1 Introduction

There are several good reasons you might want to read about uniform spanning trees, one being that spanning trees are useful combinatorial objects. Not only are they fundamental in algebraic graph theory and combinatorial geometry, but they predate both of these subjects, having been used by Kirchoff in the study of resistor networks. This article addresses the question about spanning trees most natural to anyone in probability theory, namely what does a typical spanning tree look like?

Some readers will be happy to know that understanding the basic questions requires no background knowledge or technical expertise. While the model is elementary, the answers are surprisingly rich. The combination of a simple question and a structurally complex answer is sometimes taken to be the quintessential mathematical experience. This nonwithstanding, I think the best reason to set out on a mathematical odyssey is to enjoy the ride. Answering the basic questions about spanning trees depends on a sort of vertical integration of techniques and results from diverse realms of probability theory and discrete mathematics. Some of the topics encountered en route are random walks, resistor networks, discrete harmonic analysis, stationary Markov chains, circulant matrices, inclusion-exclusion, branching processes and the method of moments. Also touched on are characters of abelian groups, entropy and the infamous incipient infinite cluster.

The introductory section defines the model and previews some of the connections to these other topics. The remaining sections develop these at length. Explanations of jargon and results borrowed from other fields are provided whenever possible. Complete proofs are given in most cases, as appropriate.

1.1 Defining the model

Begin with a finite graph G. That means a finite collection $V(G)$ of *vertices* along with a finite collection $E(G)$ of *edges*. Each edge either connects two vertices v and $w \in V(G)$ or else is a *self-edge*, connecting some $v \in V(G)$ to itself. There may be more than one edge connecting a pair of vertices. Edges are said to be *incident* to the vertices they connect. To make the notation less cumbersome we will write $v \in G$ and $e \in G$ instead of $v \in V(G)$ and $e \in E(G)$. For $v, w \in G$ say v is a *neighbor* of w, written $v \sim w$ if and only if some edge connects v and w. Here is an example of a graph G_1 which will serve often as an illustration.
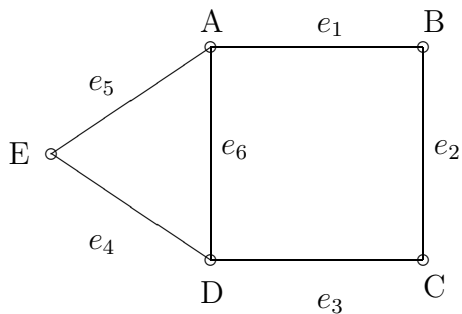

figure 1

Its vertex set is $\{A, B, C, D, E\}$ and it has six edges $e_1, \ldots, e_6$, none of which is a self-edge.

A subgraph of a graph G will mean a graph with the same vertex set but only a subset of the edges. (This differs from standard usage which allows the vertex set to be a subset as well.) Since G_1 has 6 edges, there are $2^6 = 64$ possible different subgraphs of G_1. A subgraph $H \subseteq G$ is said to be a *forest* if there are no cycles, i.e. you cannot find a sequence of vertices $v_1, \ldots, v_k$ for which there are edges in H connecting v_i to v_{i+1} for each $i < k$ and an edge connecting v_k to v_1. In particular ($k = 1$) there are no self-edges in a forest. A *tree* is a forest that is connected, i.e. for any v and w there is a path of edges that connects them. The *components* of a graph are the maximal connected subgraphs, so for example the components of a forest are trees. A *spanning* forest is a

forest in which every vertex has at least one incident edge; a *spanning tree* is a tree in which every vertex has at least one incident edge. If G is connected (and all our graphs will be) then a spanning tree is just a subgraph with no cycles such that the addition of any other edge would create a cycle. From this it is easy to see that every connected graph has at least one spanning tree.

Now if G is any finite connected graph, imagine listing all of its spanning trees (there are only finitely many) and then choosing one of them at random with an equal probability of choosing any one. Call this random choice $\mathbf{T}$ and say that $\mathbf{T}$ is a *uniform random spanning tree* for G. In the above example there are eleven spanning trees for G_1 given (in the obvious notation) as follows:

$$e_1 e_2 e_3 e_4 \quad e_1 e_2 e_3 e_5 \quad e_1 e_2 e_4 e_5 \quad e_1 e_3 e_4 e_5$$

$$e_2 e_3 e_4 e_5 \quad e_1 e_2 e_4 e_6 \quad e_1 e_3 e_4 e_6 \quad e_2 e_3 e_4 e_6$$

$$e_1 e_2 e_5 e_6 \quad e_1 e_3 e_5 e_6 \quad e_2 e_3 e_5 e_6$$

In this case, $\mathbf{T}$ is just one of these eleven trees, picked with uniform probability. The model is so simple, you may wonder what there is to say about it! One answer is that the model has some properties that are easy to state but hard to prove; these are introduced in the coming subsections. Another answer is that the definition of a uniform random spanning tree does not give us a way of readily computing local characteristics of the random tree. To phrase this as a question: can you compute probabilities of events local to a small set of edges, such as $\mathbf{P}(e_1 \in \mathbf{T})$ or $\mathbf{P}(e_1, e_4 \in \mathbf{T})$ without actually enumerating all of the spanning trees of G? In a sense, most of the article is devoted to answering this question. (Events such as e_1 being in the tree are called local in contrast to a global event such as the tree having diameter – longest path between two vertices – at most three.)

1.2 Uniform spanning trees have negative correlations

Continuing the example in figure 1, suppose I calculate the probability that $e_1 \in \mathbf{T}$. That's easy: there are 8 spanning trees containing e_1, so

$$\mathbf{P}(e_1 \in \mathbf{T}) = \frac{8}{11}.$$

Similarly there are 7 spanning trees containing e_4 so

$$\mathbf{P}(e_4 \in \mathbf{T}) = \frac{7}{11}.$$

There are only 4 spanning trees containing both e_1 and e_4, so

$$\mathbf{P}(e_1 \in \mathbf{T} \text{ and } e_4 \in \mathbf{T}) = \frac{4}{11}.$$

Compare the probability of both of these edges being in the tree with the product of the probabilities of each of the edges being in the tree:

$$\frac{8}{11} \cdot \frac{7}{11} = 56/121 > \frac{4}{11}.$$

Thus

$$\mathbf{P}(e_1 \in \mathbf{T} \,|\, e_4 \in \mathbf{T}) = \frac{\mathbf{P}(e_1 \in \mathbf{T} \text{ and } e_4 \in \mathbf{T})}{\mathbf{P}(e_4 \in \mathbf{T})} < \mathbf{P}(e_1 \in \mathbf{T})$$

or in words, the conditional probability of e_1 being in the tree if you know that e_4 is in the tree is less than the original unconditional probability. This negative correlation of edges holds in general, with the inequality not necessarily strict.

Theorem 1.1 *For any finite connected graph G, let $\mathbf{T}$ be a uniform spanning tree. If e and f are distinct edges, then $\mathbf{P}(e, f \in \mathbf{T}) \leq \mathbf{P}(e \in \mathbf{T})\mathbf{P}(f \in \mathbf{T})$.*

Any spanning tree of an n-vertx graph contains $n - 1$ edges, so it should seem intuitively plausible – even obvious – that if one edge is forced to be in the tree then any other

4

edge is less likely to be needed. Two proofs will be given later, but neither is straightforward, and in fact the only proofs I know involve elaborate connections between spanning trees, random walks and electrical networks. Sections 2 and 3 will be occupied with the elucidation of these connections. The connection between random walks and electrical networks will be given more briefly, since an excellent treatment is available [8].

As an indication that the previous theorem is not trivial, here is a slightly stronger statement, the truth or falsity of which is unknown. Think of the distribution of $\mathbf{T}$ as a probability distribution on the outcome space Ω consisting of all the $2^{|E(G)|}$ subgraphs of G that just happens to give probability zero to any subgraph that is not a spanning tree. An event A (i.e. any subset of the outcome space) is called an *up-event* – short for *upwardly closed* – if whenever a subgraph H of G has a further subgraph K and $K \in A$, then $H \in A$. An example of an up-event is the event of containing at least two of the three edges e_1, e_3 and e_5. Say an event A ignores an edge e if for every H, $H \in A \Leftrightarrow H \cup e \in A$.

Conjecture 1 *For any finite connected graph G, let $\mathbf{T}$ be a uniform spanning tree. Let e be any edge and A be any up-event that ignores e. Then*

$$\mathbf{P}(A \text{ and } e \in \mathbf{T}) \leq \mathbf{P}(A)\mathbf{P}(e \in \mathbf{T}).$$

Theorem 1.1 is a special case of this when A is the event of f being in the tree. The conjecture is known to be true for *series-parallel* graphs and it is also know to be true in the case when A is an *elementary cylinder event*, i.e. the event of containing some fixed $e_1, \ldots, e_k$. On the negative side, there are natural generalizations of graphs and spanning trees, namely *matroids* and *bases* (see [19] for definitions), and both Theorem 1.1 and Conjecture 1 fail to generalize to this setting. If you're interested in seeing the counterexample, look at the end of [15].

1.3 The transfer-impedance matrix

The next two paragraphs discuss a theorem that computes probabilities such as $\mathbf{P}(e, f \in \mathbf{T})$. These computations alone would render the theorem useful, but it appears even more powerful in the context of how strongly it constrains the probability measure governing $\mathbf{T}$. Let me elaborate.

Fix a subset $S = \{e_1, \ldots, e_k\}$ of the edges of a finite connected graph G. If $\mathbf{T}$ is a uniform random spanning tree of G then the knowledge of whether $e_i \in \mathbf{T}$ for each i partitions the space into 2^k possible outcomes. (Some of these may have probability zero if S contain cycles, but if not, all 2^k may be possible.) In any case, choosing $\mathbf{T}$ from the uniform distribution on spanning trees of G induces a probability distribution on Ω, the space of these 2^k outcomes. There are many possible probability distributions on Ω: the ways of choosing 2^k nonnegative numbers summing to one are a $2^k - 1$-dimensional space. Theorem 1.1 shows that the actual measure induced by $\mathbf{T}$ satisfies certain inequalities, so not all probability distributions on Ω can be gotten in this way. But the set of probability distributions on Ω satisfying these inequalities is still $2^k - 1$-dimensional. It turns out, however, that the set of probability distributions on Ω that arise as induced distributions of uniform spanning trees on subsets of k edges actually has at most the much smaller dimension $k(k+1)/2$. This is a consequence of the following theorem which is the bulwark of our entire discussion of spanning trees:

Theorem 1.2 (Transfer-Impedance Theorem) *Let G be any finite connected graph. There is a symmetric function $H(e, f)$ on pairs of edges in G such that for any $e_1, \ldots, e_r \in G$,*

$$\mathbf{P}(e_1, \ldots, e_r \in \mathbf{T}) = \det M(e_1, \ldots, e_r)$$

where $M(e_1, \ldots, e_r)$ is the r by r matrix whose i, j-entry is $H(e_i, e_j)$.

By inclusion-exclusion, the probability of any event in Ω may be determined from the

probabilities of $\mathbf{P}(e_{j_1}, \ldots, e_{j_r} \in \mathbf{T})$ as $e_{j_1}, \ldots, e_{j_r}$ vary over all subsets of $e_1, \ldots, e_k$. The theorem says that these are all determined by the $k(k+1)/2$ numbers $\{H(e_i, e_j) : i, j \le k\}$, which shows that there are indeed only $k(k+1)/2$ degrees of freedom in determining the measure on Ω.

Another way of saying this is that the measure is almost completely determined by its two-dimensional marginals, i.e. from the values of $\mathbf{P}(e, f \in \mathbf{T})$ as e and f vary over pairs of (not necessarily distinct) edges. To see this, calculate the values of $H(e, f)$. The values of $H(e, e)$ in the theorem must be equal to $\mathbf{P}(e \in \mathbf{T})$ since $\mathbf{P}(e, e) = \det M(e) = H(e, e)$. To see what $H(e, f)$ is for $e \ne f$, write

$$
\begin{aligned}
\mathbf{P}(e, f \in \mathbf{T}) &= \det M(e, f) \\
&= H(e, e)H(e, f) - H(e, f)^2 \\
&= \mathbf{P}(e \in \mathbf{T})\mathbf{P}(f \in \mathbf{T}) - H(e, f)^2
\end{aligned}
$$

and hence

$$
H(e, f) = \pm\sqrt{\mathbf{P}(e \in \mathbf{T})\mathbf{P}(f \in \mathbf{T}) - \mathbf{P}(e, f \in \mathbf{T})}.
$$

Thus the two dimension marginals determine H up to sign, and H determines the measure. Note that the above square root is always real, since by Theorem 1.1 the quantity under the radical is nonnegative. Section 4 will be devoted to proving Theorem 1.2, the proof depending heavily on the connections to random walks and electrical networks developed in Sections 2 and 3.

1.4 Applications of transfer-impedance to limit theorems

Let K_n denote the complete graph on n vertices, i.e. there are no self-edges and precisely one edge connecting each pair of distinct vertices. Imagine picking a uniform random

spanning tree of K_n and letting n grow to infinity. What kind of limit theorem might we expect? Since a spanning tree of K_n has only $n-1$ edges, each of the $n(n-1)/2$ edges should have probability $2/n$ of being in the tree (by symmetry) and is hence decreasingly likely to be included as $n \to \infty$. On the other hand, the number of edges incident to each vertex is increasing. Say we fix a particular vertex v_n in each K_n and look at the number of edges incident to v_n that are included in the tree. Each of $n-1$ incident edges has probability $2/n$ of being included, so the expected number of of such edges is $2(n-1)/n$, which is evidently converging to 2. If the inclusion of each of these $n-1$ edges in the tree were independent of each other, then the number of edges incident to v_n in $\mathbf{T}$ would be a binomial random variable with parameters $(n-1, 2/n)$; the well known Poisson limit theorem would then say that the random variable $D_{\mathbf{T}}(v_n)$ counting how many edges incident to v_n are in $\mathbf{T}$ converged as $n \to \infty$ to a Poisson distribution with mean two. (A quick explanation: integer-valued random variables X_n are said to converge to X in distribution if $\mathbf{P}(X_n = k) \to \mathbf{P}(X = k)$ for all integers k. In this instance, convergence of $D_{\mathbf{T}}(v_n)$ to a Poisson of mean two would mean that for each k, $\mathbf{P}(D_{\mathbf{T}}(v_n) = k) \to e^{-2}k^2/2$ as $n \to \infty$ for each integer k.) Unfortunately this can't be true because a Poisson(2) is sometimes zero, whereas $D_{\mathbf{T}}(v_n)$ can never be zero. It has however been shown [2] that $D_{\mathbf{T}}(v_n)$ converges in distribution to the next simplest thing: one plus a Poisson of mean one.

To show you why this really is the next best thing, let me point out a property of the mean one Poisson distribution. Pretend that if you picked a family in the United States at random, then the number of children in the family would have a Poisson distribution with mean one (population control having apparently succeeded). Now imagine picking a child at random instead of picking a family at random, and asking how many children in the family. You would certainly get a different distribution, since you couldn't ever get the answer zero. In fact you would get one plus a Poisson of mean one. (Poisson distributions are the only ones with this property.) Thus a Poisson-plus-one distribution is a more natural distribution than it looks at first. At any rate, the convergence theorem

is

Theorem 1.3 *Let $D_{\mathbf{T}}(v_n)$ be the random degree of the vertex v_n in a uniform spanning tree of K_n. Then as $n \to \infty$, $D_{\mathbf{T}}(v_n)$ converges in distribution to X where X is one plus a Poisson of mean one.*

Consider now the n-cube B_n. Its vertices are defined to be all strings of zeros and ones of length n, where two vertices are connected by an edge if and only if they differ in precisely one location. Fix a vertex $v_n \in B_n$ and play the same game: choose a uniform random spanning tree and let $D_{\mathbf{T}}(v_n)$ be the random degree of v_n in the tree. It is not hard to see again that the expected value, $\mathbf{E}D$, converges to 2 as $n \to \infty$. Indeed, for any graph the number of vertices in a spanning tree is one less than the number of vertices, and since each edge has two endpoints the average degree of the vertices will be ≈ 2; if the graph is symmetric, each vertex will then have the same expected degree which must be 2. One could expect Theorem 1.3 to hold for B_n as well as K_n and in fact it does. A proof of this for a class of sequences of graphs that includes both K_n and B_n and does not use transfer-impedances appears in [2] along with the conjecture that the result should hold for more general sequences of graphs. This can indeed be established, and in Section 5 we will discuss the proof of Theorem 1.3 via transfer-impedances which can be extended to more general sequences of graphs.

The convergence in distribution of $D_{\mathbf{T}}(v_n)$ in these theorems is actually a special case of a stronger kind of convergence. To begin discussing this stronger kind of convergence, imagine that we pick a uniform random spanning tree of a graph, say K_n, and want to write down what it looks like "near v_n". Interpret "near v_n" to mean within a distance of r of v_n, where r is some arbitrary positive integer. The answer will be a rooted tree of height r. (A *rooted* tree is a tree plus a choice of one of its vertices, called the root. The *height* of a rooted tree is the maximum distance of any vertex from the root.) The

rooted tree representing $\mathbf{T}$ near v_n will be the tree you get by picking up $\mathbf{T}$, dangling it from v_n, and ignoring everything more than r levels below the top.

Call this the *r-truncation* of $\mathbf{T}$, written $\mathbf{T} \wedge_{v_n} r$ or just $\mathbf{T} \wedge r$ when the choice of v_n is obvious. For example, suppose $r = 2$, v_n has 2 neighbors in $\mathbf{T}$, w_1 and w_2, w_1 has 3 neighbors other than v_n in $\mathbf{T}$ and w_2 has none. This information is encoded in the following picture. The picture could also have been drawn with left and right reversed, since we consider this to be the same abstract tree, no matter how it is drawn.
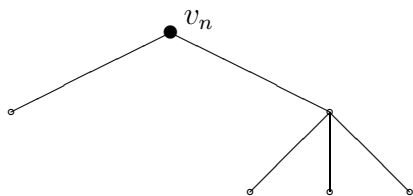

figure 2

When $r = 1$, the only information in $\mathbf{T} \wedge r$ is the number of children of the root, i.e. $D_{\mathbf{T}}(v_n)$. Thus the previous theorem asserts the convergence in distribution of $\mathbf{T} \wedge_{v_n} 1$ to a root with a (1+Poisson) number of vertices. Generalizing this is the following theorem, proved in Section 5.

Theorem 1.4 *For any $r \geq 1$, as $n \to \infty$, $\mathbf{T} \wedge_{v_n} r$ converges in distribution to a particular random tree, $\mathcal{P}_1 \wedge r$ to be defined later.*

Convergence in distribution means that for any fixed tree t of height at most r, $\mathbf{P}(\mathbf{T} \wedge_{v_n} r = t)$ converges as $n \to \infty$ to the probability of the random tree $\mathcal{P}_1 \wedge r$ equalling t. As the notation indicates, the random tree $\mathcal{P}_1 \wedge r$ is the r-truncation of an infinite random tree. It is in fact the tree of a Poisson(1) branching process conditioned to live forever, but these terms will be defined later, in Section 5. The theorem is stated here only for

10

the sequence K_n, but is in fact true for a more general class of sequences, which includes B_n.

2 Spanning trees and random walks

Unless G is a very small graph, it is virtually impossible to list all of its spanning trees. For example, if $G = K_n$ is the complete graph on n vertices, then the number of spanning trees is n^{n-2} according to the well known Prüfer bijection [17]. If n is much bigger than say 20, this is too many to be enumerated even by the snazziest computer that ever will be. Luckily, there are shortcuts which enable us to compute probabilities such as $\mathbf{P}(e \in \mathbf{T})$ without actually enumerating all spanning trees and counting the proportion containing e. The shortcuts are based on a close correspondence between spanning trees and random walks, which is the subject of this section.

2.1 Simple random walk

Begin by defining a simple random walk on G. To avoid obscuring the issue, we will place extra assumptions on the graph G and later indicate how to remove these. In particular, in addition to assuming that G is finite and connected, we will often suppose that it is D-regular for some positive integer D, which means that every vertex has precisely D edges incident to it. Also suppose that G is simple, i.e. it has no self-edges or parallel edges (different edges connecting the same pair of vertices). For any vertex $x \in G$, define a simple random walk on G starting at x, written SRW_x^G, intuitively as follows. Imagine a particle beginning at time 0 at the vertex x. At each future time $1, 2, 3, \ldots$, it moves along some edge, always choosing among the D edges incident to the vertex it is currently at with equal probability. When G is not D-regular, the definition will be the same: each of the edges leading away from the current position

will be chosen with probability $1/\text{degree}(v)$. This defines a sequence of random positions $SRW_x^G(0), SRW_x^G(1), SRW_x^G(2), \ldots$ which is thus a random function SRW_x^G (or just SRW if x and G may be understood without ambiguity) from the nonnegative integers to the vertices of G. Formally, this random function may be defined by its finite-dimensional marginals which are given by $\mathbf{P}(SRW_x^G(0) = y_0, SRW_x^G(1) = y_1, \ldots, SRW_x^G(k) = y_k) = D^{-k}$ if $y_0 = x$ and for all $i = 1, \ldots, k$ there is an edge from y_{i-1} to y_i, and zero otherwise. For an illustration of this definition, let G be the following 3-regular simple graph.
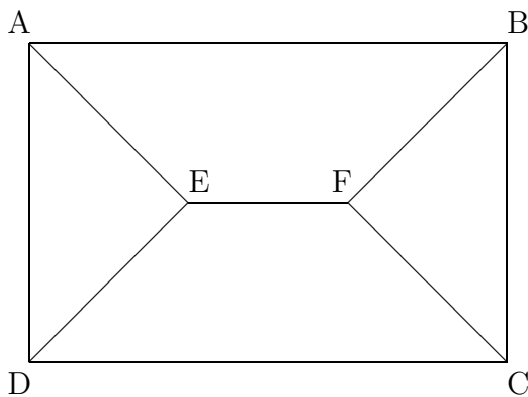

figure 3

Consider a simple random walk SRW_A^G starting at the vertex A. The probability of a particular beginning, say $SRW(1) = B$ and $SRW(2) = F$ is just $(1/3)^2$. The random position at time 2, $SRW(2)$, is then equal to F with probability $2/9$, since each of the two ways, ABF and AEF, of getting to F in two steps has probability $1/9$.

Another variant of random walk we will need is the *stationary Markov chain* corresponding to a simple random walk on G. I will preface this definition with a quick explanation of Markov chains; since I cannot do justice to this large topic in two paragraphs, the reader is referred to [11], [9] or any other favorite introductory probability text for further details.

A (time-homogeneous) *Markov chain* on a finite state space S is a sequence of ran-

dom variables $\{X_i\}$ taking values in S, indexed by either the integers or the nonnegative integers and having the Markov property: there is a set of transition probabilities $\{p(x, y) : x, y \in S\}$ so that the probability of X_{i+1} being y, conditional upon $X_i = x$, is always equal to $p(x, y)$ regardless of how much more information about the past you have. (Formally, this means $\mathbf{P}(X_{i+1} = y \mid X_i = x$ and any values of X_j for $j < i)$ is still $p(x, y)$.) An example of this is SRW_x^G, where S is the set of vertices of G and $p(x, y) = D^{-1}$ if $x \sim y$ and 0 otherwise (recall that $x \sim y$ means x is a neighbor of y). The values $p(x, y)$ must satisfy $\sum_y p(x, y) = 1$ for every x in order to be legitimate conditional probabilities. If in addition they satisfy $\sum_x p(x, y) = 1$ for every y, the Markov chain is said to be *doubly stochastic*. It will be useful later to know that the Markov property is time-reversible, meaning if $\{X_i\}$ is a Markov chaing then so is the sequence $\{\tilde{X}_i = X_{-i}\}$, and there are backwards transition probabilities $\tilde{p}(x, y)$ for which $\mathbf{P}(X_{i-1} = y \mid X_i = x) = \tilde{p}(x, y)$.

If it is possible eventually to get from every state in S to every other, then there is a unique *stationary distribution* which is a set of probabilities $\{\pi(x) : x \in S\}$ summing to one and having the property that $\sum_x \pi(x) p(x, y) = \pi(y)$ for all y. Intuitively, this means that if we build a Markov chain with transition probabilities $p(x, y)$ and start it by randomizing X_0 so that $\mathbf{P}(X_0 = x) = \pi(x)$ then it will also be true that $\mathbf{P}(X_i = x) = \pi(x)$ for every $i > 0$. A *stationary* Markov chain is one indexed by the integers (as opposed to just the positive integers), in which $\mathbf{P}(X_i = x) = \pi(x)$ for some, hence every i. If a Markov chain is doubly stochastic, it is easy to check that the uniform distribution U is stationary:

$$\sum_x U(x) p(x, y) = \sum_x |S|^{-1} p(x, y) = |S|^{-1} = U(y).$$

The stationary distribution π is unique (assuming every state can get to every other) and is hence uniform over all states.

Now we define a stationary simple random walk on G to be a stationary Markov chain with state space $V(G)$ and transition probabilities $p(x, y) = D^{-1}$ if $x \sim y$ and 0 otherwise. Intuitively this can be built by choosing X_0 at random uniformly over $V(G)$,

then choosing the X_i for $i > 0$ by walking randomly from X_0 along the edges and choosing the X_i for $i < 0$ also by walking randomly from X_0, thinking of this latter walk as going backwards in time. (For SRW, $p(x, y) = p(y, x) = \tilde{p}(x, y)$ so the walk looks the same backwards as forwards.)

2.2 The random walk construction of uniform spanning trees

Now we are ready for the random walk construction of uniform random spanning trees. What we will actually get is a directed spanning tree, which is a spanning tree together with a choice of vertex called the root and an orientation on each edge (an arrow pointing along the edge in one of the two possible directions) such that following the arrows always leads to the root. Of course a directed spanning tree yields an ordinary spanning tree if you ignore the arrows and the root. Here is an algorithm to generate directed trees from random walks.

GROUNDSKEEPER'S ALGORITHM

Let G be a finite, connected, D-regular, simple graph and let x be any vertex of G. Imagine that we send the groundskeeper from the local baseball diamond on a walk along the edges of G starting from x; later we will take to be the walk SRW_x^G. She brings with her the wheelbarrow full of chalk used to draw in lines. This groundskeeper is so eager to choose a spanning tree for G that she wants to chalk a line over each edge she walks along. Of course if that edge, along with the edges she's already chalked, would form a cycle (or is already chalked), she is not allowed to chalk it. In this case she continues walking that edge but temporarily – and reluctantly – shuts off the flow of chalk. Every time she chalks a new edge she inscribes an arrow pointing from the new vertex back to the old.

Eventually every vertex is connected to every other by a chalked path, so no more can be added without forming a cycle and the chalking is complete. It is easy to see that the subgraph consisting of chalked edges is always a single connected component. The first time the walk reaches a vertex y, the edge just travelled cannot form a cycle with the other chalked edges. Conversely, if the walk moves from z to some y that has been reached before, then y is connected to z already by some chalked path, so adding the edge zy would create a cycle and is not permitted. Also it is clear that following the arrows leads always to vertices that were visited previously, and hence eventually back to the root. Furthermore, every vertex except x has exactly one oriented edge leading out of it, namely the edge along which the vertex was first reached.

Putting this all together, we have defined a function – say τ – from walks on G (infinite sequences of vertices each consecutive pair connected by an edge) to directed spanning trees of G. Formally $\tau(y_0, y_1, y_2, \ldots)$ is the subgraph $H \subseteq G$ such that if e is an oriented edge from w to z then

$e \in H \Leftrightarrow$ for some $k > 0, y_k = z, y_{k-1} = w$, and there is no $j < k$ such that $y_j = z$.

As an example, suppose SRW_A^G in figure 2.1 begins ABFBCDAE. Then applying τ gives the tree with edges BA, FB, CB, DC and EA.

To be completely formal, I should admit that the groundskeeper's algorithm never stops if there is a vertex that the walk fails to hit in finite time. This is not a problem since we are going to apply τ to the path of a SRW, and this hits every vertex with probability one. As hinted earlier, the importance of this construction is the following equivalence.

Theorem 2.1 *Let G be any finite, connected, D-regular, simple graph and let x be any vertex of G. Run a simple random walk SRW_x^G and let $\mathbf{T}$ be the random spanning tree gotten by ignoring the arrows and root of the random directed spanning tree $\tau(SRW_x^G)$. Then $\mathbf{T}$ has the distribution of a uniform random spanning tree.*

15

To prove this it is necessary to consider a stationary simple random walk on G ($SSRW^G$). It will be easy to get back to a SRW_x^G because the sites visited in positive time by a $SSRW^G$ conditioned on being at x at time zero form a SRW_x^G. Let T_n be the tree $\tau(SSRW(n), SSRW(n+1), \ldots)$; in other words, T_n is the directed tree gotten by applying the groundskeeper's algorithm to the portion of the stationary simple random walk from time n onwards. The first goal is to show that the random collection of directed trees T_n forms a time-homogeneous Markov chain as n ranges over all integers.

Showing this is pretty straightforward because the transition probabilities are easy to see. First note that if t and u are any two directed trees on disjoint sets of vertices, rooted respectively at v and w, then adding any arrow from v to a vertex in u combines them into a single tree rooted at w. Now define two operations on directed spanning trees of G as follows.

Operation $F(t, x)$: *Start with a directed tree t rooted at v. Choose one of the the D neighbors of v in G, say x. Take away the edge in t that leads out of x, separating t into two trees, rooted at v and x. Now add an edge from v to x, resulting in a single tree $F(t, x)$.*

Operation $F^{-1}(t, w)$: *Start with a directed tree t rooted at x. Choose one of the the D neighbors of x in G, say w. Follow the path from w to x in t and let v be the last vertex on this path before x. Take away the edge in t that leads out of v, separating t into two trees, rooted at x and v. Now add an edge from x to w, resulting in a single directed tree $F^{-1}(t, w)$.*

It is easy to see that these operations really are inverse to each other, i.e. if t is rooted at v then $F^{-1}(F(t, x), w) = t$ for any $x \sim v$, where w is the other endpoint of the edge leading out of x in t. Here is a pictorial example.
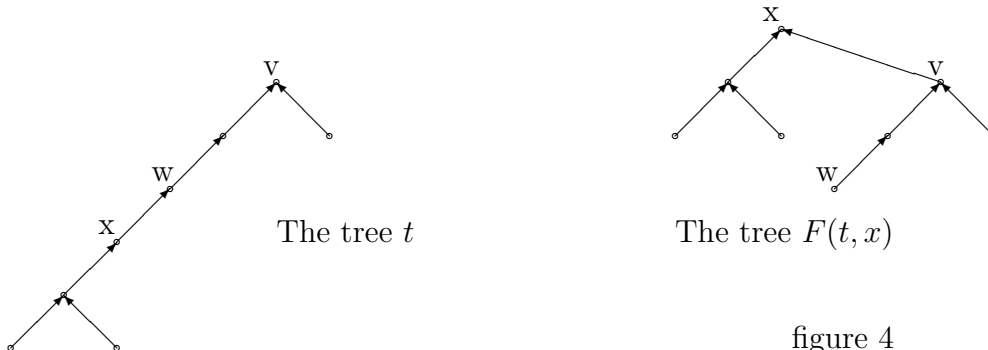
The tree t The tree $F(t,x)$

figure 4

I claim that for any directed trees t and u, the backward transition probability $\tilde{p}(t,u)$ is equal to D^{-1} if $u = F(t,x)$ for some x and zero otherwise. To see this, it is just a matter of realizing where the operation F comes from. Remember that T_n is just $\tau(SSRW(n), SSRW(n+1), \ldots)$, so in particular the root of T_n is $SSRW(n)$. Now $SSRW$ really is a Markov chain. We already know that $\mathbf{P}(SSRW(n-1) = x \,|\, SSRW(n) = v)$ is D^{-1} if $x \sim v$ and zero otherwise. Also, this is unaffected by knowledge of $SSRW(j)$ for any $j > n$. Suppose it turns out that $SSRW(n-1) = x$. Then knowing only T_n and x (but not the values of $SSRW(j)$ for $j > n$) it is possible to work out what T_{n-1} is. Remember that T_n and T_{n-1} come from applying τ to the respective sequences $SSRW(n), SSRW(n+1), \ldots$ and $SSRW(n-1), SSRW(n), \ldots$ whose only difference is that the second of these has an extra x tacked on the beginning. Every time the first sequence reaches a vertex for the first time, so does the second, unless that vertex happens to be x. So the T_{n-1} has has all the oriented edges of T_n except the one out of x. What it has instead is an oriented edge from v to x, chalked in by the groundskeeper at her very first step. Adding in the edge from v to some neighbor x and erasing the edge out of x yields precisely $F(t,x)$. So we have shown that $T_{n-1} = F(T_n, SSRW(n-1))$. But $SSRW(n-1)$ is uniformly distributed among the neighbors of $SSRW(n)$ no matter what other information we know about the future. This proves the claim and the time-homogeneous Markov property.

17

The next thing to show is that the stationary distribution is uniform over all directed trees. As we've seen, this would follow if we knew that $\{T_n\}$ was doubly stochastic. Since $p(t, u)$ is D^{-1} whenever $u = F(t, x)$ for some x and zero otherwise, this would be true if for every tree u there are precisely D trees t for which $F(t, x) = u$ for some x. But the trees t for which $F(t, x) = u$ for some x are precisely the trees $F^{-1}(u, x)$ for some neighbor x of the root of u, hence there are D such trees and transition probabilities for $SSRW$ are doubly stochastic.

Now that the stationary distribution for $\{T_n\}$ has been shown to be uniform, the proof of Theorem 2.1 is almost done. Note that the event $SSRW(0) = x$ is the same as the event of $\tau(SSRW(0), SSRW(1), \ldots)$ being rooted at x. Since SRW_x^G is just $SSRW$ conditioned on $SSRW(0) = x$, $T_0(SRW_x^G)$ is distributed as a uniform directed spanning tree conditioned on being rooted at x. That is to say, $T_0(SRW_x^G)$ is uniformly distributed over all directed spanning trees rooted at x. But ordinary spanning trees are in a one to one correspondence with directed spanning trees rooted at a fixed vertex x, the correspondence being that to get from the ordinary tree to the directed tree you name x as the root and add arrows that point toward x. Then the tree $\mathbf{T}$ gotten from $T_0(SRW_x^G)$ by ignoring the root and the arrows is uniformly distributed over all ordinary spanning trees of G, which is what we wanted to prove. $\square$

2.3   Weighted graphs

It is time to remove the extra assumptions that G is D-regular and simple. It will make sense later to generalize from graphs to weighted graphs, and since the generalization of Theorem 2.1 is as easy for weighted graphs as for unweighted graphs, we may as well introduce weights now.

A weighted graph is just a graph to each edge e of which is assigned a positive real number called its weight and written $w(e)$. Edge weights are not allowed to be zero,

though one may conceptually identify a graph with an edge of weight zero with the same graph minus the edge in question. An unweighted graph may be thought of as a graph with all edge weights equal to one, as will be clear from the way random trees and random walks generalize. Write $d(v)$ for the sum of the weights of all edges incident to v. Corresponding to the old notion of a uniform random spanning tree is the *weight-selected* random spanning tree (WST). A WST, $\mathbf{T}$ is defined to have

$$\mathbf{P}(\mathbf{T} = t) = \frac{\prod_{e \in t} w(e)}{\sum_u \prod_{e \in u} w(e)}$$

so that the probability of any individual tree is proportional to its weight which is by definition the product of the weights of its edges.

Corresponding to a simple random walk from a vertex x is the weighted random walk from x, WRW_x^G which is a Markov Chain in which the transition probabilities from a vertex v are proportional to the weights of the edges incident to v (among which the walk must choose). Thus if v has two neighbors w and x, and there are four edges incident to v with respective weights $1, 2, 3$ and 4 that connect v respectively to itself, w, x and x, then the probabilities of choosing these four edges are respectively $1/10, 2/10, 3/10$ and $4/10$. Formally, the probability of walking along an edge e incident to the current position v is given by $w(e)/d(v)$. The bookkeeping is a little unwieldly since knowing the set of vertices $WRW(0), WRW(1), \ldots$ visited by the WRW does not necessarily determine which edges were travelled now that the graph is not required to be simple. Rather than invent some clumsy *ad hoc* notation to include the edges, it is easier just to think that a WRW includes this information, so it is <u>not</u> simply given by its positions $WRW(j) : j \geq 0$, but that we will refer to this information in words when necessary. If G is a connected weighted graph then WRW^G has a unique stationary distribution denoted by positive numbers $\pi^G(v)$ summing to one. This will not in general be uniform, but its existence is enough to guarantee the existence of a stationary Markov chain with the same transition probabilities. We call this stationary Markov chain $SWRW$ the few times the need arises. The new and improved theorem then reads:

Theorem 2.2 *Let G be any finite, connected weighted graph and let x be any vertex of G. Run a weighted random walk WRW_x^G and let $\mathbf{T}$ be the random spanning tree gotten by ignoring the arrows and root of the random directed spanning tree $\tau(WRW_x^G)$. Then $\mathbf{T}$ has the distribution of WST.*

The proof of Theorem 2.1 serves for Theorem 2.2 with a few alterations. These will now be described, thought not much would be lost by taking these details on faith and skipping to the next section.

The groundskeeper's algorithm is unchanged with the provision that the WRW brings with it the information of which edge she should travel if more than one edge connects $WRW(i)$ to $WRW(i+1)$ for some i. The operation to get from the directed tree T_n to a candidate for T_{n-1} is basically the same only instead of there being D choices for how to do this there is one choice for each edge incident to the root v of T_n: choose such an edge, add it to the tree oriented from v to its other endpoint x and remove the edge out of x. It is easy to see again that $\{T_n\}$ is a time-homogeneous Markov chain with transition probability from t to u zero unless u can be gotten from t by the above operation, and if so the probability is proportional to the weight of the edge that was added in the operation. (This is because if $T_n = t$ then $T_{n-1} = u$ if and only if u can be gotten from this operation and WRW travelled along the edge added in this operation between times $n-1$ and n.)

The uniform distribution on vertices is no longer stationary for WRW since we no longer have D-regularity, but the distribution $\pi(v) = d(v)/\sum_x d(x)$ is easily seen to be stationary: start a WRW with $WRW(0)$ having distribution π; then

$$\mathbf{P}(WRW(1) = v) = \sum_x \mathbf{P}(WRW(0) = x \text{ and } WRW(1) = v)$$

$$= \sum_x \frac{d(x)}{\sum_y d(y)} \left(\sum_{e \text{ connecting } x \text{ to } v} w(e)/d(x) \right)$$

20

$$= \frac{1}{\sum_y d(y)} \sum_{e \text{ incident to } v} w(e)$$

$$= \pi(v).$$

The stationary distribution π for the Markov chain $\{T_n\}$ gives a directed tree t rooted at v probability

$$\pi(t) = Kd(v) \prod_{e \in t} w(e),$$

where $K = (\sum_t d(\text{root}(t)) \prod_{e \in t} w(e))^{-1}$ is a normalizing constant. If t, rooted at v, can go to u, rooted at x, by adding an edge e and removing the edge f, then $\pi(u)/\pi(t) = d(x)w(e)/d(v)w(f)$. To verify that π is a stationary distribution for T_n write $\mathcal{C}(u)$ for the class of trees from which it is possible to get to u in one step and for each $t \in \mathcal{C}(u)$ write v_t, e_t and f_t for the root of t, edge added to t to get u and edge taken away from t to get u respectively. If u is rooted at x, then

$$\mathbf{P}(T_{n-1} = u) = \sum_t \mathbf{P}(T_n = t \text{ and } T_{n-1} = u)$$

$$= \sum_{t \in \mathcal{C}(u)} \pi(t)w(e_t)/d(v_t)$$

$$= \sum_{t \in \mathcal{C}(u)} [\pi(u)d(v_t)w(f_t)/d(x)w(e_t)]w(e_t)/d(v_t)$$

$$= \pi(u) \sum_{t \in \mathcal{C}(u)} w(f_t)/d(x)$$

$$= \pi(u),$$

since as t ranges over all trees that can get to u, f_t ranges over all edges incident to x.

Finally, we have again that $\tau(WRW_x^G(0))$ is distributed as $\tau(SWRW^G(0))$ conditioned on having root x, and since the unconditioned π is proportional to $d(x)$ times the weight of the tree (product of the edge weights), the factor of d is constant and

21

$\mathbf{P}(\tau(WRW_x^G(0)) = t)$ is proportional to $\prod_{e \in t} w(e)$ for any t rooted at x. Thus $\tau(WRW_x^G(0))$ is distributed identically to WST. $\square$

2.4   Applying the random walk construction to our model

Although the benefit is not yet clear, we have succeeded in translating the question of determining $\mathbf{P}(e \in \mathbf{T})$ from a question about uniform spanning trees to a question about simple random walks. To see how this works, suppose that e connects the vertices x and y and generate a uniform spanning tree by the random walk construction starting at x: $\mathbf{T}$ = the tree gotten from $\tau(SRW_x^G)$ by ignoring the root and arrows. If $e \in \mathbf{T}$ then its orientation in $\tau(SRW_x^G)$ must be from y to x, and so $e \in \mathbf{T}$ if and only if $SRW(k-1) = x$ where k is the least k for which $SRW(k) = y$. In other words,

$$\mathbf{P}(e \in \mathbf{T}) = \mathbf{P}(\text{first visit of } SRW_x^G \text{ to } y \text{ is along } e). \tag{1}$$

The computation of this random walk probability turns out to be tractable.

   More important is the fact that this may be iterated to get probabilities such as $\mathbf{P}(e, f \in \mathbf{T})$. This requires two more definitions. If G is a finite connected graph and e is an edge of G whose removal does not disconnect G, then the *deletion* of G by e is the graph $G \setminus e$ with the same vertex set and the same edges minus e. If e is any edge that connects distinct vertices x and y, then the *contraction* of G by e is the graph G/e whose vertices are the vertices of G with x and y replaced by a single vertex $x * y$. There is an edge $\rho(f)$ of G/e for every edge of f of G, where if one or both endpoints of f is x or y then that endpoint is replaced by $x * y$ in $\rho(f)$. We write $\rho(z)$ for the vertex corresponding to z in this correspondence, so $\rho(x) = \rho(y) = x * y$ and $\rho(z) = z$ for every $z \neq x, y$. The following example shows G_1 and G_1/e_4. The edge e_4 itself maps to a self-edge under ρ, e_5 becomes parallel to e_6 and D and E map to $D * E$.
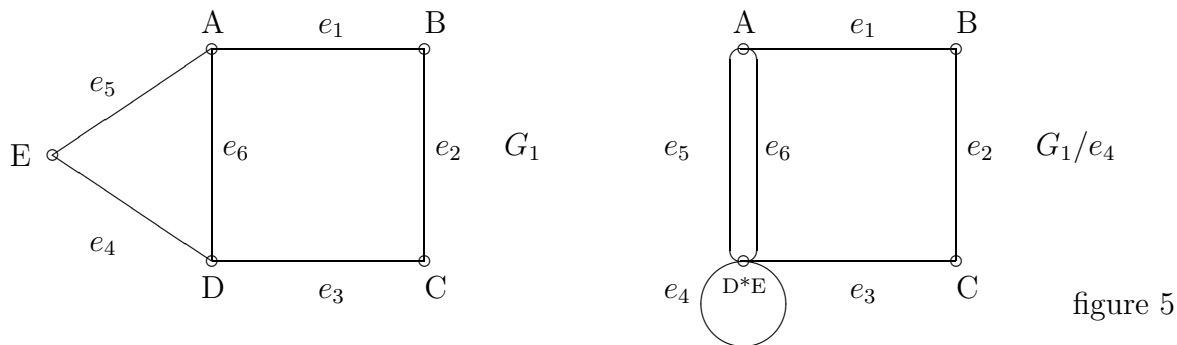
figure 5

It is easy to see that successive deletions and contractions may be performed in any order with the same result. If $e_1, \ldots, e_r$ are edges of G whose joint removal does not disconnect G then the successive deletion of these edges is permissible. Similarly if $\{f_1, \ldots, f_s\}$ is a set of edges of G that contains no cycle, these edges may be successively contracted and the graph $G \setminus e_1, \ldots, e_r / f_1, \ldots, f_s$ is well-defined. It is obvious that the spanning trees of $G \setminus e$ are just those spanning trees of G that do not contain e. Almost as obvious is a one to one correspondence between spanning trees of G containing e and spanning trees of G/e: if t is a spanning tree of G containing e then there is a spanning tree of G/e consisting of $\{\rho(f) : f \neq e \in t\}$.

To translate $\mathbf{P}(e, f \in \mathbf{T})$ to the random walk setting, write this as $\mathbf{P}(e \in \mathbf{T})\mathbf{P}(f \in \mathbf{T} \,|\, e \in \mathbf{T})$. The first term has already been translated. The conditional distribution of a uniform random spanning tree given that it contains e is just uniform among those trees containing e, which is just $\mathbf{P}_{G/e}(\rho(f) \in \mathbf{T})$ where the subscript G/e refers to the fact that $\mathbf{T}$ is now taken to be a uniform random spanning tree of G/e. If f connects z and x then this is in turn equal to $\mathbf{P}(SRW_{\rho(x)}^{G/e}$ first hits $\rho(z)$ along $\rho(f))$. Both the terms have thus been translated; in general it should be clear how this may be iterated to translate the probability of any elementary event, $\mathbf{P}(e_1, \ldots, e_r \in \mathbf{T}$ and $f_1, \ldots, f_s \notin \mathbf{T})$ into a product of random walk probabilities. It remains to be seen how these probabilities may

be calculated.

3 Random walks and electrical networks

Sections 3.1 - 3.3 contain a development of the connection between random walks and electrical networks. The right place to read about this is in [8]; what you will see here is necessarily a bit rushed. Sections 3.5 and 3.6 contain similarly condensed material from other sources.

3.1 Resistor circuits

The electrical networks we discuss will have only two kinds of elements: resistors and voltage sources. Picture the resistors as straight pieces of wire. A resistor network will be built by soldering resistors together at their endpoints. That means that a diagram of a resistor network will just look like a finite graph with each edge bearing a number: the resistance. Associated with every resistor network H is a weighted graph G_H which looks exactly like the graph just mentioned except that the weight of an edge is not the resistance but the *conductance*, which is the reciprocal of the resistance. The distinction between H and G_H is only necessary while we are discussing precise definitions and will then be dropped. A voltage source may be a single battery that provides a specified voltage difference (explained below) across a specified pair of vertices or is may be a more complicated device to hold various voltages fixed at various vertices of the network. Here is an example of a resistor network on a familiar graph, with a one volt battery drawn as a dashed box. Resistances on the edges (made up arbitrarily) are given in ohms.
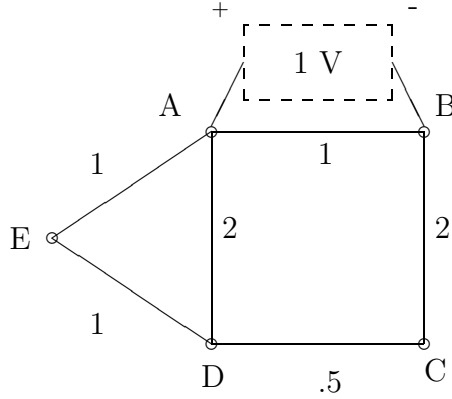
figure 6

The electrical properties of such a network are given by Kirchoff's laws. For the sake of exposition I will give the laws numbers, although these do not correspond to the way Kirchoff actually stated the laws. The first law is that every vertex of the network has a voltage which is a real number. The second law gives every oriented edge (resistor) a current. Each edge has two possible orientations. Say an edge connects x and y. Then the current through the edge is a real number whose sign depends on which orientation you choose for the edge. In other words, the current $I(\vec{xy})$ that flows from x to y is some real number and the current $I(\vec{yx})$ is its negative. (Note though that the weights $w(e)$ are always taken to positive; weights are functions of unoriented edges, whereas currents are functions of oriented edges.) If $I(e)$ denotes the current along an oriented edge $e = \vec{xy}$, $V(x)$ denotes the voltage at x and $R(e)$ denotes the resistance of e, then quantatively, the second law says

$$I(\vec{xy}) = [V(x) - V(y)]R(e)^{-1}. \tag{2}$$

Kirchoff's third law is that the total current flowing into a vertex equals the total current flowing out, or in other words

$$\sum_{y \sim x} I(\vec{xy}) = 0. \tag{3}$$

This may be rewritten using (2). Recalling that in the weighted graph G_H, the weight $w(e)$ is just $R(e)^{-1}$ and that $d(v)$ denotes the sum of $w(e)$ over edges incident to v, we

25

get at every vertex x an equation

$$0 = \sum_{y \sim x}[V(x) - V(y)]w(xy) = V(x)d(x) - \sum_{y \sim x} V(y)w(xy). \tag{4}$$

Since a voltage source may provide current, this may fail to hold at any vertex connected to a voltage source. The above laws are sufficient to specify the voltages of the network – and hence the currents – except that a constant may be added to all the voltages (in other words, it is the voltage differences that are determined, not the absolute voltages). In the above example the voltage difference across AB is required to be one. Setting the voltage at B to zero (since the voltages are determined only up to an additive constant) the reader may check that the voltages at A, C, D and E are respectively $1, 4/7, 5/7$ and $6/7$ and the currents through AB, AE, ED, AD, DC, CB are respectively $1, 1/7, 1/7, 1/7, 2/7, 2/7$.

3.2   Harmonic functions

The voltages in a weighted graph G (which we are now identifying with the resistor network it represents) under application of a voltage source are calculated by finding a solution to Kirchoff's laws on G with specified boundary conditions. For each vertex x there is an unknown voltage $V(x)$. There is also a linear equation for every vertex not connected to a voltage source, and an equation given by the nature of each voltage source. Will these always be enough information so that Kirchoff's laws have a unique solution? The answer is yes and it is most easily seen in the context of harmonic functions. [1]

---

[1]There is also the question of whether any solution exists, but addressing that would take us too far afield. If you aren't convinced of its existence on physical grounds, wait until the next subsection where a probabilistic interpretation for the voltage is given, and then deduce existence of a solution from the fact that these probabilities obey Kirchoff's laws.

If f is a function on the vertices of a weighted graph G, define the *excess* of f at a vertex v, written $\triangle f(v)$ by

$$\triangle f(v) = \sum_{y \sim v}[f(v) - f(y)]w(vy).$$

You can think of $\triangle$ as an operator that maps functions f to other functions $\triangle f$ that is a discrete analog of the Laplacian operator. A function f from the vertices of a finite weighted graph G to the reals is said to be *harmonic* at a vertex v if and only if $\triangle f(v) = 0$. Note that for any function f, the sum of the excesses $\sum_{v \in G}\triangle f(v) = 0$, since each $[f(x) - f(y)]w(xy)$ cancels a $[f(y) - f(x)]w(yx)$ due to $w(xy) = w(yx)$. To see what harmonic functions are intuitively, consider the special case where G is unweighted, i.e. all of the edge weights are one. Then a function is harmonic if and only if its value at a vertex x is the average of the values at the neighbors of x. In the weighted case the same is true, but with a weighted average! Here is an easy but important lemma about harmonic functions.

Lemma 3.1 (Maximum principle) *Let V be a function on the vertices of a finite connected weighted graph, harmonic everywhere except possibly at vertices of some set $X = \{x_1, \ldots, x_k\}$. Then V attains its maximum and minimum on X. If V is harmonic everywhere then it is constant.*

Proof: Let S be the set of vertices where V attains its maximum. Certainly S is nonempty. If $x \in S$ has a neighbor $y \notin S$ then V cannot be harmonic at x since $V(x)$ would then be a weighted average of values less than or equal to $V(x)$ with at least one strictly less. In the case where V is harmonic everywhere, this shows that no vertex in S has a neighbor not in S, hence since the graph is connected every vertex is in S and V is constant. Otherwise, suppose V attains its maximum at some $y \notin X$ and pick a path connecting y to some $x \in X$. The entire path must then be in S up until

and including the first vertex along the path at which V is not harmonic. This is some $x' \in X$. The argument for the minimum is just the same. □

Kirchoff's third law (4) says that the voltage function is harmonic at every x not connected to a voltage source. Suppose we have a voltage source that provides a fixed voltage at some specified set of vertices. Say for concreteness that the vertices are $x_1, \ldots, x_k$ and the voltages produced at these vertices are $c_1, \ldots, c_k$. We now show that Kirchoff's laws determine the voltages everywhere else, i.e. there is at most one solution to them.

Theorem 3.2 *Let V and W be real-valued functions on the vertices of a finite weighted graph G. Suppose that $V(x_i) = W(x_i) = c_i$ for some set of vertices $x_1, \ldots, x_k$ and $1 \leq i \leq k$ and that V and W are harmonic at every vertex other than $x_1, \ldots, x_k$. Then $V = W$.*

Proof: Consider the function $V - W$. It is easy to check that being harmonic at x is a linear property, so $V - W$ is harmonic at every vertex at which both V and W are harmonic. Then by the Maximum Principle, $V - W$ attains its maximum and minimum at some x_i. But $V - W = 0$ at every x_i, so $V - W \equiv 0$. □

Suppose that instead of fixing the voltages at a number of points, the voltage source acts as a current source and supplies a fixed amount of current I_i to vertices $x_i, 1 \leq i \leq k$. This is physically reasonable only if $\sum_{i=1}^{k} I_i = 0$. Then a net current of I_i will have to flow out of each x_i into the network. Using (2) gives

$$I_i = \sum_{y \sim x} w(x, y)(V(x) - V(y)) = \triangle V(x).$$

From this it is apparent that the assumption $\sum_i I_i = 0$ is algebraically as well as physically necessary since the excesses must sum to zero. Kirchoff's laws also determine the voltages (up to an additive constant) of a network with current sources, as we now show.

Theorem 3.3 *Let V and W be real-valued functions on the vertices of a finite weighted graph G. Suppose that V and W both have excess c_i at x_i for some set of vertices x_i and reals c_i, $1 \leq i \leq k$. Suppose also that V and W are harmonic elsewhere. Then $V = W$ up to an additive constant.*

Proof: Excess is linear, so the excess of $V - W$ is the excess of V minus the excess of W. This is zero everywhere, so $V - W$ is harmonic everywhere. By the Maximum Principle, $V - W$ is constant. □

3.3   Harmonic random walk probabilities

Getting back to the problem of random walks, suppose G is a finite connected graph and x, a, b are vertices of G. Let's say that I want to calculate the probability that SRW_x reaches a before b. Call this probability $h_{ab}(x)$. It is not immediately obvious what this probability is, but we can get an equation by watching where the random walk takes its first step. Say the neighbors of x are $y_1, \ldots, y_d$. Then $\mathbf{P}(SRW_x(1) = y_i) = d^{-1}$ for each $i \leq d$. If we condition on $\mathbf{P}(SRW_x(1) = y_i)$ then the probability of the walk reaching a before b is (by the Markov property) the same as if it had started out at y_i. This is just $h_{ab}(y_i)$. Thus

$$
\begin{aligned}
h_{ab}(x) &= \sum_i \mathbf{P}(SRW_x(1) = y_i) h_{ab}(y_i) \\
&= d^{-1} \sum_i h_{ab}(y_i).
\end{aligned}
$$

In other words, h_{ab} is harmonic at x. Be careful though, if x is equal to a or b, it doesn't make sense to look one step ahead since $SRW_x(0)$ already determines whether the walk hit a or b first. In particular, $h_{ab}(a) = 1$ and $h_{ab}(b) = 0$, with h_{ab} being harmonic at every $x \neq a, b$.

Theorem 3.2 tells us that there is only one such function h_{ab}. This same function solves Kirchoff's laws for the unweighted graph G with voltages at a and b fixed at 1 and 0 respectively. In other words, the probability of SRW_x reaching a before b is just the voltage at x when a one volt battery is connected to a and b and the voltage at b is taken to be zero. If G is a weighted graph, we can use a similar argument: it is easy to check that the first-step transition probabilities $p(x, y) = w(\vec{xy})/\sum_z w(\vec{xz})$ show that $h_{ab}(x)$ is harmonic in the sense of weighted graphs. Summarizing this:

Theorem 3.4 *Let G be a finite connected weighted graph. Let a and b be vertices of G. For any vertex x, the probability of SRW_x^G reaching a before b is equal to the voltage at x in G when the voltages at a and b are fixed at one and zero volts respectively.*

Although more generality will not be needed we remark that this same theorem holds when a and b are taken to be sets of vertices. The probability of SRW_x reaching a vertex in a before reaching a vertex in b is harmonic at vertices not in $a \cup b$, is zero on b and one on a. The voltage when vertices in b are held at zero volts and vertices in a are held at one volt also satisfies this, so the voltages and the probabilities must coincide.

Having given an interpretation of voltage in probabilistic terms, the next thing to find is a probabilistic interpretation of the current. The arguments are similar so they will be treated briefly; a more detailed treatment appears in [8]. First we will need to find an electrical analogue for the numbers $u_{ab}(x)$ which are defined probabilistically as the expected number of times a SRW_a hits x before the first time it hits b. This is defined to be zero for $x = b$. For any $x \neq a, b$, let $y_1, \ldots, y_r$ be the neighbors of x. Then the number of visits to x before hitting b is the sum over i of the number of times SRW_a hits y_i before b and goes to x on the next move (the walk had to be somewhere the move before it hit x). By the Markov property, this quantity is $u_{ab}(y_i)p(y_i, x) = u_{ab}(y_i)w(\vec{xy_i})/d(y_i)$. Letting $\phi_{ab}(z)$ denote $u_{ab}(z)/d(z)$ for any z, this yields

$$\phi_{ab}(x) = d(x)u_{ab}(x) = \sum_i u_{ab}(y_i)w(\vec{xy_i})/d(y_i) = \sum_i w(\vec{xy_i})\phi_{ab}(y_i).$$

In other words ϕ_{ab} is harmonic at every $x \neq a, b$. Writing K_{ab} for $\phi_{ab}(a)$ we then have that ϕ_{ab} is K_{ab} at a, zero at b and harmonic elsewhere, hence it is the same function as the the voltage induced by a battery of K_{ab} volts connected to a and b, with the voltage at b taken to be zero. Without yet knowing what K_{ab} is, this determines ϕ_{ab} up to a constant multiple. This in turn determines u_{ab}, since $u_{ab}(x) = d(x)\phi_{ab}(x)$.

Now imagine that we watch SRW_a to see when it crosses over a particular edge $\vec{xy}$ and count plus one every time it crosses from x to y and minus one every time it crosses from y to x. Stop counting as soon as the walk hits b. Let $H_{ab}(\vec{xy})$ denote the expected number of signed crossings. (H now stands for harmonic, not for the name of a resistor network.) We can calculate H in terms of u_{ab} by counting the plusses and the minuses separately. The expected number of plus crossings is just the expected number of times the walk hits x, mulitplied by the probability on each of these occasions that the walk crosses to y on the next move. This is $u_{ab}(x)w(\vec{xy})/d(x)$. Similarly the expected number of minus crossings is $u_{ab}(y)w(\vec{xy})/d(y)$. Thus

$$
\begin{aligned}
H_{ab}(\vec{xy}) &= u_{ab}(x)w(\vec{xy})/d(x) - u_{ab}(y)w(\vec{xy})/d(y) \\
&= w(\vec{xy})[\phi_{ab}(x) - \phi_{ab}(y)].
\end{aligned}
$$

But $\phi_{ab}(x) - \phi_{ab}(y)$ is just the voltage difference across $\vec{xy}$ induced by a K_{ab}-volt battery across a and b. Using (2) and $w(\vec{xy}) = R(\vec{xy})^{-1}$ shows that the expected number of signed crossings of $\vec{xy}$ is just the current induced in $\vec{xy}$ by a K_{ab}-volt battery connected to a and b. A moment's thought shows that the expected number of signed crossings of all edges leading out of a must be one, since the walk is guaranteed to leave a one more time than it returns to a. So the current supplied by the K_{ab}-volt battery must be one amp. Another way of saying this is that

$$
\triangle\phi_{ab} = \delta_a - \delta_b. \tag{5}
$$

Instead of worrying about what K_{ab} is, we may just as well say that the expected number

31

of crossings of $\vec{xy}$ by SRW_a before hitting b is the current induced when one amp is supplied to a and drawn out at b.

3.4 Electricity applied to random walks applied to spanning trees

Finally we can address the random walk question that relates to spanning trees. In particular, the claim that the probability in equation (1) is tractable will be borne out several different ways. First we will see how the probability may be "calculated" by an analog computing device, namely a resistor network. In the next subsection, the computation will be carried out algebraically and very neatly, but only for particularly nice symmetric graphs. At the end of the section, a universal method will be given for the computation which is a little messier. Finally in Section 4 the question of the individual probabilities in (1) will be avoided altogether and we will see instead how values for these probabilities (wherever they might come from) determine the probabilities for all contractions and deletions of the graph and therefore determine all the joint probabilities $\mathbf{P}(e_1, \ldots, e_k \in \mathbf{T})$ and hence the entire measure.

Let $e = \vec{xy}$ be any edge of a finite connected weighted graph G. Run SRW_x^G until it hits y. At this point either the walk just moved along e from x to y – necessarily for the first time – and e will be in the tree $\mathbf{T}$ given by $\tau(SRW_x^G)$, or else the walk arrived at y via a different edge in which case the walk never crossed e at all and $e \notin \mathbf{T}$. In either case the walk never crossed from y to x since it stops if it hits y. Then the expected number of signed crossings of $e = \vec{xy}$ by SRW_x up to the first time it hits y is equal to the probability of first reaching y along e which equals $\mathbf{P}(e \in \mathbf{T})$. Putting this together with the electrical interpretation of signed crossings give

Theorem 3.5 $\mathbf{P}(e \in \mathbf{T}) =$ *the fraction of the current that goes through edge e when a*

battery is hooked up to the two endpoints of e.

$\square$

This characterization leads to a proof of Theorem 1.1 provided we are willing to accept a proposition that is physically obvious but not so easy to prove, namely

Theorem 3.6 (Rayleigh's monotonicity law) *The effective resistance of a circuit cannot increase when a new resistor is added.*

The reason this is physically obvious is that adding a new resistor provides a new path for current to take while allowing the current still to flow through all the old paths. Theorem 1.1 says that the conditional probability of $e \in \mathbf{T}$ given $f \in \mathbf{T}$ must be less than or equal to the unconditional probability. Using Theorem 3.5 and the fact that the probabilities conditioned on $f \notin \mathbf{T}$ are just the probabilities for WST on $G \setminus f$, this boils down to showing that the fraction of current flowing directly across e is no greater on G than it is on $G \setminus f$. The battery across e meets two parallel resistances: e and the effective resistance of the rest of G. The fraction of current flowing through e is inversely proportional to the ratio of these two resistances. Rayleigh's theorem says that the effective resistance of the rest of G including f is at most the effective resistance of $G \setminus f$, so the fraction flowing through e on G is at most the fraction flowing through e on $G \setminus f$. In Section 4, a proof will be given that does not rely on Rayleigh.

3.5   Algebraic calculations for the square lattice

If G is a finite graph, then the functions from the vertices of G to the reals form a finite-dimensional real vector space. The operator $\triangle$ that maps a function V to its excess is a linear operator on this vector space. In this language, the voltages in a resistor network

33

with one unit of current supplied to a and drawn out at b are the unique (up to additive constant) function V that solves $\triangle V = \delta_a - \delta_b$. Here δ_x is the function that is one at x and zero elsewhere. This means that V can be calculated simply by inverting $\triangle$ in the basis $\{\delta_x; x \in G\}$. Although $\triangle$ is technically not invertible, its nullspace has dimension one so it can be inverted on a set of codimension one. A classical determination of V for arbitrary graphs is carried out in the next subsection. The point of this subsection is to show how the inverse can be obtained in a simpler way for nice graphs.

The most general "nice" graphs to which the method will apply are the infinite $\mathbb{Z}^d$-periodic lattices. Since in this article I am restricting attention to finite graphs, I will not attempt to be general but will instead show a single example. The reader may look in [6] for further generality. The example considered here is the square lattice. This is just the graph you see on a piece of graph paper, with vertices at each pair of integer coordinates and four edges connecting each point to its nearest neighbors. The exposition will be easiest if we consider a finite square piece of this and impose wrap-around boundary conditions. Formally, let T_n (T for torus) be the graph whose vertices are pairs of integers $\{(i,j) : 0 \leq i, j \leq n-1\}$ and for which two points are connected if and only if they agree in one component and differ by one mod n in the other component. Here is a picture of this with $n = 3$ and the broken edges denoting edges that wrap around to the other side of the graph. The graph is unweighted (all edge weights are one.)
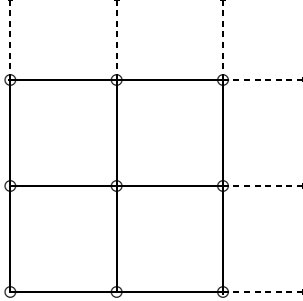
Let $\zeta = e^{2\pi i/n}$ denote the first n^{th} root of unity. To invert $\triangle$ we exhibit its eigenvectors. Since the vector space is a space of functions, the eigenvectors are called eigenfunctions. For each pair of integers $0 \le k, l \le n - 1$ let f_{kl} be the function on the vertices of T_n defined by

$$f_{kl}(i, j) = \zeta^{ki+lj}.$$

If you have studied group representations, you will recognize f_{kl} as the representations of the group $T_n = (\mathbb{Z}/n\mathbb{Z})^2$ and in fact the rest of this section may be restated more compactly in terms of characters of this abelian group.

It is easy to calculate

$$
\begin{aligned}
\triangle f_{kl}(i, j) &= 4\zeta^{ki+lj} - \zeta^{ki+l(j+1)} - \zeta^{ki+l(j-1)} - \zeta^{k(i+1)+lj} - \zeta^{k(i-1)+lj} \\[2mm]
&= \zeta^{ki+lj}(4 - \zeta^k - \zeta^{-k} - \zeta^l - \zeta^{-l}) \\[2mm]
&= \zeta^{ki+lj}(4 - 2\cos(2\pi k/n) - 2\cos(2\pi l/n)).
\end{aligned}
$$

Since the multiplicative factor $(4 - 2\cos(2\pi k/n) - 2\cos(2\pi l/n))$ does not depend on i or j, this shows that f_{kl} is indeed an eigenfunction for $\triangle$ with eigenvalue $\lambda_{kl} = 4 - 2\cos(2\pi k/n) - 2\cos(2\pi l/n)$.

35

Now if $\{v_k\}$ are eigenvectors for some linear operator A with eigenvalues $\{\lambda_k\}$, then for any constants $\{c_k\}$,

$$A^{-1}(\sum_k c_k v_k) = \sum_k \lambda_k^{-1} c_k v_k. \tag{6}$$

If some λ_k is equal to zero, then the range of A does not include vectors w with $c_k \neq 0$, so $A^{-1}w$ does not exist for such w and indeed the formula blows up due to the λ_k^{-1}. In our case $\lambda_{kl} = 4 - 2\cos(2\pi k/n) - 2\cos(2\pi l/n) = 0$ only when $k = l = 0$. Thus to calculate $\triangle^{-1}(\delta_a - \delta_b)$ we need to figure out coefficents c_{kl} for which $\delta_a - \delta_b = \sum_{kl} c_{kl} f_{kl}$ and verify that $c_{00} = 0$. For this puropose, it is fortunate that the eigenfunctions $\{f_{kl}\}$ are actually a unitary basis in the inner product $< f, g >= \sum_{ij} f(i,j)\overline{g(i,j)}$. You can check this by writing

$$< f_{kl}, f_{k'l'} >= \sum_{ij} \zeta^{ki+lj}\overline{\zeta^{k'i+l'j}};$$

elementary algebra show this to be one if $k = k'$ and $l = l'$ and zero otherwise, which what it means to be unitary. Unitary bases are great for calculation because the coefficients $\{c_{kl}\}$ of any V in a unitary eigenbasis $\{f_{kl}\}$ are given by $c_{kl} =< V, f_{kl} >$. In our case, this means $c_{kl} = \sum_{ij} V(i,j)\overline{f_{kl}(i,j)}$. Letting a be the vertex $(0,0)$, b be the vertex $(1,0)$ and $V = \delta_a - \delta_b$, this gives $c_{kl} = 1 - \overline{\zeta}^k$ and hence

$$\delta_a - \delta_b = \sum_{k,l}(1 - \overline{\zeta}^k)f_{kl}.$$

We can now plug this into equation (6), since clearly $c_{00} = 0$. This gives

$$\triangle V \;\; = \;\; \delta_a - \delta_b$$

$$\Leftrightarrow \;\; V(i,j) \;\; = \;\; cf_{00}(i,j) + \sum_{(k,l)\neq(0,0)}(1 - \zeta^k)\lambda_{kl}^{-1}f_{kl}(i,j)$$

$$= \;\; c + \sum_{(k,l)\neq(0,0)} \frac{1 - \zeta^k}{4 - 2\cos(2\pi k/n) - 2\cos(2\pi l/n)}\zeta^{ki+lj}. \tag{7}$$

36

This sum is easy to compute exactly and to approximate efficiently when n is large. In particular as $n \to \infty$ the sum may be replaced by an integral which by a small miracle admits an exact computation. Details of this may be found in [16, page 148]. You may check your arithmetic against mine by using (7) to derive the voltages for a one volt battery placed across the bottom left edge e of T_3 and across the bottom left edge e' of T_4:

5/8	3/8	1/2		56/90	34/90	40/90	50/90
5/8	3/8	1/2		50/90	40/90	42/90	48/90
1	0	1/2		56/90	34/90	40/90	50/90
				1	0	34/90	56/90

Section 5 shows how to put these numbers to good use, but we can already make one calculation based on Theorem 3.5. The four currents flowing out of the bottom left vertex under the voltages shown are given by the voltage differences: $1, 3/8, 1/2$ and $3/8$. The fraction of the current flowing directly through the bottom left edge e is $8/18$, and according to Theorem 3.5, this is $\mathbf{P}(e \in \mathbf{T})$. An easy way to see this is right is by the symmetry of the graph T_3. Each of the 18 edges should be equally likely to be in $\mathbf{T}$, and since every spanning tree has 8 edges, the probability of any given edge being in the tree must be $8/18$.

3.6   Electrical networks and spanning trees

The order in which topics have been presented so far makes sense from an expository viewpoint but is historically backwards. The first interest in enumerating spanning trees came from problems in electrical network theory. To set the record straight and also to close the circle of ideas

$$\text{spanning trees} \to \text{random walks} \to \text{electrical networks} \to \text{spanning trees}$$

I will spend a couple of paragraphs on this remaining connection.

Let G be a finite weighted graph. Assume there are no voltage sources and the quantity of interest is the *effective resistance* between two vertices a and b. This is defined to be the voltage it is necessary to place across a and b to induce a unit current flow. A classical theorem known to Kirchoff is:

Theorem 3.7 *Say s is an a, b-spanning bitree if s is a spanning forest with two components, one containing a and the other containing b. The effective resistance between a and b may be computed from the weighted graph G by taking the quotient N/D where*

$$D = \sum_{\text{spanning trees } t} \left(\prod_{e \in t} w(e) \right)$$

is the sum of the weights of all spanning trees of G and

$$N = \sum_{a, b\text{-spanning bitrees } s} \left(\prod_{e \in s} w(e) \right)$$

is the analogous sum over a, b-spanning bitrees. $\square$

To see that how this is implied by Theorem 3.5 and equation (1), imagine adding an extra one ohm resistor from a to b. The probability of this edge being chosen in a WST on the new graph is by definition given by summing the weights of trees containing the new edge and dividing by the total sum of the weights of all spanning trees. Clearly D is the sum of the weights of trees not containing the extra edge. But the trees containing the extra edge are in one-to-one correspondence with a, b-spanning bitrees (the correspondence being to remove the extra edge). The extra edge has weight one, so the sum of the weights of trees that do contain the extra edge is N and the probability of a WST containing the extra edge is $N/(N + D)$. By equation (1) and Theorem 3.5, this must then be the fraction of current flowing directly through the extra edge when a battery is placed across a and b. Thinking of the new circuit as consisting of the extra

38

edge in parallel with G, the fractions of the current passing through the two components are proportional to the inverses of their resistances, so the ratio of the resistance of the extra edge to the rest of the circuit must be $D : N$. Since the extra edge has resistance one, the effective resistance of the rest of the circuit is N/D.

The next problem of course was to efficiently evaluate the sum of the weights of all spanning trees of a weighted graph. The solution to this problem is almost as well known and can be found, among other places in [7].

Theorem 3.8 (Matrix-Tree Theorem) *Let G be a finite, simple, connected, weighted graph and define a matrix indexed by the vertices of G by letting $M(x,x) = d(x)$, $M(x,y) = -w(\vec{xy})$ if x and y are connected by an edge, and $M(x,y) = 0$ otherwise. Then for any vertex x, the sum of the weights of all spanning trees of G is equal to the determinant of the matrix gotten from M by deleting by the row and column corresponding to x.*

The matrix M is nothing but a representation of $\triangle$ with respect to the basis $\{\delta_x\}$. Recalling that the problem essentially boils down to inverting $\triangle$, the only other ingredient in this theorem is the trick of inverting the action of a singular matrix on an element on its range by inverting the largest invertible principal minor of the matrix. Details can be found in [7]. □

4   Transfer-impedances

In the last section we saw how to calculate $\mathbf{P}(e \in \mathbf{T})$ in several ways: by Theorems 3.5 or 3.7 in general and by equations such as (7) in particularly symmetric cases. By repeating the calculations in Theorem 3.5 and 3.7 for contractions and deletions of a graph (see Section 2.4), we could then find enough conditional probabilities to determine

the probability of any elementary event $\mathbf{P}(e_1, \ldots, e_r \in \mathbf{T}$ and $f_1, \ldots, f_s \notin \mathbf{T})$. Not only is this inefficient, but it fails to apply to the symmetric case of equation (7) since contracting or deleting the graph breaks the symmetry. The task at hand is to alleviate this problem by showing how the data we already know how to get – current flows on G – determine the current flows on contractions and deletions of G and thereby determine all the elementary probabilities for WST on G. This will culminate in a proof of Theorem 1.2, which encapsulates all of the necessary computation into a single determinant.

4.1   An electrical argument

To keep notation to a minimum this subsection will only deal with unweighted, D-regular graphs. Begin by stating explicitly the data that will be used to determine all other probabilities. For oriented edges $e = \vec{xy}$ and $f = \vec{zw}$ in a finite connected graph G, define the *transfer-impedance* $H(e, f) = \phi_{xy}(z) - \phi_{xy}(w)$ which is equal to the voltage difference across f, $V(z) - V(w)$, when one amp of current is supplied to x and drawn out at y. We will assume knowledge of $H(e, f)$ for every pair of edges in G (presumably via some analog calculation, or in a symmetric case by equation (7) or something similar) and show how to derive all other probabilities from these transfer-impedances.

Note first that $H(e, e)$ is the voltage across e for a unit current flow supplied to one end of e and drawn out of the other. This is equal to the current flowing directly along e under a unit current flow and is thus $\mathbf{P}(e \in \mathbf{T})$. The next step is to try a computation involving a single contraction. For notation, recall the map ρ which projects vertices and edges of G to vertices and edges of G/f. Fix edges $e = \vec{xy}$ and $f = \vec{zw}$ and let $\{V(v) : v \in G/f\}$ be the voltages we need to solve for: voltages at vertices of G/f when a unit current is supplied to $\rho(x)$ and drawn out at $\rho(y)$. As we have seen, this means $\triangle V(v) = +1, -1$ or 0 according to whether $v = x, y$ or neither. Suppose we lift this to a function $\overline{V}$ on the vertices of G by letting $\overline{V}(x) = V(\rho(x))$. Let's calculate the excess

$\triangle \overline{V}$ of $\overline{V}$. Each edge of G corresponds to an edge in G/f, so for any $v \neq z, w$ in G, $\triangle \overline{V}(v) = \triangle V(\rho(v))$; this is equal to $+1$ if $v = x$, -1 if $v = y$ and zero otherwise. Since ρ maps both z and w onto the same vertex $v * w$, we can't tell what the $\triangle \overline{V}$ is at z or w individually, but $\triangle \overline{V}(z) + \triangle \overline{V}(w)$ will equal $\triangle V(z * w)$ which will equal $+1$ if z or w coincides with x, -1 if z or w coincides with y and zero otherwise (or if both coincide!). The last piece of information we have is that $\overline{V}(z) = \overline{V}(w)$. Summarizing,

(i) $\triangle \overline{V} = \delta_x - \delta_y + c(\delta_z - \delta_w)$;

(ii) $\overline{V}(z) = \overline{V}(w)$,

where c is some unknown constant. To see that this uniquely defines $\overline{V}$ up to an additive constant, note that the difference between any two such functions has excess $c(\delta_z - \delta_w)$ for some c, hence by the maximum principle reaches its maximum and minimum on $\{z, w\}$; on the other hand the values at z and w are equal, so the difference is constant.

Now it is easy to find $\overline{V}$. Recall from equation (5) that ϕ satisfies $\triangle \phi_{ab} = \delta_a - \delta_b$. The function $\overline{V}$ we are looking for is then $\phi_{xy} + c\phi_{zw}$ where c is chosen so that

$$\phi_{xy}(z) + c\phi_{zw}(z) = \phi_{xy}(w) + c\phi_{zw}(w).$$

In words, $\overline{V}$ gives the voltages for a battery supplying unit current in at x and out at y plus another battery across z and w just strong enough to equalize the voltages at z and w. How strong is that? The battery supplying unit current to x and y induces by definition a voltage $H(\vec{xy}, \vec{zw})$ across z and w. To counteract that, we need a $-H(\vec{xy}, \vec{zw})$-volt battery across z and w. Since supplying one unit of current in at z and out at w produces a voltage across z and w of $H(\vec{zw}, \vec{zw})$, the current supplied by the counterbattery must be $c = -H(\vec{xy}, \vec{zw})/H(\vec{zw}, \vec{zw})$. We do not need to worry about $H(\vec{zw}, \vec{zw})$ being zero since this means that $\mathbf{P}(f \in \mathbf{T}) = 0$ so we shouldn't be conditioning on $f \in \mathbf{T}$. Going

back to the original problem,

$$
\begin{aligned}
\mathbf{P}(e \in \mathbf{T} \,|\, f \in \mathbf{T}) &= V(\rho(x)) - V(\rho(y)) \\[2mm]
&= \overline{V}(x) - \overline{V}(y) \\[2mm]
&= H(\vec{xy}, \vec{xy}) + H(\vec{zw}, \vec{xy}) \frac{-H(\vec{xy}, \vec{zw})}{H(\vec{zw}, \vec{zw})} \\[2mm]
&= \frac{H(\vec{xy}, \vec{xy}) H(\vec{zw}, \vec{zw}) - H(\vec{xy}, \vec{zw}) H(\vec{zw}, \vec{xy})}{H(\vec{zw}, \vec{zw})}.
\end{aligned}
$$

Multiplying this conditional probability by the unconditional probability $\mathbf{P}(f \in \mathbf{T})$ gives the probability of both e and f being in $\mathbf{T}$ which may be written as

$$
\mathbf{P}(e, f \in \mathbf{T}) = \left| \begin{array}{cc} H(\vec{xy}, \vec{xy}) & H(\vec{xy}, \vec{zw}) \\ H(\vec{zw}, \vec{xy}) & H(\vec{zw}, \vec{zw}) \end{array} \right|.
$$

Thus $\mathbf{P}(e, f \in \mathbf{T}) = \det M(e, f)$ where M is the matrix of values of H as in Theorem 1.2.

Theorem 1.2 has in fact now been proved for $r = 1, 2$. The procedure for general r will be similar. Write $\mathbf{P}(e_1, \ldots e_r \in \mathbf{T})$ as a product of conditional probabilities $\mathbf{P}(e_i \in \mathbf{T} \,|\, e_{i+1}, \ldots, e_r \in \mathbf{T})$. Then evaluate this conditional probability by solving for voltages on $G/e_{i+1} \cdots e_r$. This is done by placing batteries across $e_1, \ldots, e_r$ so as to equalize voltages across all $e_{i+1}, \ldots, e_r$ simultaneously. Although in the $r = 2$ case it was not necessary to worry about dividing by zero, this problem does come up in the general case which causes an extra step in the proof. We will now summarily generalize the above discussion on how to solve for voltages on contractions of a graph and then forget about electricity altogether.

Lemma 4.1 *Let G be a finite D-regular connected graph and let $f_1, \ldots, f_r$ and $e = \vec{xy}$ be edges of G that form no cycle. Let ρ be the map from G to $G/f_1 \ldots f_r$ that maps edges to corresponding edges and maps vertices of G to their equivalence classes under the relation*

of being connected by edges in $\{f_1, \ldots, f_r\}$. Let $\overline{V}$ be a function on the vertices of G such that

(i) *If $z\vec{w} = f_i$ for some i then $\overline{V}(z) = \overline{V}(w)$;*

(ii) *$\sum_{z \in \rho^{-1}(v)} \triangle\overline{V}(z) = +1$ if $\rho(x) = v$, -1 if $\rho(y) = v$ and zero otherwise.*

If $\mathbf{T}$ is a uniform spanning tree for G then $\mathbf{P}(e \in \mathbf{T} \mid f_1, \ldots, f_r \in \mathbf{T}) = \overline{V}(x) - \overline{V}(y)$.

Proof: As before, we know that $\mathbf{P}(e \in \mathbf{T} \mid f_1, \ldots, f_r \in \mathbf{T})$ is given by $V(\rho(x)) - V(\rho(y))$ where V is the voltage function on $G/f_1 \cdots f_r$ for a unit current supplied in at x and out at y. Defining $\overline{V}(v)$ to be $V(\rho(v))$, the lemma will be proved if we can show that $\overline{V}$ is the unique function on the vertices of G satisfying (i) and (ii). Seeing that $\overline{V}$ satisfies (i) and (ii) is the same as before. Since ρ provides a one to one correspondence between edges of G and edges of $G/f_1, \ldots, f_r$, the excess of $\overline{V}$ at vertices of $\rho^{-1}(v)$ is the sum over edges leading out of vertices in $\rho^{-1}(v)$ of the difference of $\overline{V}$ across that edge, which is the sum over edges leading out of $\rho(v)$ of the difference of V across that edge; this is the excess of V at $\rho(v)$ which is $= 1, -1$ or 0 according to whether x or y or neither is in $\rho^{-1}(v)$.

Uniqueness is also easy. If $\overline{W}$ is any function satisfying (i), define a function W on the vertices of $G/f_1 \cdots f_r$ by $W(\rho(v)) = \overline{W}(v)$. If $\overline{W}$ satisfies (ii) as well then it is easy to check that W satisfies $\triangle W = \delta_{\rho(x)} - \delta_{\rho(y)}$ so that $W = V$ and $\overline{W} = \overline{V}$. $\qquad\square$

4.2   Proof of the transfer-impedance theorem

First of all, though is is true that the function H in the previous subsection and the statement of the theorem is symmetric, I'm not going to include a proof – nothing else

we talk about relies on symmetry of H and a proof may be found in any standard treatment of the Green's function, such as [16]. Secondly, it is easiest to reduce the problem to the case of D-regular graphs immediately so as to be able to use the previous lemma. Suppose G is any finite connected graph. Let D be the maximum degree of any vertex in G and to any vertex of lesser degree k, add $D - k$ self-edges. The resulting graph is D-regular (though not simple) and furthermore it has the same spanning trees as G. To prove Theorem 1.2 for finite connected graphs, it therefore suffices to prove the theorem for finite, connected, D-regular graphs. Restating what is to be proved:

Theorem 4.2 *Let G be any finite, connected, D-regular graph and let $\mathbf{T}$ be a uniform random spanning tree of G. Let $H(\vec{xy}, \vec{zw})$ be the voltage induced across $\vec{zw}$ when one amp is supplied from x to y. Then for any $e_1, \ldots, e_r \in G$,*

$$\mathbf{P}(e_1, \ldots, e_r \in \mathbf{T}) = \det M(e_1, \ldots, e_r)$$

where $M(e_1, \ldots, e_r)$ is the r by r matrix whose i, j-entry is $H(e_i, e_j)$.

The proof is by induction on r. We have already proved it for $r = 1, 2$, so now we assume it for $r - 1$ and try to prove it for r. There are two cases. The first possibility is that $\mathbf{P}(e_1, \ldots, e_{r-1} \in \mathbf{T}) = 0$. This means that no spanning tree of G contains $e_1, \ldots, e_r$ which means that these edges contain some cycle. Say the cycle is $e_{n(0)}, \ldots, e_{n(k-1)}$ where there are vertices $v(i)$ for which $e_{n(i)}$ connects $v(i)$ to $v(i + 1 \mod k)$. For any vertices x, y, ϕ_{xy} is the unique solution up to an additive constant of $\triangle \phi_{xy} = \delta_x - \delta_y$. Thus $\triangle \left(\sum_{i=0}^{k-1} \phi_{v(i)\,v(i+1 \bmod k)} \right) = 0$ which means that $\sum_{i=0}^{k-1} \phi_{v(i)\,v(i+1 \bmod k)}$ is constant. Then for any $\vec{xy}$,

$$\sum_{i=0}^{k-1} H(e_{n(i)}, \vec{xy})$$

$$= \sum_{i=0}^{k-1} \phi_{v(i)\,v(i+1 \bmod k)}(x) - \sum_{i=0}^{k-1} \phi_{v(i)\,v(i+1 \bmod k)}(y)$$

$$= 0.$$

This says that in the matrix $M(e_1, \ldots, e_r)$, the rows $n(1), \ldots, n(k)$ are linearly dependent, summing to zero. Then $\det M(e_1, \ldots, e_r) = 0$ which is certainly the probability of $e_1, \ldots, e_r \in \mathbf{T}$.

The second possibility is that $\mathbf{P}(e_1, \ldots, e_{r-1} \in \mathbf{T}) \neq 0$. We can then write

$$\mathbf{P}(e_1, \ldots, e_r \in \mathbf{T})$$

$$= \mathbf{P}(e_1, \ldots e_{r-1} \in \mathbf{T})\mathbf{P}(e_r \in \mathbf{T} \mid e_1, \ldots e_{r-1} \in \mathbf{T})$$

$$= \det M(e_1, \ldots, e_{r-1})\mathbf{P}(e_r \in \mathbf{T} \mid e_1, \ldots e_{r-1} \in \mathbf{T})$$

by the induction hypothesis. To evaluate the last term we look for a function $\overline{V}$ satisfying the conditions of Lemma 4.1 with e_r instead of e and $e_1, \ldots, e_{r-1}$ instead of $f_1, \ldots, f_r$. For $i \leq r-1$, let x_i and y_i denote the vertices connected by e_i. For any $v \in G/e_1 \cdots e_{r-1}$ and any $i \leq r-1$, $\sum_{z \in \rho^{-1}(v)} \triangle\phi_{x_i y_i}(z) = \sum_{z \in \rho^{-1}(v)} \delta_{x_i}(z) - \delta_{y_i}(z)$ which is zero since the class $\rho^{-1}(v)$ contains both x_i and y_i or else contains neither. The excess of $\phi_{x_r y_r}$ summed over $\rho^{-1}(v)$ is just 1 if $\rho(x_r) = v$, -1 if $\rho(y_r) = v$ and zero otherwise. By linearity of excess, this implies that the sum of $\phi_{x_r y_r}$ with any linear combination of $\{\phi_{x_i y_i} : i \leq r-1\}$ satisfies (ii) of the lemma.

Satisfying part (i) is then a matter of choosing the right linear combination, but the lovely thing is that we don't have to actually compute it! We do need to know it exists and here's the argument for that. The ith row of $M(e_1, \ldots, e_r)$ lists the values of $\phi_{x_i y_i}(x_j) - \phi_{x_i y_i}(y_j)$ as j runs from 1 to r. Looking for $c_1, \ldots, c_{r-1}$ such that $\phi_{x_r y_r} + \sum_{i=1}^{r-1} \phi_{x_i y_i}$ is the same on x_j as on y_j for $j \leq r-1$ is the same as looking for c_i for which the rth row of M plus the sum of C_i times the ith row of M has zeros for every entry except the rth. In other words we want to row-reduce, using the first $r-1$ rows to clear $r-1$ zeros in the last row. There is a unique way to do this precisely when the

determinant of the upper $r-1$ by $r-1$ submatrix is nonzero, which is what we have assumed. So these $c_1, \ldots, c_{r-1}$ exist and $\overline{V}(v) = \phi_{x_r y_r}(v) + \sum_{i=1}^{r-1} \phi_{x_i y_i}(v)$.

The lemma tells us that $\mathbf{P}(e_r \in \mathbf{T} \,|\, e_1, \ldots, e_{r-1} \in \mathbf{T})$ is $\overline{V}(x_r) - \overline{V}(y_r)$. This is just the r, r-entry of the row-reduced matrix. Now calculate the determinant of the row-reduced matrix in two ways. Firstly, since row-reduction does not change the determinant of a matrix, the determinant must still be $\det M(e_1, \ldots, e_r)$. On the other hand, since the last row is all zeros except the last entry, expanding along the last row gives that the determinant is the r, r-entry times the determinant of the upper $r-1$ by $r-1$ submatrix, which is just $\mathbf{P}(e_r \in \mathbf{T} \,|\, e_1, \ldots, e_{r-1} \in \mathbf{T}) \det M(e_1, \ldots, e_{r-1})$. Setting these two equal gives

$$\mathbf{P}(e_r \in \mathbf{T} \,|\, e_1, \ldots, e_{r-1} \in \mathbf{T}) = \det M(e_1, \ldots, e_r)/\det M(e_1, \ldots, e_{r-1}).$$

The induction hypothesis says that

$$\mathbf{P}(e_1, \ldots, e_{r-1} \in \mathbf{T}) = \det M(e_1, \ldots e_{r-1})$$

and multiplying the conditional and unconditional probabilities proves the theorem. $\square$

4.3   A few computational examples

It's time to take a break from theorem-proving to see how well the machinery we've built actually works. A good place to test it is the graph T_3, since the calculations have essentially been done, and since even T_3 is large enough to prohibit enumeration of the spanning trees directly by hand (you can use the Matrix-Tree Theorem with all weights one to check that there are 11664 of them). Say we want to know the probability that the middle vertex A is connected to B, C and D in a uniform random spanning tree $\mathbf{T}$ of T_3.
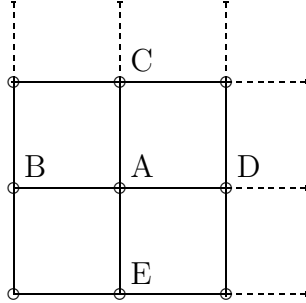
figure 8

We need then to calculate the transfer-impedance matrix for the edges AB, AC and AD. Let's say we orient them all toward A. The symmetry of T_3 under translation and 90° rotation allows us to rely completely on the voltages calculated at the end of 3.5. Sliding the picture upwards one square and multiplying the given voltages by 4/9 to produce a unit current flow from B to A gives voltages

$$
\begin{array}{ccc}
5/18 & 3/18 & 4/18 \\
8/18 & 0 & 4/18 \\
5/18 & 3/18 & 4/18
\end{array}
$$

which gives transfer-impedances $H(BA, BA) = 8/18$, $H(BA, CA) = 3/18$ and $H(BA, DA) = 4/18$. The rest of the values follow by symmetry, giving

$$
M(BA, CA, DA) = \frac{1}{18} \begin{pmatrix} 8 & 3 & 4 \\ 3 & 8 & 3 \\ 4 & 3 & 8 \end{pmatrix}.
$$

Applying Theorem 4.2 gives $\mathbf{P}(BA, CA, DA \in \mathbf{T}) = \det M(BA, CA, DA) = \dfrac{312}{5832}$, or in other words just 624 of the 11664 spanning trees of T_3 contain all these edges. Compare

47

this to using the Matrix-Tree Theorem to calculate the same probability. That does not require the preliminary calculation of the voltages, but it does require an eight by eight determinant.

Suppose we want now to calculate the probability that A is a *leaf* of $\mathbf{T}$, that is to say there is only one edge in $\mathbf{T}$ incident to A. By symmetry this edge will be AB 1/4 of the time, so we need to calculate $\mathbf{P}(BA \in \mathbf{T}$ and $CA, DA, EA \notin \mathbf{T})$ and then multiply by four. As remarked earlier, we can use inclusion-exclusion to get the answer. This would entail writing

$$\mathbf{P}(BA \in \mathbf{T} \text{ and } CA, DA, EA \notin \mathbf{T})$$

$$\begin{aligned}
= \ & \mathbf{P}(BA \in \mathbf{T}) - \mathbf{P}(BA, CA \in \mathbf{T}) - \mathbf{P}(BA, DA \in \mathbf{T}) - \mathbf{P}(BA, EA \in \mathbf{T}) \\
& + \mathbf{P}(BA, CA, DA \in \mathbf{T}) + \mathbf{P}(BA, CA, EA \in \mathbf{T}) + \mathbf{P}(BA, DA, EA \in \mathbf{T}) \\
& - \mathbf{P}(BA, CA, DA, EA \in \mathbf{T}).
\end{aligned}$$

This is barely manageable for four edges, and gets exponentially messier as we want to know about probabilities involving more edges. Here is an easy but useful theorem telling how to calculate the probability of a general *cylinder* event, namely the event that $e_1, \ldots, e_r$ are in the tree, while $f_1, \ldots, f_s$ are not in the tree.

Theorem 4.3 *Let $M(e_1, \ldots, e_k)$ be an k by k transfer-impedance matrix. Let $M^{(r)}$ be the matrix for which $M^{(r)}(i,j) = M(i,j)$ if $i \leq r$ and $M^{(r)}(i,j) = 1 - M(i,j)$ if $r+1 \leq i \leq k$. Then $\mathbf{P}(e_1, \ldots, e_r \in \mathbf{T}$ and $e_{r+1}, \ldots, e_k \notin \mathbf{T}) = \det M^{(r)}$.*

Proof: The proof is by induction on $k - r$. The initial step is when $r = k$; then $M^{(r)} = M$ so the theorem reduces to Theorem 4.2. Now suppose the theorem to be true for $k - r = s$ and let $k - r = s + 1$. Write

$$\mathbf{P}(e_1, \ldots, e_r \in \mathbf{T} \text{ and } e_{r+1}, \ldots, e_k \notin \mathbf{T})$$

$$= \mathbf{P}(e_1, \ldots, e_r \in \mathbf{T} \text{ and } e_{r+2}, \ldots, e_k \notin \mathbf{T})$$
$$-\mathbf{P}(e_1, \ldots, e_{r+1} \in \mathbf{T} \text{ and } e_{r+2}, \ldots, e_k \notin \mathbf{T})$$

$$= \det M(e_1, \ldots, e_r, e_{r+2}, \ldots e_k) - \det M(e_1, \ldots, e_{r+1}, e_{r+2}, \ldots e_k),$$

since the induction hypothesis applies to both of the last two probabilities. Call these last two matrices M_1 and M_2. The trick now is to stick an extra row and column into M_1: let M' be $M(e_1, \ldots, e+k)$ with the $r+1^{st}$ row replaced by zeros except for a one in the $r+1^{st}$ position. Then M' is M_1 with an extra row and column inserted. Expanding along the extra row gives $\det M' = \det M_1$. But M' and M_2 differ only in the $r+1^{st}$ row, so by multilinearity of the determinant,

$$\det M_1 - \det M_2 = \det M' - det M_2 = \det M''$$

where M'' agrees with M' and M_2 except that the $r+1^{st}$ row is the difference of the $r+1^{st}$ rows of M' and M_2. The induction is done as soon as you realize that M'' is just $M^{(r)}$. $\square$

Applying this to the probability of A being a leaf of T_3, we write

$$\mathbf{P}(BA \in \mathbf{T} \text{ and } CA, DA, EA \notin \mathbf{T})$$

$$= \det M^{(3)}(BA, CA, DA, EA)$$

$$= \begin{vmatrix} 8/18 & 3/18 & 4/18 & 3/18 \\ -3/18 & 10/18 & -3/18 & -4/18 \\ -4/18 & -3/18 & 10/18 & -3/18 \\ -3/18 & -4/18 & -3/18 & 10/18 \end{vmatrix} = \frac{10584}{18^4} = \frac{1176}{11664}$$

so A is a leaf of $4 \cdot 1176 = 4704$ of the 11664 spanning trees of T_3. This time, the Matrix-Tree Theorem would have required evaluation of several different eight by eight determinants. If T_3 were replaced by T_n, the transfer-impedance calculation would not

be significantly harder, but the Matrix-Tree Theorem would require several n^2 by n^2 determinants. If n goes to ∞, as it might when calculating some sort of limit behavior, these large determinants would not be tractable.

5   Poisson limits

As mentioned in the introduction, the random degree of a vertex in a uniform spanning tree of G converges in distribution to one plus a Poisson(1) random variable as G gets larger and more highly connected. This section investigates some such limits, beginning with an example symmetric enough to compute explicitly. The reason for this limit may seem clearer at the end of the section when we discuss a stronger limit theorem. Proofs in this section are mostly sketched since the details occupy many pages in [6].

5.1   The degree of a vertex in K_n

The simplest situation in which to look for a Poisson limit is on the complete graph K_n. This is pictured here for $n = 8$.

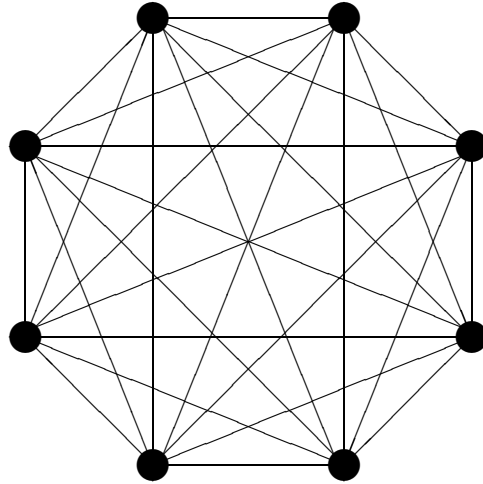Calculating the voltages for a complete graph is particularly easy because of all the symmetry. Say the vertices of K_n are called $v_1, \ldots, v_n$, and put a one volt battery across v_1 and v_2, so $V(v_1) = 1$ and $V(v_2) = 0$. By Theorem 3.4, the voltage at any other vertex v_j is equal to the probability that $SRW^{K_n}_{v_j}$ hits v_1 before v_2. This is clearly equal to $1/2$. The total current flow out of v_1 with these voltages is $n/2$, since one amp flows along the edge to v_2 and $1/2$ amp flows along each of the $n-2$ other edges out of v_1. Multiplying by $2/n$ to get a unit current flow gives voltages

$$
V(v_i) = \begin{cases} 2/n & : \quad i = 1 \\ 0 & : \quad i = 2 \\ 1/n & \quad \text{otherwise.} \end{cases}
$$

The calculations will of course come out similarly for a unit current flow supplied across any other edge of K_n.

The first distribution we are going to examine is of the degree in $\mathbf{T}$ of a vertex, say v_1. Since we are interested in which of the edges incident to v_1 are in $\mathbf{T}$, we need to

calculate $H(\overline{v_1 v_i}, \overline{v_1 v_j})$ for every $i, j \neq 1$. Orienting all of these edges away from v_1 and using the voltages we just worked out gives

$$H(\overline{v_1 v_i}, \overline{v_1 v_j}) = \begin{cases} 2/n & : & i = j \\ 1/n & & \text{otherwise} \end{cases}.$$

Denoting the edge from v_1 to v_i by e_i, we have the $n-1$ by $n-1$ matrices

$$M(e_2, \ldots, e_n) = \begin{pmatrix} \frac{2}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{2}{n} & \cdots & \frac{1}{n} \\ & & \vdots & \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{2}{n} \end{pmatrix} \qquad M^{(n-1)}(e_2, \ldots, e_n) = \begin{pmatrix} \frac{n-2}{n} & \frac{-1}{n} & \cdots & \frac{-1}{n} \\ \frac{-1}{n} & \frac{n-2}{n} & \cdots & \frac{-1}{n} \\ & & \vdots & \\ \frac{-1}{n} & \frac{-1}{n} & \cdots & \frac{n-2}{n} \end{pmatrix}.$$

There must be at least one edge in $\mathbf{T}$ incident to v_1 so Theorem 4.3 says $\det M^{(n-1)} = \mathbf{P}(e_2, \ldots, e_n \notin \mathbf{T}) = 0$. This is easy to verify: the rows sums to zero. We can use $M^{(n-1)}$ to calculate the probability that e_2 is the only edge in $\mathbf{T}$ incident to v_1 by noting that this happens if and only if $e_3, \ldots, e_n \notin \mathbf{T}$. This is the determinant of $M^{(n-2)}(e_3, \ldots, e_n)$ which is a matrix smaller by one thatn $M^{(n-1)}(e_2, \ldots, e_n)$ but which still has $(n-2)/n$'s down the diagonal and $-1/n$'s elsewhere. This is a special case of a *circulant* matrix, which is a type of matrix whose determinant is fairly easy to calculate.

A k by k circulant matrix is an M for which $M(i, j)$ is some number $a(i-j)$ depending only on $i - j \bmod k$. Thus M has a_0 all down the diagonal for some a_0, a_1 on the next diagonal, and so forth. The eigenvalues of a circulant matrix $\lambda_0, \ldots, \lambda_{k-1}$ are given by $\lambda_j = \sum_{t=0}^{k-1} a_t \zeta^{jt}$ where $\zeta = e^{2\pi i/n}$ is the nth root of unity. It is easy to verify that these are the eigenvalues, by checking that the vector $\vec{w}$ for which $w_t = \zeta^{tj}$ is an eigenvector for M (no matter what the a_i are) and has eigenvalue λ_j. The determinant is then the product of the eigenvalues. Details of this may be found in [17].

In the case of $M^{(n-2)}$, $a_0 = (n-2)/n$ and $a_j = -1/n$ for $j \neq 0$. Then $\lambda_0 = \sum_j a_j = 1/n$. To calculate the other eigenvalues note that for any $j \neq 0 \bmod n - 2$, $\sum_{t=1}^{n-3} \zeta^{jt} = 0$. Then $\lambda_j = (n-2)/n \sum_{t=1}^{n-3}(-1/n)\zeta^{jt} = (n-1)/n - (1/n)\sum_{t=0}^{n-3} \zeta^{tj} = (n-1)/n$. This gives

$$\det M^{(n-2)} = \prod_{j=0}^{n-3} \lambda_j = \frac{1}{n}\left(\frac{n-1}{n}\right)^{n-3} = \frac{1+o(1)}{ne}$$

as $n \to \infty$. [2] Part of the Poisson limit has emerged: the probability that v_1 has degree one in $\mathbf{T}$ is (by symmetry) $n-1$ times the probability that the particular edge e_2 is the only edge in $\mathbf{T}$ incident to v_1; this is $(n-1)(1+o(1))/en$ so it converges to e^{-1} as $n \to \infty$. This is $\mathbf{P}(X = 1)$ where X is one plus a Poisson(1) , i.e. a Poisson of mean one.

Each further part of the Poisson limit requires a more careful evaluation of the limit. To illustrate, we carry out the second step. Use one more degree of precision in the Taylor series for $\ln(x)$ and $\exp(x)$ to get

$$n^{-1}\left(\frac{n-1}{n}\right)^{n-3}$$
$$= n^{-1}\exp[(n-3)(-n^{-1} - n^{-2}(1/2 + o(1)))]$$
$$= n^{-1}\exp[-1 + (5/2 + o(1))n^{-1}]$$
$$= n^{-1}e^{-1}[1 + (5/2 + o(1))n^{-1}].$$

The reason we need this precision is that we are going to calculate the probability of v_1 having degree 2 by summing the $\mathbf{P}(e, f$ are the only edges incident to v_1 in $\mathbf{T})$ over all pairs of edges e, f coming out of v_1. By symmetry this is just $(n-1)(n-2)/2$ times the probability that the particular edges e_2 and e_3 are the only edges in $\mathbf{T}$ incident to v_1. This probability is the determinant of a matrix which is not a circulant, and to avoid calculating a difficult determinant it is better to write this probability as the following

---

[2]Here, $o(1)$ signifies a quantity going to zero as $n \to \infty$. This is a convenient and standard notation that allows manipulation such as $(2 + o(1))(3 + o(1)) = 6 + o(1)$.

difference: the probability that no edges other that e_2 and e_3 are incident to v_1 minus the probability that e_2 is the only edge incident to v_1 minus the probability that e_2 is the only edge incident to v_3. Since the final probability is this difference multiplied by $(n-1)(n-2)/2$, the difference should be of order n^{-2}, which explains why this degree of precision is required for the latter two probabilities.

The probability of $\mathbf{T}$ containing no edges incident to v_1 other than e_2 and e_3 is the determinant of $M^{(n-3)}(e_4, \ldots, e_n)$, which is an $n-3$ by $n-3$ circulant again having $(n-2)/n$ on the diagonal and $-1/n$ elsewhere. Then $\lambda_0 = \sum_{j=0}^{n-4} a_j = 2/n$ and $\lambda_j = (n-1)/n$ for $j \neq 0 \bmod n-3$, yielding

$$\det M^{(n-3)} = 2n^{-1}\left(\frac{n-1}{n}\right)^{n-4} = 2n^{-1}e^{-1}[1 + (7/2 + o(1))n^{-1}]$$

in the same manner as before. Subtracting off the probabilities of e_2 or e_3 being the only edge in $\mathbf{T}$ incident to v_1 gives

$$\mathbf{P}(e_2, e_3 \in \mathbf{T}, e_4, \ldots, e_n \notin \mathbf{T})$$

$$= \quad 2n^{-1}e^{-1}[1 + (7/2 + o(1))n^{-1}] - 2n^{-1}e^{-1}[1 + (5/2 + o(1))n^{-1}] = (2 + o(1))n^{-2}e^{-1}.$$

Multiplying by $(n-1)(n-2)/2$ gives

$$\mathbf{P}(v_1 \text{ has degree 2 in } \mathbf{T}) \to e^{-1}$$

as $n \to \infty$, which is $\mathbf{P}(X = 2)$ where X is one plus a Poisson(1).

5.2   Another point of view

The calculations of the last section may be continued *ad infinitum*, but each step requires a more careful estimate so it pays to look for a way to do all the steps at once. The right alternative method will be more readily apparent if we generalize to graphs other than

K_n which do not admit such a precise calculation (if a tool that is difficult to use breaks, you may discover a better one).

The important feature about K_n was that the voltages were easy to calculate. There is a large class of graphs for which the voltages are just as easy to calculate approximately. The term "approximately" can be made more rigorous by considering sequences of graphs G_n and stating approximations in terms of limits as $n \to \infty$. Since I've always wanted to name a technical term after my dog, call a sequence of graphs G_n *Gino-regular* if there is a sequence D_n such that

(i) The maximum and minimum degree of a vertex in G_n are $(1 + o(1))D_n$ as $n \to \infty$; and

(ii) The maximum and minimum over vertices $x \neq y, z$ of G_n of the probability that $SRW_x^{G_n}$ hits y before z are $1/2 + o(1)$ as $n \to \infty$.

Condition (ii) implies that $D_n \to \infty$, so the graphs G_n are growing locally. It is not hard to see that the voltage $V(z)$ in a unit current flow across any edge $e = \vec{xy}$ of a graph G_n in a Gino-regular sequence is $(1 + o(1))D_n^{-1}(\delta_x - \delta_y)(z)$ uniformly over all choices of $x, y, z \in G_n$ as $n \to \infty$. The complete graphs K_n are Gino-regular. So are the n-cubes, B_n, whose vertex sets are all the n-long sequences of zeros and ones and whose edges connect sequences differing in only one place.

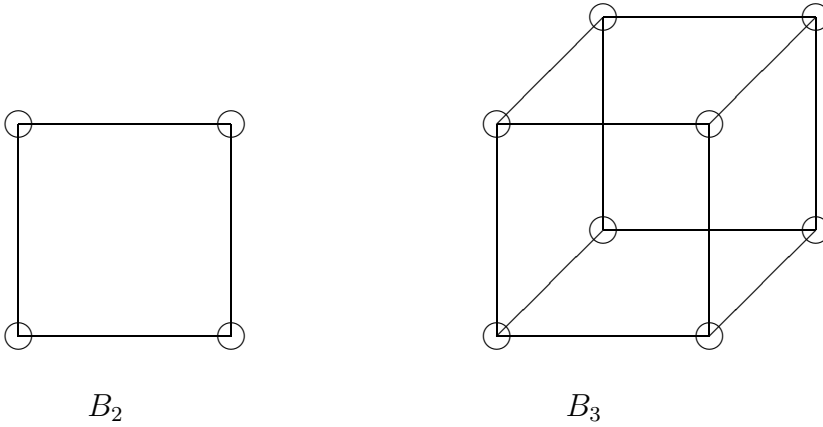B_2 B_3

figure 10

To see why $\{B_n\}$ is Gino-regular, consider the "worst case" when x is a neighbor of y. There is a small probability that $SRW_x(1)$ will equal y, small because this is degree$(x)^{-1} = (1 + o(1))D_n^{-1}$ which is going to zero. There are even smaller probabilities of reaching y in the next few steps; in general, unless SRW_x hits y in one step, it tends to get "lost" and by the time it comes near y or z again it is thoroughly random and is equally likely to hit y or z first. In fact Gino-regular sequences may be thought of as graphs that are nearly degree-regular, which SRW gets lost quickly.

The approximate voltages give approximate transfer-impedances $H(e, f) = (2 + o(1))/n$ if $e = f$, $(1 + o(1))/n$ if e and f meet at a single vertex (choose orientations away from the vertex) and $o(1)/n$ if e and f do not meet. The determinant of a matrix is continuous in its entries, so it may seem that we have everything necessary to calculate limiting probabilities as limits of determinants of transfer-impedance matrices. If v is a vertex in G_k and $e_1, \ldots, e_n$ are the edges incident to v in G_k (so $n \approx D_k$), then the

56

probability of e_2 being the only edge in $\mathbf{T}$ incident to v is the determinant of

$$M^{(n-1)}(e_2,\ldots,e_n) = \begin{pmatrix} (n-2+o(1))/n & (-1+o(1))/n & \cdots & (-1+o(1))/n \\ (-1+o(1))/n & (n-2+o(1))/n & \cdots & (-1+o(1))/n \\ & & & \\ & & \vdots & \\ & & & \\ (-1+o(1))/n & (-1+o(1))/n & \cdots & (n-2+o(1))/n \end{pmatrix}.$$

Unfortunately, the matrix is changing size as $n \to \infty$, so convergence of each entry to a known limit does not give us the limit of the determinant.

If the matrix were staying the same size, the problem would disappear. This means we can successfully take the limit of probabilities of events as long as they involve a bounded number of edges. Thus for any fixed edge e_1, $\mathbf{P}(e_1 \in \mathbf{T}) = \det M(e_1) = (1+o(1))(2/n)$. For any fixed pair of edges e_1 and e_2 incident to the same vertex,

$$\mathbf{P}(e_1, e_2 \in \mathbf{T}) \;=\; \det M(e_1, e_2) = \begin{vmatrix} (2+o(1))/n & (1+o(1))/n \\ (1+o(1))/n & (2+o(1))/n \end{vmatrix} \;=\; (3+o(1))n^{-2}.$$

In general if $e_1,\ldots,e_r$ are all incident to v then the transfer-impedance matrix is n^{-1} times an r by r matrix converging to the matrix with 2 down the diagonal and 1 elsewhere. The eigenvalues of this circulant are $\lambda_0 = r+1$ and $\lambda_j = 1$ for $j \neq 0$, yielding

$$\mathbf{P}(e_1,\ldots,e_r \in \mathbf{T}) = (r+1+o(1))n^{-r}.$$

What can we do with these probabilities? Inclusion-exclusion fails for the same reason as the large determinants fail – the $o(1)$ errors pile up. On the other hand, these probabilities determine certain expectations. Write $e_1,\ldots,e_n$ again for the edges adjacent to v and I_i for the indicator function which is one when $e_i \in \mathbf{T}$ and zero otherwise; then

$$\sum_i \mathbf{P}(e_i \in \mathbf{T}) = \sum_i \mathbf{E} I_i = \mathbf{E} \sum_i I_i = \mathbf{E} \deg(v).$$

This tells us that $\mathbf{E}\deg(v) = n(2 + o(1))n^{-1} = 2 + o(1)$. If try this with ordered pairs of edges, we get

$$\sum_{i \neq j} \mathbf{P}(e_i, e_j \in \mathbf{T}) = \sum_{i \neq j} \mathbf{E}I_i I_j = \mathbf{E}\sum_{i \neq j} I_i I_j.$$

This last quantity is the sum of all distinct ordered pairs of edges incident to v of the quantity: 1 if they are both in the tree and 0 otherwise. If $\deg(v) = r$ then a one occurs in this sum $r(r-1)$ times, so the sum is $\deg(v)(\deg(v) - 1)$. The determinant calculation gave $\mathbf{P}(e_i, e_j \in \mathbf{T}) = (3 + o(1))n^{-2}$ for each i, j, so

$$\mathbf{E}[\deg(v)(\deg(v) - 1)] = n(n-1)(3 + o(1))n^{-2} = 3 + o(1).$$

In general, using ordered r-tuples of distinct edges gives

$$\mathbf{E}[\deg(v)(\deg(v) - 1)\cdots(\deg(v) - r + 1)]$$
$$= n(n-1)\cdots(n - r + 1)(r + 1 + o(1))n^{-r}$$
$$= r + 1 + o(1).$$

Use the notation $(A)_r$ to denote $A(A - 1)\cdots(A - r + 1)$ which is called the rth lower factorial of A. If Y_n is the random variable $\deg(v)$ then we have succinctly,

$$\mathbf{E}(Y_n)_r = r + 1 + o(1). \tag{8}$$

$\mathbf{E}(Y_n)_r$ is called the rth factorial moment of Y_n.

If you remember why we are doing these calculations, you have probably guessed that $\mathbf{E}(X)_r = r + 1$ when X is one plus a Poisson(1). This is indeed true and can be seen easily enough from the logarithmic moment generating function $\mathbf{E}t^X$ via the identity

$$\mathbf{E}(X)_r = \left(\frac{d}{dt}\right)^r\bigg|_{t=1} \mathbf{E}t^X,$$

using $\mathbf{E}t^X = \mathbf{E}e^{X\ln(t)} = \phi(\ln(t)) = te^{t-1}$; consult [14, page 301] for details. All that we need now for a Poisson limit result is a theorem saying that if the factorial moments of

Y_n are each converging to the factorial moments of X, then Y_n is actually converging in distribution to X. This is worth spending a short subsection on because it is algebraically very neat.

5.3 The method of moments

A standard piece of real analysis shows that if all the factorial moments of a sequence of random variables converging to a limit are finite, then for each r, the limit of the rth factorial moments is the rth factorial moment of the limit. (This is essentially the Lebesgue-dominated convergence theorem.) Another standard result is that if the moments of a sequence of random variables converge, then the sequence, or at least some subsequence is converging in distribution to some other random variable whose moments are the limits of the moments in the sequence. Piecing together these straight-forward facts leaves a serious gap in our prospective proof: What if there is some random variable Z distributed differently from X with the same factorial moments? If this could happen, then there would be no reason to think that Y_n converged in distribution to X rather than Z. This scenario can actually happen – there really are differently distributed random variables with the same moments! (See the discussion of the lognormal distribution in [9].) Luckily this only happens when X is badly behaved, and a Poisson plus one is not badly behaved. Here then is a proof of the fact that the distribution of X is the only one with r^{th} factorial moment $r + 1$ for all r. I will leave it to you to piece together, look up in [9] or take on faith how this fact plus the results from real analysis imply $Y \xrightarrow{\mathcal{D}} X$.

Theorem 5.1 *Let X be a random variable with $\mathbf{E}(X)_r \leq e^{kr}$ for some k. Then no random variable distributed differently from X has the same factorial moments.*

Proof: The factorial moments $\mathbf{E}(X)_r$ determine the regular moments $\mu_r = \mathbf{E}X^r$ and *vice versa* by the linear relations $(X)_1 = X^1; (X)_2 = X^2 - X^1$, etc. From these linear

relations it also follows that factorial moments are bounded by some e^{kr} if and only if regular moments are bounded by some e^{kr}, thus it suffices to prove the theorem for regular moments. Not only do the moments determine the distribution, it is even possible to calculate $\mathbf{P}(X = j)$ directly from the moments of X in the following manner.

The *characteristic function* of X is the function $\phi(t) = \mathbf{E}e^{itX}$ where $i = \sqrt{-1}$. This is determined by the moments since $\mathbf{E}e^{itX} = \mathbf{E}(1 + (itX) + (itX)^2/2! + \cdots) = 1 + it\mu_1 + (it)^2\mu_2/2! + \cdots$. We use the exponential bound on the growth of μ_r to deduce that this is absolutely convergent for all t (though a somewhat weaker condition would do). The growth condition also shows that $\mathbf{E}e^{itX}$ is bounded and absolutely convergent for $y \in [0, 2\pi]$. Now $\mathbf{P}(X = j)$ can be determined by Fourier inversion:

$$\frac{1}{2\pi} \int_0^{2\pi} \mathbf{E}e^{itX} e^{-ijt} dt$$

$$= \frac{1}{2\pi} \int_0^{2\pi} [\sum_{r \geq 0} e^{itr} \mathbf{P}(X = r)] e^{-ijt} dt$$

$$= \frac{1}{2\pi} \sum_{r \geq 0} \mathbf{P}(X = r) \int_0^{2\pi} e^{itr} e^{-ijt} dt$$

(switching the sum and integral is OK for bounded, absolutely convergent integrals)

$$= \frac{1}{2\pi} \sum_{r \geq 0} \mathbf{P}(X = r) \delta_0(r - j)$$

$$= \mathbf{P}(X = j).$$

$\square$

5.4 A branching process

In the last half of section 1.4 I promised to explain how convergence in distribution of $\deg(v)$ was a special case of convergence of $\mathbf{T}$ near v to a distribution called $\mathcal{P}_1$. (You might want to go back and reread that section before continuing.) The infinite tree $\mathcal{P}_1$ is interesting in its own right and I'll start making good on the promise by describing $\mathcal{P}_1$.

This begins with a short description of *Galton-Watson branching processes*. You can think of a Galton-Watson process as a family tree for some fictional amoebas. These fictional amoebas reproduce by splitting into any number of smaller amoebas (unlike real amoebas that can only split into two parts at a time). At time $t = 0$ there is just a single amoeba, and at each time $t = 1, 2, 3, \ldots$, each living amoeba $\mathcal{A}$ splits into a random number $N = N_t(\mathcal{A})$ of amoebas, where the random numbers are independent and all have the same distribution $\mathbf{P}(N_t(\mathcal{A}) = j) = p_j$. Allow the possibility that $N = 0$ (the amoeba died) or that $N = 1$ (the amoeba didn't do anything). Let $\mu = \sum_j j p_j$ be the mean number of amoebas produced in a split. A standard result from the theory of branching processes [4] is that if $\mu > 1$ then there is a positive probability that the family tree will survive forever, the population exploding exponentially as in the usual Malthusian forecasts for human population in the twenty-first century. Conversely when $\mu < 1$, the amoeba population dies out with probability 1 and in fact the chance of it surviving n generations decreases exponentially with n. When $\mu = 1$ the branching process is said to be *critical*. It must still die out, but the probability of it surviving n generations decays more slowly, like a constant times $1/n$. The theory of branching processes is quite large and you can find more details in [4] or [10].

Specialize now to the case where the random number of offspring has a Poisson(1) distribution, i.e. $p_j = e^{-1}/j!$. Here's the motivation for considering this case. Imagine a graph G in which each vertex has N neighbors and N is so large it is virtually infinite. Choose a subgraph U by letting each edge be included independently with probability

N^{-1}. Fix a vertex $v \in G$ and look at the vertices connected to v in U. The number of neighbors of v in U has a Poisson(1) distribution by the standard characterization of a Poisson as the limit of number of occurrences of rare events. For each neighbor y of v in U, there are $N - 1$ edges out of y other than the one to v, and the number of those in U will again be Poisson(1) (since $N \approx \infty$, subtracting one does not matter) and continuing this way shows that the connected component of v in U is distributed as a Galton-Watson process with Poisson(1) offspring.

Of course U is not distributed like a uniform spanning tree $\mathbf{T}$. For one thing, U may with probability e^{-1} fail to have any edges out of v. Even if this doesn't happen, the chance of U having more than n vertices goes to zero as $n \to \infty$ (a critical Galton-Watson process dies out) whereas $\mathbf{T}$, being a spanning tree of an almost infinite graph, goes on as far as the eye can see. The next hope is that $\mathbf{T}$ looks like U conditioned not to die out. This should in fact seem plausible: you can check that U has no cycles near v since virtually all of the N edges out of each neighbor of v lead further away from v; then a uniform spanning tree should be a random cycle-free graph U that treats each edge as equally likely, conditioned on being connected.

The conditioning must be done carefully, since the probability of U living forever is zero, but it turns out fine if you condition on U living for at least n generations and take the limit as $n \to \infty$. The random infinite tree $\mathcal{P}_1$ that results is called the *incipient infinite cluster* at v, so named by percolation theorists (people who study connectivity properties of random graphs). It turns out there is an alternate description for the incipient infinite cluster. Let $v = v_0, v_1, v_2, \ldots$ be a single line of vertices with edges $\overline{vv_1}, \overline{v_1v_2}, \ldots$. For each of the vertices v_i independently, make a separate independent copy U_i of the critical Poisson(1) branching process U with v_i as the root and paste it onto the line already there. Then this collage has the same distribution as $\mathcal{P}_1$. This fact is the "whole tree" version of the fact that a Poisson(1) conditioned to be nonzero is distributed as one plus a Poisson(1) (you can recover this fact from the fact about $\mathcal{P}_1$ by

looking just at the neighbors of v).

5.5 Tree moments

To prove that a uniform spanning tree $\mathbf{T}_n$ of G_n converges in distribution to $\mathcal{P}_1$ when G_n is Gino-regular, we generalize factorial moments to trees. Let t be a finite tree rooted at some vertex x and let W be a tree rooted at v. W is allowed to be infinite but it must be locally finite – only finitely many edges incident to any vertex. Say that a map f from the vertices of t to the vertices of W is a *tree-map* if f is one to one, maps x to v and neighbors to neighbors. Let $N(W;t)$ count the number of tree-maps from t into W. For example in the following picture, $N(W;t) = 4$, since C and D can map to H and I in either order with A mapping to E, and B can map to F or G.
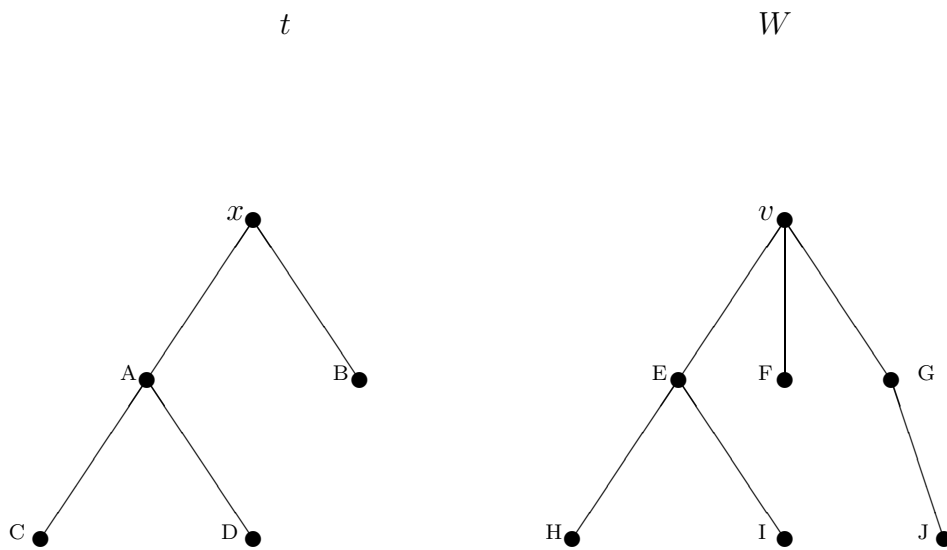

figure 11

Define the tth *tree-moment* of a random tree Z rooted at v to be $\mathbf{E}N(Z;t)$. If t is an n-star, meaning a tree consisting of n edges all emanating from x, then a tree-map from t to W is just a choice of n distinct neighbors of v in order, so $N(W;t) = (\deg(v))_n$. Thus $\mathbf{E}N(Z;t) = \mathbf{E}(\deg(v))_n$, the nth factorial moment of $\deg(v)$. This is to show you that tree-moments generalize factorial moments. Now let's see what the tree-moments of $\mathcal{P}_1$ are. Let t be any finite tree and let $|t|$ denote the number of vertices in t.

Lemma 5.2 *Let U be a Galton-Wastson process rooted at v with Poisson(1) offspring. Then $\mathbf{E}N(U;t) = 1$ for all finite trees t.*

Proof: Use induction on t, the lemma being clear when t is a single vertex. The way the induction step works for trees is to show that if a fact is true for a collection of trees $t_1, \ldots, t_n$ then it is true for the tree t_* consisting of a root x with n neighbors $x_1, \ldots, x_n$ having subtrees $t_1, \ldots t_n$ respectively as in the following illustration.
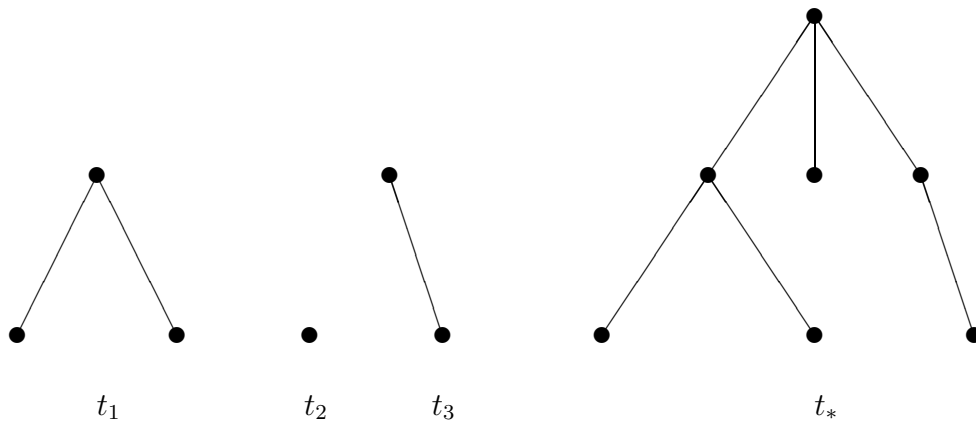

figure 12

So let $t_1, \ldots, t_n$ and t_* be as above. Any tree-map $f : t_* \to U$ must map the n neighbors of v into distinct neighbors of U and the expected number of ways to do this

is $\mathbf{E}(\deg(v))_n$ which is one for all n since $\deg(v)$ is a Poisson(1) [9]. Now for any such assignment of f on the neighbors of v, the number of ways of completing the assignment to a tree-map is the product over $i = 1, \ldots n$ of the number of ways of mapping each t_i into the subtree of U below $f(x_i)$. After conditioning on what the first generation of U looks like, the subtrees below any neighbors of v are independent and themselves Galton-Watsons with Poisson(1) offspring. (This is what it means to be Galton-Watson.) By induction then, the expected number of ways of completing the assignment of f is the product of a bunch of ones and is therefore one. Thus $\mathbf{E}N(U; t) = \mathbf{E}(\deg(v))_n \prod_{i=1}^{n} 1 = 1$.
$\square$

Back to calculating $\mathbf{E}N(\mathcal{P}_1; t)$. Recall that $\mathcal{P}_1$ is a line $v_0, v_1, \ldots$ with Poisson(1) branching processes U_i stapled on. Each tree-map $f : t \rightarrow \mathcal{P}_1$ hits some initial segment $v_0, \ldots v_k$ of the original line, so there is some vertex $y_f \in t$ such that $f(y_f) = v_k$ for some k but v_{k+1} is not in the image of f. For each $y \in t$, we count the expected number of tree-maps f for which $y_f = y$. There is a path $x = f^{-1}(v_0), \ldots, f^{-1}(v_k) = y$ in t going from the root x to y. The remaining vertices of t can be separated into $k + 1$ subtrees below each of the $f^{-1}(v_i)$. These subtrees must then get mapped respectively into the U_i. By the lemma, the expected number of ways of mapping anything into a U_i is one, so the expected number of f for which $y_f = y$ is $\prod_{i=1}^{k} 1 = 1$. Summing over y then gives

$$\mathbf{E}N(\mathcal{P}_1; t) = |t| \tag{9}$$

The last thing we are going to do to in proving the stronger Poisson convergence theorem is to show

Lemma 5.3 *Let G_n be a Gino-regular sequence of graphs, and let $\mathbf{T}_n$ be a uniform spanning tree of G_n rooted at some v_n. Then for any finite rooted tree t, $\mathbf{E}N(\mathbf{T}_n; t) \rightarrow |t|$ as $n \rightarrow \infty$.*

It is not trivial from here to establish that $\mathbf{T} \wedge r$ converges in distribution to $\mathcal{P}_1 \wedge r$ for

every r. The standard real analysis facts I quoted in section 5.3 about moments need to be replace by some not-so-standard (but not too hard) facts about tree-moments. Suffice it to say that the previous two lemmas do in the end prove (see [6] for details)

Theorem 5.4 *Let G_n be a Gino-regular sequence of graphs, and let $\mathbf{T}_n$ be a uniform spanning tree of G_n rooted at some v_n. Then for any r, $\mathbf{T}_n \wedge r$ converges in distribution to $\mathcal{P}_1 \wedge r$ as $n \to \infty$.*

Sketch of proof of Lemma 5.3: Fix a finite t rooted at x. To calculate the expected number of tree-maps from t into $\mathbf{T}_n$ we will sum over every possible image of a tree-map the probability that all of those edges are actually present in $\mathbf{T}_n$. By an image of a tree-map, I mean two things: (1) a collection $\{v_x : x \in t\}$ of vertices of G_n indexed by the vertices of t for which $v_x \sim v_y$ in G whenever $x \sim y$ in t; (2) a collection of edges e_ϵ connecting v_x and v_y for every edge $\epsilon \in t$ connecting some x and y. Fix such an image.

The transfer-impedance theorem tells us that the probability of finding all the edges v_e in $\mathbf{T}$ is the determinant of $M(e_\epsilon : \epsilon \in t)$. Now for edges $e, e' \in G$, Gino-regularity gives that $H(e, e') = D_n^{-1}(o(1) + \kappa)$ uniformly over edges of G_n, where κ is $2, 1$ or 0 according to whether $e = e'$, they share an endpoint, or they are disjoint. The determinant is then well approximated by the corresponding determinant without the $o(1)$ terms, which can be worked out as exactly $|t| D_n^{1-|t|}$.

This must now be summed over all possible images, which amounts to multiplying $|t| D_n^{1-|t|}$ by the number of possible images. I claim the number of possible images is approximately $D_n^{|t|-1}$. To see this, imagine starting at the root x, which must get mapped to v_n, and choosing successively where to map each nest vertex of t. Since there are approximately D_n edges coming out of each vertex of G_n, there are always about D_n choices for the image of the next vertex (the fact that you are not allowed to choose any vertex already chosen is insignificant as D_n gets large). There are $|t| - 1$ choices, so the number of maps is about $D_n^{|t|-1}$. This proves the claim. The claim implies that the

66

expected number of tree-maps from t to $\mathbf{T}_n$ is $|t|D_n^{1-|t|}D_n^{|t|-1} = |t|$, proving the lemma.
$\square$

6 Infinite lattices, dimers and entropy

There is, believe it or not, another model that ends up being equivalent to the uniform spanning tree model under a correspondence at least as surprising as the correspondence between spanning trees and random walks. This is the so-called *dimer* or *domino tiling* model, which was studied by statistical physicists quite independently of the uniform spanning tree model. The present section is intended to show how one of the fundamental questions of this model, namely calculating its entropy, can be solved using what we know about spanning trees. Since it's getting late, there will be pictures but no detailed proofs.

6.1 Dimers

A dimer is a substance that on the molecular level is made up of two smaller groups of atoms (imagine two spheres of matter) adhering to each other via a covalent bond; consequently it is shaped like a dumbbell. If a bunch of dimer molecules are packed together in a cold room and a few of the less significant laws of physics are ignored, the molecules should array themselves into some sort of regular lattice, fitting together as snugly as dumbbells can. To model this, let r be some positive real number representing the length of one of the dumbbells. Let L be a lattice, i.e. a regular array of points in three-space, for which each point in L has some neighbors at distance r. For example r could be 1 and L could be the standard integer lattice $\{(x, y, z) : x, y, z \in \mathbb{Z}\}$, so r is the minimum distance between any two points of L (see the picture below). Alternatively r could be $\sqrt{2}$ or $\sqrt{3}$ for the same L. Make a graph G whose vertices are the points of L, with an edge between any pair of points at distance r from each other. Then the possible

packings of dimers in the lattice are just the ways of partitioning the lattice into pairs of vertices, each pair (representing one molecule) being the two enpoints of some edge. The following picture shows part of a packing of the integer lattice with nearest-neighbor edges.
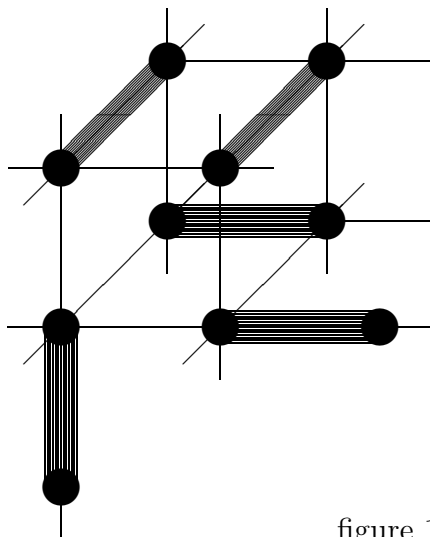

figure 13

Take a large finite box inside the lattice, containing N vertices. If N is even and the box is not an awkward shape, there will be not only one but many ways to pack it with dimers. There will be several edges incident to each vertex v, representing a choice to be made as to which other vertex will be covered by the molecule with one atom covering v. These choices obviously cannot be made independently, but it should be plausible from this that the total number of configurations is approximately γ^N for some $\gamma > 1$ as N goes to infinity. This number can be written alternatively as e^{hN} where $h = \ln(\gamma)$ is called the entropy of the packing problem. The thermodynamics of the resulting substance depend on, among other things, the entropy h.

The case that has been studied the most is where L is the two-dimensional integer lattice with $r = 1$. The graph G is then the usual nearest-neighbor square lattice. Physically this corresponds to packing the dimers between two slides. You can get the same

packing problem by attempting to tile the plane with dominos – vertical and horizontal 1 by 2 rectangles – which is why the model also goes by the name of domino tiling.

6.2 Dominos and spanning trees

We have not yet talked about spanning trees of an infinite graph, but the definition remains the same: a connected subgraph touching each vertex and contaning no cycles. If the subgraph need not be connected, it is a spanning forest. Define an *essential spanning forest* or ESF to be a spanning forest that has no finite components. Informally, an ESF is a subgraph that you can't distinguish from a spanning tree by only looking at a finite part of it (since it has no cycles or *islands*).

Let G_2 denote the nearest-neighbor graph on the two dimensional integer lattice. Since G_2 is a planar graph, it has a *dual* graph G_2^*, which has a vertex in each cell of G_2 and an edge e^* crossing each edge e of G_2. In the following picture, filled circles and heavy lines denote G_2 and open circles and dotted lines denote G_2^*. Note that G_2, together with G_2^* and the points where edges cross dual edges, forms another graph $\tilde{G}_2$ that is just G_2 scaled down by a factor of two.
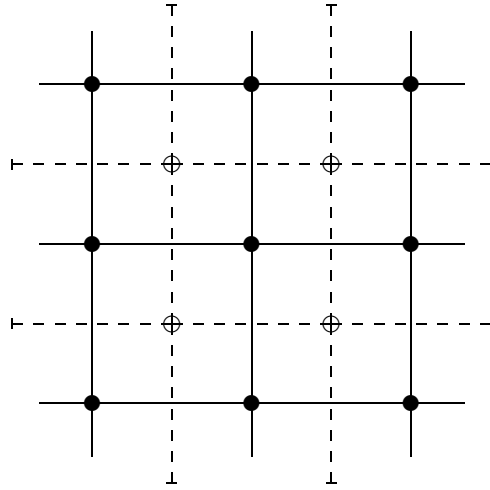
figure 14

Each subgraph H of G has a dual subgraph H^* consisting of all edges e^* of G^* dual to edges e <u>not</u> in H. If H has a cycle, then the duals of all edges in the cycle are absent from H^* which separates H^* into two components: the interior and exterior of the cycle. Similarly, an island in H corresponds to a cycle in H^* as in the picture:
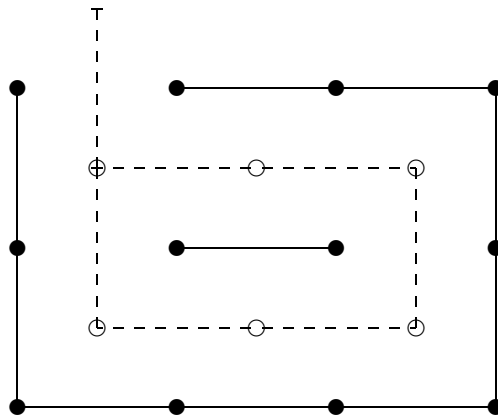

figure 15

From this description, it is clear that T is an essential spanning forest of G_2 if and only if T^* is an essential spanning forest of $G_2^* \cong G_2$.

70

Let T now an infinite tree. We define *directed* a little differently than in the finite case: say T is directed if the edges are oriented so that every vertex has precisely one edge leading out of it. Following the arrows from any vertex gives an infinite path and it is not hard to check that any two such paths from different vertices eventually merge. Thus directedness for infinite trees is like directedness for finite trees, toward a vertex at infinity.

Say an essential spanning forest of G_2 is directed if a direction has been chosen for each of its components and each of the components of its dual. Here then is the connection between dominos and essential spanning forests.

Let T be a directed essential spanning forest of G_2, with dual T^*. Construct a domino tiling of $\tilde{G}_2$ as follows. Each vertex $v \in V(G_2) \subseteq V(\tilde{G}_2)$ is covered by a domino that also covers the vertex of $\tilde{G}_2$ in the middle of the edge of T that leads out of v. Similarly, each vertex $v^* \in V(G_2^*)$ is covered by a domino also covering the middle of the edge of T^* leading out of v. It is easy to check that this gives a legitimate domino tiling: every domino covers two neighboring vertices, and each vertex is covered by precisely one domino.

Conversely, for any domino tiling of $\tilde{G}_2$, directed essential spanning forests T and T^* for G_2 and G_2^* can be constructed as follows. For each $v \in V(G_2)$, the oriented edge leading out of v in T is the one along which the domino covering v lies (i.e. the one whose midpoint is the other vertex of $\tilde{G}_2$ covered by the domino covering v). Construct T^* analogously. To show that T and T^* are directed ESF's amounts to showing there are no cycles, since clearly T and T^* will have one edge coming out of each vertex. This is true because if you set up dominos in such a way as to create a cycle, they will always enclose an odd number of vertices (check it yourself!). Then there is no way to extend this configuration to a legitimate domino tiling of $\tilde{G}_2$.

It is easy to see that the two operations above invert each other, giving a one to one correspondence between domino tilings of $\tilde{G}_2$ and directed essential spanning forests of G_2. To bring this back into the realm of finite graphs requires ironing out some technicalities which I am instead going to ignore. The basic idea is that domino tilings of the $2n$-torus T_{2n} correspond to spanning trees of T_n almost as well as domino tilings of $\tilde{G}_2$ correspond to spanning trees of G_2. Going from directed essential spanning forests to spanning trees is one of the details glossed over here, but explained somewhat in the next subsection. The entropy for domino tilings is then one quarter the entropy for spanning trees, since T_{2n} has four times as many vertices as T_n. Entropy for spanning trees just means the number h for which T_n has approximately e^{hn^2} spanning trees. To calculate this, we use the matrix-tree theorem.

The number of spanning trees of T_n according to this theorem is the determinant of a minor of the matrix indexed by vertices of T_n whose v, w-entry is 4 if $v = w$, -1 if $v \sim w$ and 0 otherwise. If T_n were replaced by n edges in a circle, then this would be a circulant matrix. As is, it is a generalized circulant, with symmetry group $T_n = (\mathbb{Z}/n\mathbb{Z})^2$ instead of $Z/n\mathbb{Z}$. The eigenvalues can be gotten via group representations of T_n, resulting in eigenvalues $4 - 2\cos(2\pi k/n) - 2\cos(2\pi l/n)$ as k and l range from 0 to $n - 1$. The determinant we want is the product of all of these except for the zero eigenvalue at $k = l = 0$. The log of the determinant divided by n^2 is the average of these as k and l vary, and the entropy is the limit of this as $n \to \infty$ which is given by

$$\int_0^1 \int_0^1 \ln(4 - 2\cos(2\pi x) - 2\cos(2\pi y)) \, dx \, dy.$$

6.3 Miscellany

The limit theorems in Section 5 involved letting G_n tend to infinity locally, in the sense that each vertex in G_n had higher degree as n grew larger. Instead, one may consider a sequence such as $G_n = T_n$; clearly the n-torus converges in some sense to G_2 as $n \to \infty$,

so there ought to be some limit theorem. Let $\mathbf{T}_n$ be a uniform spanning tree of G_n. Since G_n is not Gino-regular, the limit may not be $\mathcal{P}_1$ and in fact cannot be since the limit has degree bounded by four. It turns out that $\mathbf{T}_n$ converges in distribution to a random tree $\mathbf{T}$ called the uniform random spanning tree for the integer lattice. This works also for any sequence of graphs converging to the three or four dimensional integer lattices [13]. Unfortunately the process breaks down in dimensions five and higher. There the uniform spanning spanning trees on G_n do converge to a limiting distribution but instead of a spanning tree of the lattice, you get an essential spanning forest that has infinitely many components. If you can't see how the limit of spanning trees could be a spanning forest, remember that an essential spanning forest is so similar to a spanning tree that you can't tell them apart with any finite amount of information.

Another result from this study is that in dimensions $2, 3$ and 4, the uniform random spanning tree $\mathbf{T}$ has only one path to infinity. What this really means is that any two infinite paths must eventually join up. Not only that, but $\mathbf{T}^*$ has the same property. That means there is only one way to direct $\mathbf{T}$, so that each choice of $\mathbf{T}$ uniquely determines a domino tiling of $\tilde{G}_2$. In this way it makes sense to speak of a uniform random domino tiling of the plane: just choose a uniform random spanning tree and see what domino tiling it corresponds to.

That takes care of one of the details glossed over in the previous subsection. It also just about wraps up what I wanted to talk about in this article. As a parting note, let me mention an open problem. Let G be the infinite nearest neighbor graph on the integer lattice in d dimensions and let $\mathbf{T}$ be the uniform spanning tree on G gotten by taking a distributional limit of uniform spanning trees on d-dimensional n-tori as $n \to \infty$ as explained above.

Conjecture 2 *Suppose $d \geq 5$. Then with probability one, each component of the essential spanning forest has only one path to infinity, in the sense that any two infinite paths*

must eventually merge.

References

[1] Aldous, D. (1990). Asymptotic fringe distributions for general families of random trees. *Preprint.*

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

[3] Anatharam, V. and Tsoucas, P. (1988). A proof of the Markov chain tree theorem. *Preprint.*

[4] Athreya, K. and Ney, P. (1972). Branching Processes. Springer-Verlag: New York.

[5] Broder, A. (1988). Generating random spanning trees. In: Symp. foudations of computer sci., Institue for Electrical and Electronic Engineers, New York 442 - 447.

[6] Burton, R. and Pemantle, R. (1991). Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer impedances. *Preprint.*

[7] Cvetkovic, D., Doob, M. and Sachs, H. (1980). Spectra of graphs: theory and application. Academic Press: New York.

[8] Doyle, P. and Snell, J. L. (1984). Random walks and electric networks. Carus Mathematical Monograph number 22: Mathematical Association of America.

[9] Feller, W. (1967). An introduction to probability theory and its applications, volumes I and II. John Wiley and Sons: New York.

[10] Harris, T. (1963). The theory of branching processes. Grundlehren # 119. Springer-Verlag: New York.

[11] Isaacson, D. and Madsen, R. (1976). Markov chains: theory and applications. John Wiley and Sons: New York.

[12] Kastelyn, P. (1961). The statistics of dimers on a lattice. *Physica* **27** 1209 - 1225.

[13] Pemantle, R. (1992). Choosing a spanning tree for the integer lattice uniformly. *Ann. Probab. to appear.*

[14] Ross, S. (1988). A first course in probability, 3rd edition. Macmillan: New York.

[15] Seymour, P. and Welsh, D. (1975). Combinatorial applications of an inequality from statistical mechanics. *Math. Proc. Camb. Phil. Soc.* **77** 485 - 495.

[16] Spitzer, F. (1964). Principles of random walk. Van Nostrand: Princeton.

[17] Stanley, R. (1986). Enumerative combinatorics. Wadsworth: Monterey.

[18] Weinberg, L. (1962). Network analysis and synthesis. McGraw-Hill: New York.

[19] White, N. (1986). Theory of Matroids. Encyclopedia of mathematics and its applications, volume 26. Cambridge University Press: New York.