delta_x)$. Since $\xi$ can be arbitrarily large, it follows that $a_x\eta$ is recurrent. □

**Corollary 10.2.** *For $\eta \in \mathcal{R}_G$, $a_x\eta$ is the unique element of $\mathcal{R}_G$ equivalent to $\eta + \delta_x$ mod $\Delta$.*

**Corollary 10.3.** *$a_x : \mathcal{R}_G \to \mathcal{R}_G$ is one-to-one.*

*Proof.* The map $\eta \mapsto \eta + \delta_x$ respects equivalence classes. □

**Proposition 10.2.** *With the operation*

$$\eta \oplus \zeta := \mathcal{S}(\eta + \zeta), \qquad \eta, \zeta \in \mathcal{R}_G,$$

*$\mathcal{R}_G$ is an Abelian group isomorphic to $K_G$.*

*Proof.* Commutativity is clear from the definition. Associativity is clear from

$$(\eta \oplus \zeta) \oplus \xi = \mathcal{S}(\eta + \zeta + \xi) = \eta \oplus (\zeta \oplus \xi).$$

Let $I \in \mathcal{R}_G$ be equivalent to 0 mod $\Delta$. Then $\eta + I \overset{\Delta}{\sim} \eta$, and hence by Corollary 10.2, $\mathcal{S}(\eta + I) = \eta$. Hence $I$ is an identity element. Given $\eta \in \mathcal{R}_G$, let $\bar{\eta} \in \mathcal{R}_G$ be equivalent to $-\eta$ mod $\Delta$. Then $\eta + \bar{\eta} \overset{\Delta}{\sim} 0$, hence $\mathcal{S}(\eta + \bar{\eta}) = I$. Hence inverses exist.

The map $\phi : \mathcal{R}_G \to \mathbb{Z}^{n-1}/\mathbb{Z}^{n-1}\Delta$ given by $\eta \mapsto [\eta]_\Delta$ is one-to-one by Proposition 10.1. It is clearly a homomorphism, and hence an isomorphism. □

*Remark.* It is an intriguing question what the configuration $I$ looks like. See [16] for some pictures.

We now define a natural Markov chain with statespace $\Omega_G$. At each timestep, pick $x \in V_0$ uniformly at random, add a chip at $x$ to the current configuration, and stabilize (equivalently, apply $a_x$).

**Proposition 10.3.** *(i) The set of states that are recurrent in the Markov chain sense is precisely $\mathcal{R}_G$.*
*(ii) The stationary distribution is uniform on $\mathcal{R}_G$.*

*Proof.* (i) Suppose $\eta \in \Omega_G$ is recurrent in the Markov chain sense. Given any $\xi$, there is positive probability for the Markov chain to go from $\mathcal{S}(\eta+\xi)$ to $\eta$, which implies that there exists $\alpha$ such that $\eta = \mathcal{S}(\eta + \xi + \alpha)$. Since $\eta + \xi + \alpha$ can be arbitrarily large, we get $\eta \in \mathcal{R}_G$. We have proved that all states in $\Omega_G \setminus \mathcal{R}_G$ are transient, hence the Markov chain, starting from any state will enter $\mathcal{R}_G$ with probability 1 and stay there. We prove that all of $\mathcal{R}_G$ is a single recurrent class, by showing that the restriction of the Markov chain to $\mathcal{R}_G$ is irreducible. Indeed, given any $\eta, \zeta \in \mathcal{R}_G$, there exists $\xi$ (for example $\zeta \ominus \eta \in \mathcal{R}_G$), such that $\mathcal{S}(\eta + \xi) = \zeta$. This implies irreducibility.

(ii) The transition matrix on $\mathcal{R}_G$ is doubly stochastic:

$$\sum_{\eta \in \mathcal{R}_G} p(\eta, \xi) = \sum_{\substack{\eta \in \mathcal{R}_G \\ \exists x \in V_0 : a_x \eta = \xi}} \frac{1}{|V_0|} = \sum_{x \in V_0} \frac{1}{|V_0|} = 1.$$

□

*Remark.* Statement (ii) above is an example of the general fact that a (invariant) random walk on a group always has the uniform measure as stationary distribution.

We have $|\mathcal{R}_G| = |K_G| = \det(\Delta)$. The matrix-tree theorem in combinatorics [6, Corollary II.13] states that $\det(\Delta)$ is also the number of spanning trees of $G$. Our goal in the remainder of this section is to give a mapping between $\mathcal{R}_G$ and $T_G =$ set of spanning trees of $G$.

Let

$$\beta_x := \deg_G(x) - \deg_{V_0}(x) = \#\text{ edges from } x \text{ to } \delta, \qquad x \in V_0.$$

The following proof of the so called buring algorithm is from [16].

**Lemma 10.4** (Burning algorithm)**.** *$\eta \in \mathcal{R}_G$ if and only if $\mathcal{S}(\eta + \beta) = \eta$. If $\eta \in \mathcal{R}_G$, each vertex $x \in V_0$ fires exactly once in stabilizing $\eta + \beta$.*

*Proof.* (i) Suppose $\mathcal{S}(\eta+\beta) = \eta$. Then $\mathcal{S}(\eta+M\beta) = \eta$ for all $M \geq 1$. Starting from $\eta + M\beta$, we can selectively fire vertices and obtain a configuration that can be made arbitrarily large by choosing $M$ large. Hence $\eta$ is recurrent.

(ii) Suppose now that $\eta \in \mathcal{R}_G$. Observe that $\beta = \sum_{x \in V_0} \Delta_x$. Hence we have $\eta + \beta \overset{\Delta}{\sim} \eta$, therefore, $\mathcal{S}(\eta + \beta) = \eta$.

(iii) Let $c_x = \#$ times $x$ fires in stabilizing $\eta + \beta$. Then

$$\eta = \mathcal{S}(\eta + \beta) = \eta + \beta - \sum_{x \in V_0} c_x \Delta_x.$$

Since the rows of $\Delta$ are linearly independent, $c_x = 1$, $x \in V_0$. □

We can view the statement of the lemma like this: start from a configuration $\eta$. First "fire the sink", meaning that $\beta_x$ chips are placed on each neighbour $x$ of the sink. Then fire as many vertices as possible. If all vertices fired, we get back $\eta$ and $\eta$ was recurrent.

Following the sequence of firings allows us to define a map from $\mathcal{R}_G$ to $T_G$. Fix $\eta \in \mathcal{R}_G$. When stabilizing $\eta + \beta$, at each time step $t = 1, 2, \dots$ fire all vertices, simultaneously, that can be fired. Let

$$\begin{aligned} &W_0 := \{\delta\} \\ &W_t := \text{vertices that fire at time } t, \qquad t = 1, 2, \dots \end{aligned}$$

We of course have eventually $W_t = \emptyset$, and $V = \cup_{t \geq 0} W_t$ a disjoint union. A spanning tree $\psi(\eta)$ will be defined by connecting a vertex $x \in W_t$ $(t \geq 1)$ with a vertex $\bar{x} \in W_{t-1}$, as follows. Fix $x \in W_t$, $t \geq 1$. Let

$$\begin{aligned} N_x &:= \{y \sim x : y \in \cup_{s<t} W_s\} \\ M_x &:= \{y \sim x : y \in W_{t-1}\}. \end{aligned}$$

Up to time $t-1$, $x$ has received $N_x$ chips, and it fires at time $t$. Therefore, $\eta_x + N_x \geq \deg_G(x)$. Before time $t-1$, $x$ has received $N_x - M_x$ chips, and it does not fire at time $t-1$, therefore, $\eta_x + N_x - M_x < \deg_G(x)$. Hence

$$\deg_G(x) - N_x \leq \eta_x < \deg_G(x) - N_x + M_x.$$

The number of possible values of $\eta_x$ equals $M_x$, which equals the number of neighbours of $x$ in $W_{t-1}$. Associate each possible value with a neighbour $\bar{x} \in W_{t-1}$, and draw an edge $x\bar{x}$. Then $\{x\bar{x} : x \in V_0\}$ is a spanning tree of $G$.

**Theorem 10.1** (Majumdar-Dhar correspondence [28]). *The map $\psi$ defined above is one-to-one.*

Note that under the correspondence, the stationary distribution of the Markov chain is mapped to the uniform spanning tree measure. This opens up the possibility of studying the sandpile model using the uniform spanning tree. See [3, 17, 18] for result in this direction.

# References

[1] Aldous, D.: The random walk construction of uniform spanning trees and uniform labelled trees. SIAM J. Disc. Math. **3**, (1990) 450–465.

[2] Aldous, D. and Fill, J.: *Reversible Markov chains and random walks on graphs.* Book in preparation.
`http://www.stat.berkeley.edu/~aldous/RWG/book.html`

[3] Athreya, S.R. and Járai, A.A.: Infinite volume limit for the stationary distribution of Abelian sandpile models. *Commun. Math. Phys.* **249**, (2004) 197–213.

[4] Benjamini, I., Kesten H., Lyons, R. and Schramm, O.: Geometry of the uniform spanning forest: transitions in dimensions $4, 8, 12, \ldots$. *Ann. of Math.* **160**, (2004) 465–491.

[5] Benjamini, I., Lyons, R., Peres, Y. and Schramm, O.: Uniform spanning forests. Ann. Probab.**29**, (2001) 1–65.

[6] Bollobás, B.: *Modern graph Theory.* Graduate Texts in Mathematics, Vol. **184**, Springer (1998).

[7] Broder, A.: Generating random spanning trees. In *Symp. Foundations of Computer Sci.*, IEEE, New York, (1989) 442–447.

[8] Burton, R. and Pemantle, R.: Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances. Ann. Probab.**21**, (1993) 1329–1371.

[9] Dhar, D.: Self-organized critical state of sandpile automaton models. *Phys. Rev. Lett.* **64**, (1990) 1613–1616.

[10] Dhar, D.: Theoretical studies of self-organized criticality. *Phys. A* **369**, (2006) 29–70.

[11] Doyle, P.G. and Snell, J.L.: *Random walks and electric networks.* Mathematical Association of America, Washington (1984).

[12] Durrett, R.: *Probability: Theory and Examples.* Second edition. Duxbury Press, Belmont, CA, USA (1996).

[13] Følner, E.: On groups with full Banach mean value. *Math. Scand.* **3**, (1955) 243–254.

[14] Gabrielov, A.: Abelian avalanches and Tutte polynomials. *Phys. A* **195**, (1993) 253–274.

[15] Häggström, O.: Random-cluster measures and uniform spanning trees. *Stochastic Process. Appl.* **59**, (1995) 267–275.

[16] Holroyd, A.E., Levine, L., Mészáros, K., Peres, Y., Propp, J. and Wilson, D.B.: Chip-firing and rotor-routing on directed graphs. In: Sidoravicius, V. and Vares, M.E. (eds.) *In and out of equilibrium. 2.* Progr. Probab. Vol. **60**, Brikhäuser, Basel (2008) 331–364.

[17] Járai, A.A.: Thermodynamic limit of the Abelian sandpile model on $\mathbb{Z}^d$. *Marov Process. Related Fields* **11**, (2005) 313–336.

[18] Járai A.A. and Redig, F.: Infinite volume limit of the Abelian sandpile model in dimensions $d \geq 3$. *Probab. Theory Related Fields* **141**, (2008) 181–212.

[19] Kirchhoff, G.: Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird. *Ann. Phys. Chem.* **72**, (1847) 497–508.

[20] Kolmogorov, A.N. and Fomīn, S.V.: *Introductory real ananlysis.* Dover Publications, Inc., New York (1975).

[21] Kozma, G.: The scaling limit of loop-erased random walk in three dimensions. `arXiv:math/0508344v2`.

[22] Lawler, G.F.: *Intersections of random walks.* Probability and its Applications. Birkhäuser, Boston (1991).

[23] Lawler, G.F.: Loop-erased random walk. In: Bramson, M. and Durrett, R. (eds.) *Perplexing Problems in Probability.* Birkhäuser (1999).

[24] Lawler, G.F., Schramm, O. and Werner, W.: Conformal invariance of planar loop-erased random walks and uniform spanning trees. *Ann. Probab.* **32** (2004) 939–995.

[25] Lax, P.D.: *Functional Analysis.* Wiley (2002).

[26] Lyons, R. with Peres, Y.: Probability on trees and networks. Cambridge University Press. In preparation. Current version available at `http://mypage.iu.edu/~rdlyons/`.

[27] Lyons, R., Peres, Y. and Schramm, O.: Markov chain intersections and the loop-erased walk. *Ann. Inst. H. Poincaré Probab. Statist.* **39**, (2003) 779–791.

[28] Majumdar, S.N. and Dhar, D.: equivalence between the Abeliand sandpile model and the $q \to 0$ limit of the Potts model. *Phys. A* **185**, (1992) 129–145.

[29] Mörters, P. and Peres, Y.: *Brownian motion.* Cambridge Series in Statistical and Probabilisitic Mathematics Vol. **30**. To appear.

[30] Pemantle, R.: Choosing a spanning tree for the integer lattice uniformly. Ann. Probab.**19**, (1991) 1559–1574.

[31] Redig, F.: Mathematical apsects of the abelian sandpile model. In: Bovier, A., van Enter, A. and Dunlop, F. (eds.): *Mathematical statistical physics: Lecture notes of the Les Houches Summer School 2005.* Elsevier (2006).

[32] Strassen, V.: The existence of probability measures with given marginals. *Ann. Math. Statist.* **36**, (1965) 423–439.

[33] Werner, W.: *Random planar curves and Schramm-Loewner evolutions.* Lecture Notes on Mathematics, Vol. **1840**, Springer (2004).

[34] Wilson, D.B.: Generating random spanning trees more quickly than the cover time. *Proceedings of the Twenty-eights Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996)*, ACM, New York, (1996) 296–303.