The Annals of Probability
2011, Vol. 39, No. 5, 1864–1895
DOI: 10.1214/10-AOP618

PERCOLATION ON A PRODUCT OF TWO TREES[1]

BY GADY KOZMA

Weizmann Institute of Science

We show that critical percolation on a product of two regular trees of degree ≥ 3 satisfies the triangle condition. The proof does not examine the degrees of vertices and is not "perturbative" in any sense. It relies on an unpublished lemma of Oded Schramm.

1. Introduction.

1.1. *Schramm's lemma.* In 1998, while working on the exponential decay of correlations problem (a problem which is still open), Oded Schramm proved a lemma which solved the problem in an averaged sense. The lemma generated a lot of excitement in the community of researchers of percolation on groups at the time, and there was hope that it would lead to a full solution of the exponential decay problem, the $p_c < p_u$ problem and other related problems on nonamenable Cayley graphs. That hope never materialized, Schramm never published the lemma (you will see it mentioned here and there in papers of the period, e.g., in the last paragraph of [5]) and moved on to other topics.

Let us describe the settings of Schramm's lemma in its original formulation. We are given a nonamenable Cayley graph[2] G. Denote the identity element of the group by 0 (by which I definitely do not insinuate that the group is Abelian) and for any $x \in G$ let $|x|$ denote the graph distance between 0 and x (which is the same as the distance in the word metric on the group with the given generators). Let p_c be the critical probability for percolation on G. The exponential decay of correlations problem is the conjecture that at p_c one has that the connection probability $\mathbb{P}(0 \leftrightarrow x)$ decays exponentially in $|x|$. Schramm's lemma states that this is true in the following interesting averaged sense. Take random walk R on the original Cayley graph G. Then

$$\mathbb{P}\big(0 \leftrightarrow R(n)\big) \leq c^n. \tag{1}$$

Received March 2010; revised October 2010.

[1]Supported in part by the Israel Science Foundation.

MSC2010 subject classifications. Primary 60K35; secondary 60B99.

Key words and phrases. Percolation on groups, triangle condition, nonamenable groups, mean-field, product of trees.

[2]Cayley graphs, amenable groups, and any other notion not explicitly defined may be found in Wikipedia.

Let us stress again that the random walk is on the original graph G and is only some way to average $\mathbb{P}(0 \leftrightarrow x)$. The probability in (1) is over both the walk and the percolation.

We will now sketch Schramm's argument. If the sketch is too dense do not despair—a very close result, Lemma 5, will be presented below, page 1878, with all details (perhaps too many details). For general background on percolation, see the book [6]. For percolation, random walk and branching processes on transitive graphs see the book [13], especially Chapters 7 and 8. Finally, note that all notations and conventions used in this paper are collected in Section 1.6, page 1871, for convenience.

Proof sketch. Examine the following process on our graph G. Fix some m. We start with $m + 1$ "particles" at 0. Each particle does n steps of simple random walk on G, independently of all other particles. After n steps, it divides into m particles, and each one does n additional steps, divides and so on. In other words, a more-or-less standard branching random walk. The particles can be mapped to the edges of an $m + 1$ regular tree $\mathbf{T}$. The original $m + 1$ particles are mapped to the edges coming out of the root of $\mathbf{T}$. Then for each of these particles, say it corresponds to the edge $(\mathbf{r}, \mathbf{v})$ with $\mathbf{r}$ being the root of the tree and $\mathbf{v}$ one of its children, map the m descendents of the particle to the m edges coming out of $\mathbf{v}$ not in the direction of the root. Continue inductively.

We now throw the percolation into the mix as follows. We define a random subgraph $\mathbf{W}$ of our tree $\mathbf{T}$ as follows. Examine an edge $(\mathbf{v}, \mathbf{w})$ of the tree. It corresponds to some particle in the branching process which started from some v, walked n steps and landed on w. We declare that $(\mathbf{v}, \mathbf{w}) \in \mathbf{W}$ if and only if $v \leftrightarrow w$. At the formal level, $\mathbf{W}$ is a random subgraph of $\mathbf{T}$ which is a deterministic function of both the branching process and the percolation.

Note that $\mathbf{W}$ is, considered as a measure over subgraphs of $\mathbf{T}$, invariant to the automorphisms of $\mathbf{T}$. Please reflect on this fact for a minute as it utilizes a number of features of the construction, with a particular emphasis on automorphisms of $\mathbf{T}$ which do not preserve the root $\mathbf{r}$. First, notice that it is crucial to start with $m + 1$ particles at time 0, but split to only m particles at each subsequent time. Next, note that we used the reversibility of random walk: let φ be an automorphism which sends, say, some $\mathbf{v}$ which is a child of $\mathbf{r}$ into $\mathbf{r}$. For $\mathbf{W}$ to be invariant to φ it is necessary that getting from the child to the father has the same distribution as getting from the father to the child. So our process must be time-reversible.

It is now time to fix m. We fix it such that the branching process is transient, that is, such that with probability 1 only a finite number of particles ever return to 0. With this definition of m, we get that the configuration $\mathbf{W}$ contains only finite components. Indeed, by [5] the percolation clusters are all finite. Any cluster $\mathbf{C}$ of $\mathbf{W}$ is (a subset of) all returns of the branching process to some percolation cluster, that is, a finite collection of finite sets, so $\mathbf{C}$ is finite.

At this point, we apply the mass transport principle to the tree $\mathbf{T}$ (which is of course itself the Cayley graph of a nonamenable group) and we get that the expected degree of $\mathbf{r}$ is equal to the expected average degree of the cluster of $\mathbf{r}$. But the cluster of $\mathbf{r}$ is a finite subgraph of $\mathbf{T}$ so it is a finite tree and for every finite tree the average degree is <2! Hence, we get that the expected degree of $\mathbf{r}$ in the configuration $\mathbf{W}$ is <2.

But what is the expected degree of $\mathbf{r}$ in $\mathbf{W}$? It is exactly $m \cdot \mathbb{P}(0 \leftrightarrow R(n))$! We get

$$\mathbb{P}(0 \leftrightarrow R(n)) < \frac{2}{m}$$

so Schramm's argument terminates with the observation that if $m < \lambda^{-n}$ with λ being the spectral radius of the random walk then the branching process is transient. We get that $\mathbb{P}(0 \leftrightarrow R(n)) < 2\lambda^n$ which decays exponentially as for any nonamenable Cayley graph $\lambda < 1$.

Notes.

(1) In fact, the process is transient even when $m = \lambda^{-n}$ for λ being the spectral radius. This property is called "λ-transience" and holds for all nonamenable Cayley graphs. See [17], Section II/7B.

(2) FKG implies that the function $f(n) = \mathbb{P}(0 \leftrightarrow R(n))$ is supermultiplicative, that is, $f(a+b) \geq f(a)f(b)$. Therefore, if we know that $f(n) < 2\lambda^n$ then we actually know that $f(n) \leq \lambda^n$, that is, the 2 may be dropped.

(3) A topic of some interest would be to generalize this argument from a Cayley graph to a general transitive (nonamenable) graph. One obvious point in the proof which would not apply to a general transitive graph is the application of [5]. We remark that [5] holds in larger generality than Cayley graphs, it holds for any unimodular transitive graph. See [5] or [13], Section 8.2, for details on unimodularity. Unimodularity or a similar requirement might also be necessary if one wants to generalize the claim that $\mathbf{W}$ is invariant to the automorphisms of $\mathbf{T}$.

1.2. *The triangle condition.* We say that a transitive graph (which may be amenable) satisfies the triangle condition at some p if

$$\nabla_p := \sum_{x,y} \mathbb{P}_p(0 \leftrightarrow x)\mathbb{P}_p(x \leftrightarrow y)\mathbb{P}_p(0 \leftrightarrow y) < \infty.$$

We will usually be interested in behavior at the critical p, and denote $\nabla = \nabla_{p_c}$. The triangle condition was suggested in 1984 by Aizenman and Newman [2] as a marker for "mean-field behavior," a term from statistical physics which in our context means that various quantities behave at or near p_c as they would on a regular tree. In particular, Aizenman and Newman proved that under the triangle condition $\nabla < \infty$ one has that $\mathbb{E}_p|\mathcal{C}(0)| \approx (p_c - p)^{-1}$ as p tends to p_c from below. Here, $\mathcal{C}(0) = \{x : 0 \leftrightarrow x\}$ is the cluster of 0, and $|\mathcal{C}(0)|$ is its size. See [4] for a proof that (again under $\nabla < \infty$) $\mathbb{P}_{p_c}(|\mathcal{C}(0)| > n) \approx n^{-1/2}$ and $\mathbb{P}_p(|\mathcal{C}(0)| = \infty) \approx (p - p_c)^+$. See [14] for gap exponents and [11] for intrinsic (a.k.a. "chemical") exponents and the behavior of random walk on large critical clusters. In short, under the triangle condition we have a very fine picture of the behavior at and near criticality.

It is conjectured that the triangle condition holds in great generality. A "folk" conjecture suggests that it holds for every transitive graph for which the random walk triangle condition $\sum G(0,x)G(x,y)G(0,y) < \infty$ holds, where G is the random walk Green function. I was not able to find a reference for this precise formulation but a weaker one is in [15], Conjecture 1.2. Progress on this has been slow. The most spectacular result is that the triangle condition holds for the Cayley graph of $\mathbb{Z}^d$ with $d > 6$ if one takes sufficiently many generators; or with the standard generators $\pm e_i$ if d is sufficiently large. This was achieved by Hara and Slade [7] using a technology known as "lace expansion." See [8] for generalizations to long-range models. The lace expansion is a perturbative technique and generally requires the number of generators be large. A different artifact of the perturbative nature is that it seems that lace expansion is not suitable to show that $\nabla < \infty$ unless in fact ∇ is quite close to 1 ($\nabla \geq 1$ always due to the term $x = y = 0$).

Going beyond $\mathbb{Z}^d$ there are two papers I am aware of that establish the triangle condition. The first, [18], establishes it for $T \times \mathbb{Z}$ where T is a regular tree with degree ≥ 5. The proof, very roughly, utilizes the fact that for a tree of degree d, $p_c = 1/(d-1)$ but $\nabla_p < \infty$ for all $p < 1/\sqrt{d-1}$. This allows to make a relatively rough estimate of ∇ by path counting. A far more general result was achieved in [15] which showed (among other things) that for any nonamenable group, if one takes sufficiently many generators then the resulting Cayley graph satisfies the triangle condition. Both results are significantly easier than using lace expansion.

We may now state our result.

THEOREM 1. *Let T be a regular tree of degree ≥ 3. Then the product graph $T \times T$ satisfies the triangle condition at p_c.*

One may of course wonder whether one can achieve this result from the "highly nonamenable" condition of [15], but an inspection shows that this would require T to have degree ≥ 7. Also, the approach taken here can be used also for "stretched" trees, namely, suppose one replaces every edge with a path of length 100. Naturally the resulting graph is no longer transitive (the degree of vertices in the product of two stretched trees may be $2d$, $d + 2$ or 4), but it is quasi-transitive, that is, its group of automorphisms acts with finite orbits, and the argument presented below will work, mutatis mutandis. Such a graph can have arbitrarily low Cheeger constant. Similarly, it may have ∇ arbitrarily large. It is probably also worth comparing to [16], which shows for *planar* nonamenable transitive graphs, not quite the triangle condition, but many mean-field exponents. This result is also nonperturbative but relies critically on the planarity. It is easy to check that $T \times T$ is not planar except when both have degree 2.

The important property of $T \times T$ which is used is the large number of symmetries. Examine the sphere $\{x \in T \times T : |x| = r\}$. Clearly, $\mathbb{P}(0 \leftrightarrow x)$ depends only on $|x_1|$ and $|x_2|$ where x_i are the projections of x to the two trees. Thus, the sphere breaks down into $r + 1$ of classes each of which has exponentially many vertices. This is the crucial property. Thus, the argument works for a product of two trees with different degrees, or for products of three trees or more. It is probably possible to formulate the result abstractly in terms of symmetries of the graph, but I did not have a good formulation or a second interesting example, so we will restrict ourselves to the simplest nontrivial example: $T \times T$. In addition, let us remark that I could not make the argument work for $T \times \mathbb{Z}$. Even though most vertices do have many "clones," some (namely, those on the same copy of $\mathbb{Z}$ as 0) have only one clone, and this is enough to break the argument in its current form. We will return to this topic in Section 1.5.

1.3. *What has Schramm's lemma to do with the triangle condition*? Let us go back to the proof of Schramm's lemma. As already remarked, the random walk is just some way to average the connection probability. One may take any averaging method as long as it is time-reversible. Taking a 3-regular tree T as an example, one may replace the random walk of length n in Schramm's original argument with simply choosing a random element of distance n from the root. One gets that

$$\mathbb{P}(0 \leftrightarrow R) < \frac{2}{m},$$

where R is a random element of distance n from the root, and m needs to satisfy that the corresponding branching process is transient. Of course, by the symmetries of the tree all elements of distance n from the root have the same connection probability so the "averaging" performed by taking a random element has no effect.

We will calculate the largest m one may take below (it is the claim in the proof of Lemma 5), but we note for now that the number of particles which return to 0 after two steps of the branching process is $m^2/2^n$. Therefore, it is reasonable to assume that $m \simeq \sqrt{2^n}$ is the threshold for the branching process to be transient, so we should have $\mathbb{P}(0 \leftrightarrow x) \lesssim 2^{-|x|/2}$ and indeed a slightly more precise calculation shows (still for the tree) that

$$\mathbb{P}(0 \leftrightarrow x) < C|x|2^{-|x|/2}.$$

This can now be summed explicitly to give

$$\sum_{|x|,|y| \le r} \mathbb{P}(0 \leftrightarrow x)\mathbb{P}(x \leftrightarrow y)\mathbb{P}(0 \leftrightarrow y) \le Cr^8. \tag{2}$$

The 8 is of course not important—what is important is that the sum grows only polynomially, even though the ball of radius r grows exponentially. In other words, Schramm's lemma "almost" gives the triangle condition. This is the crucial observation on which we rely.

1.4. *Proof sketch.* To show the triangle condition, it is enough to show that $\mathbb{P}(0 \leftrightarrow x) \le C2^{-|x|(1/2+\varepsilon)}$ as then one can sum these explicitly. By the symmetries of our graph, if we show that

$$\mathbb{E}(\{x : 0 \leftrightarrow x, |x| \le r\}) \le C2^{r(1/2-\varepsilon)}$$

we will be done. The set $\{x : 0 \leftrightarrow x, |x| \le r\}$ is known as the "ball in the extrinsic metric"—meaning that we measure the distance to x by the distance inherited from the surrounding graph, $|x|$.

If there is anything metamathematical to be learned from comparing Section 3.2 in [11] to [12], it is that it is easier to work with the intrinsic distance. In other words, rather than looking at $|x|$, we should look at $d_{\text{int}}(x, y)$, the length of the shortest path of open edges between x and y. This idea has a checkered past—in two dimensions the behavior of $d_{\text{int}}(x, y)$ is still wide open—but in mean-field setting it has proved to be a useful tool. Denote $x \overset{r}{\leftrightarrow} y$ as a short for $d_{\text{int}}(x, y) \le r$. Denote by $B_{\text{int}}(r)$ the ball in the intrinsic distance, that is, the random set defined by $B_{\text{int}}(r) = \{x : 0 \overset{r}{\leftrightarrow} x\}$. Denote $G(r) = \mathbb{E}|B_{\text{int}}(r)|$. This will be our main object of study, and we will get better and better estimates for it. As before, assume for the sake of this sketch that $d = 3$.

Step 1. In any transitive graph, $G(r) = e^{o(r)}$. This is due to Russo's formula, since the number of pivotal edges for the event $0 \overset{r}{\leftrightarrow} x$ is always $\le r$—only edges on the path can be pivotal! See Lemma 1, page 1873.

Step 2. We apply Schramm's lemma to $T \times T$ as explained in Section 1.3, and get that $\mathbb{P}(0 \leftrightarrow x) \le C|x|^2 2^{-|x|/2}$. See Lemma 5, page 1878.

Step 3. We now repeat the argument of [12], Theorem 1.2(i). The claim there was that under the triangle condition $G(r) \le Cr$. The skeleton of the argument is as follows. It is enough to prove that

$$G(2r) \ge c\frac{G(r)^2}{r}, \tag{3}$$

because if $G(r) > (2/c)r$ then it starts growing exponentially, contradicting the information we gathered at step 1. Denote for the purpose of this sketch by w a vertex of the graph quite close to 0. w is the "opening," a standard step in any use of the triangle condition. Examine the following quantity

$$\mathbb{E}\bigl|\{(x, y) : 0 \overset{r}{\leftrightarrow} x, xw \overset{r}{\leftrightarrow} y, x \not\leftrightarrow xw\}\bigr|,$$

where xw stands for the product in the relevant group. You should think about x as being "roughly pivotal" for the event $0 \leftrightarrow y$, with the measure of roughness related to w. A calculation using the Aizenman and Newman "off method" which may be found in [12], Lemma 3.2, or in Lemma 7, page 1882 below, gives

$$\begin{aligned}(4)\qquad &\mathbb{E}\left|\{(x,y):0 \overset{r}{\leftrightarrow} x, xw \overset{r}{\leftrightarrow} y, x \not\leftrightarrow xw\}\right| \\ &\quad\ge G(r)^2\Bigl(1-\sum_{u,v}\mathbb{P}(0 \overset{r}{\leftrightarrow} u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(v \overset{r}{\leftrightarrow} w)\Bigr).\end{aligned}$$

The sum inside the parenthesis is not *quite* the triangle sum because $0 \neq w$. This expression (without the r's) is known as the *open triangle sum* and it is known that if the triangle condition holds (recall that we are still in the setting of [11] where the triangle condition is assumed) then the open triangle sum tends to 0 as the point w is sent to infinity ([4], Lemma 2.1 for $\mathbb{Z}^d$ and [10] for a general transitive graphs). Taking it to be $\le\frac{1}{2}$ gives

$$\mathbb{E}\left|\{(x,y):0 \overset{r}{\leftrightarrow} x, xw \overset{r}{\leftrightarrow} y, x \not\leftrightarrow xw\}\right| \ge \tfrac{1}{2}G(r)^2.$$

At this point, a simple modification argument that allows to connect x and xw while paying only a constant. The modification puts y into $\mathcal{C}(0)$ and also turns x from "roughly pivotal" to being properly pivotal. This proves (3)—the modification costs only a constant, but the counting over x costs another r which explains the $1/r$ factor in (3). This finishes the proof in [11].

How does all this apply in our case? Schramm's lemma gives only that the triangle condition grows moderately, not that it is finite. This means simply that it is necessary to open the triangle wider. A calculation (done in Lemma 6, page 1881) shows that it is enough to take w in distance $\approx \log r$. This is good, but not as good as it sounds because once one tries to apply the modification argument one loses the exponent of the distance namely an r^C factor and gets instead of (3)

$$G(2r) \ge \frac{G(r)^2}{r^C},$$

which only shows that $G(r) \le r^C$—a vast improvement over $e^{o(r)}$ but still not what we want. See Lemmas 8 and 9, starting from page 1884, for the modification argument.

Step 4. Because of the symmetries of the graph, our newly acquired knowledge $G(r) \le r^C$ allows to get much better estimates for $0 \overset{r}{\leftrightarrow} x$ namely because x has $2^{|x|}$ clones we get

$$\mathbb{P}(0 \overset{r}{\leftrightarrow} x) \le r^C 2^{-|x|},$$

which is better than the $2^{-|x|/2}$ given by Schramm's lemma whenever $|x| \ge C\log r$. This allows to separate 0 and w by only $\approx \log\log r$ yielding a final $G(r) \le r(\log r)^C$. See Lemmas 10 and 11, page 1886.

Step 5. We now find ourselves in a rather ridiculous situation. We have a very good estimate for the ball in the intrinsic distance—$\mathbb{E}|B_{\text{int}}(r)| \le r(\log r)^C$—but still absolutely no estimate for the extrinsic ball, for all we know we might have $\mathbb{E}|\{x : 0 \leftrightarrow x \text{ and } |x| \le r\}| \simeq 2^{r/2}$, which would of course imply that it intersects an enormous intrinsic ball, all tightly curled up. We need to contradict this possibility and we use the property that in a nonamenable graph, for any x, if one examines the ball of radius r around x then the vast majority of it is further from 0 than x. Using the fact that $B_{\text{int}}(x, r)$ is quite small and the symmetries of the graph, we get that we can show that the parts of $B_{\text{int}}(x, r)$ that go "back," that is, are closer to 0 than x are dominated by a subcritical process. We get that the process is "ballistic" in the sense that points x with $d_{\text{int}}(0, x) = r$ have also $|x| \approx r$. This shows that

$$\mathbb{E}\big|\{x : 0 \leftrightarrow x \text{ and } |x| \le r\}\big| \le Cr^3$$

(the 3 is just an artifact of sloppiness) and by the symmetries of the graph one last time

$$\mathbb{P}(0 \leftrightarrow x) \le C|x|^3 2^{-|x|},$$

which shows the triangle condition by a direct calculation. See Lemmas 12 and 13, starting page 1888.

1.5. *The case of $T \times \mathbb{Z}$*. Where does all this break for $T \times \mathbb{Z}$? We used the existence of clones in every step. Hence, this argument cannot give any estimate for $\mathbb{P}(0 \leftrightarrow (0, n))$ where $(0, n)$ stands for a vertex in the same copy of $\mathbb{Z}$ as 0. This does not seem like a big deal because there are not many of those. But in fact, it breaks the argument at step 4. In other words, you can get that $G(r) \le r^C$ but cannot progress beyond that, which breaks the final step, the subcriticality of the backward process.

There is a different way to view this. Let our percolation have different p in the different coordinates. We get a two-parameter family of processes with a critical curve separating the regime of only finite clusters and the regime of infinite clusters. See Figure 1. We immediately note the following difference between the $T \times T$ and $T \times \mathbb{Z}$ case. In the $T \times T$ case, ∇ is bounded uniformly on the critical curve. On $T \times \mathbb{Z}$ it diverges as you approach the point $(0, 1)$. Similarly, $\mathbb{P}(0 \leftrightarrow (0, n))$ is not bounded away from 1 uniformly on the critical curve, it converges to 1 as you approach $(0, 1)$. I do not claim that this is a significant hurdle, just note that all ideas in this paper (including Schramm's lemma) work uniformly on the entire critical line so perhaps a new idea is needed.

1.6. *Notation and conventions*. All graphs in this paper will have one vertex denoted by 0. For Cayley graphs, 0 will be the identity element. We denote by $d(x, y)$ the graph distance of x from y, that is, the the number of edges in the shortest path between x and y. Denote also $|x| = d(x, 0)$ and balls by $B(x, r) =$

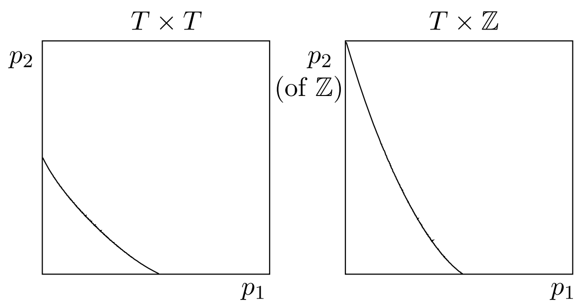

FIG. 1. *The critical curves for $T \times T$ and $T \times \mathbb{Z}$. Values were calculated by an invasion percolation algorithm run until cluster size reached* 10^6.

$\{y : d(x, y) \le r\}$ and $B(r) = B(0, r)$. For a subset of vertices A, we denote by ∂A the set of all edges with one vertex in A and one outside A.

For percolation, we denote by $d_{\text{int}}(x, y)$ the length of the shortest open path between x and y, or ∞ if $x \not\leftrightarrow y$. We denote by $x \overset{r}{\leftrightarrow} y$ the event $d_{\text{int}}(x, y) \le r$. We denote $B_{\text{int}}(x, r) = \{y : x \overset{r}{\leftrightarrow} y\}$ and $B_{\text{int}}(r) = B_{\text{int}}(0, r)$. Be careful not to confuse $B(r)$ (which is a deterministic quantity) and $B_{\text{int}}(r)$ which is a random variable. We denote $G(r) = \mathbb{E}|B_{\text{int}}(r)|$. Denote also the triangle sum

$$\nabla = \sum_{u,v} \mathbb{P}(0 \leftrightarrow u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(0 \leftrightarrow v)$$

and the restricted open triangle sum with opening w and distance r

$$\nabla(w; r) = \sum_{u,v} \mathbb{P}\big(0 \overset{r}{\leftrightarrow} u\big)\mathbb{P}(u \leftrightarrow v)\mathbb{P}\big(v \overset{r}{\leftrightarrow} w\big).$$

More standard percolation notations used are $\mathcal{C}(x) = \{y : y \leftrightarrow x\}$, and $A \circ B$ for the event that A and B "occur disjointly," see [6], Section 2.3, for the notation and for the van den Berg–Kesten inequality $\mathbb{P}(A \circ B) \le \mathbb{P}(A)\mathbb{P}(B)$. We shall denote the van den Berg–Kesten inequality by BK for short. Denote Harris' inequality ([6], Section 2.2) by FKG. For all these notations, unless p is specified explicitly it is taken to be the p_c of the relevant graph. When we want to examine a different p we will use the notations $\mathbb{P}_p$ and $\mathbb{E}_p$ for the probability and the expectation with respect to p.

We denote by T a regular tree of degree $d \ge 3$. For a vertex $x \in T \times T$, denote by x_1 and x_2 its two coordinates. It is easy to verify that $|x| = |x_1| + |x_2|$. We denote by Γ the group whose Cayley graph is $T \times T$ (see Lemma 4).

Bold letters will be used for the high-degree tree $\mathbf{T}$ that appears in the proof of Schramm's lemma (both the sketch in Section 1.1 above and the proof of Lemmas 3 and 5 below) and for vertices and subgraphs of it. Be careful not to confuse $\mathbf{T}$, which is a tree of degree m and is just an auxiliary object, with T which is a tree of degree d and the principle object of investigation.

By c and C, we will denote constants which depend only on our graph G (usually this is $T \times T$ so they only depend on the degree of T, but in Section 2 the results are general). c will denote constants which are "small enough" and C constants which are "big enough." C and c may refer to different constants in different formulas and even within the same formula. We will sometimes number them for clarity. A notation like C_{187} is specific to the lemma in which it appears. When a probability decays exponentially in some parameter n, will usually denote it by $<2e^{-cn}$.

The notation $A \approx B$ means that some constants c and C exist such that $cB \le A \le CB$. The notations $\simeq$ and $\lesssim$ mean nothing in particular. We only use them when we want to indicate that two quantities are heuristically similar, but do not want to indicate in which sense exactly. The notation $X \sim Y$ for two random variables means that they have the same distribution. For a real number x, $\lceil x \rceil$ will denote the smallest integer $\ge x$.

2. Preliminaries for transitive graphs.

THEOREM 2 (Aizenman and Barsky). *For any vertex-transitive graph G and any $p < p_c(G)$, we have $\mathbb{E}_p|\mathcal{C}(0)| < \infty$.*

Aizenman and Barsky [1] formulated their result only for $\mathbb{Z}^d$, but it is well known that it holds for any transitive graph. For example, it is mentioned in passing in [15]. A proof may be found in [3] or in [9], Appendix A.

LEMMA 1. *For any vertex transitive graph G, $G(r) = e^{o(r)}$.*

[Recall that $G(r) = \mathbb{E}_{p_c}|B_{\text{int}}(r)|$.]

PROOF OF LEMMA 1. Fix some $x \in B(r)$ and examine the event $0 \overset{r}{\leftrightarrow} x$. By Russo's formula ([6], Section 2.4), for any $0 < p < 1$,

$$\frac{d}{dp}\mathbb{P}_p(0 \overset{r}{\leftrightarrow} x) = \frac{1}{p}\mathbb{E}_p(|\{\text{open pivotal edges}\}|).$$

We are allowed to use Russo's formula, since this event is determined by a finite number edges, namely those of $B(r)$. For any configuration where $0 \overset{r}{\leftrightarrow} x$, the number of pivotal edges is $\le r$ since clearly any edge off the path between 0 and x is not pivotal. Hence, we get

$$\frac{d}{dp}\mathbb{P}_p(0 \overset{r}{\leftrightarrow} x) \le \frac{r}{p}\mathbb{P}_p(0 \overset{r}{\leftrightarrow} x).$$

Summing over $x \in B(r)$, we get

$$\frac{d}{dp}\mathbb{E}_p|B_{\text{int}}(r)| \le \frac{r}{p}\mathbb{E}_p|B_{\text{int}}(r)|$$

or

$$\frac{d}{dp}\log \mathbb{E}_p|B_{\text{int}}(r)| \le \frac{r}{p}.$$

Assume by contradiction that $G(r) \ge e^{cr}$ for some $c > 0$ and infinitely many r's. We get for $p < p_c$

$$\log \mathbb{E}_p|B_{\text{int}}(r)| \ge cr - \frac{r}{p}(p_c - p)$$

so for $p \in (p_c/(1+c), p_c)$ we have $\mathbb{E}_p|B_{\text{int}}(r)| \to \infty$ as $r \to \infty$. This contradicts the theorem of Aizenman and Barsky. $\square$

LEMMA 2. *For any transitive graph, any $r > 0$ and any $\lambda > 0$,*

$$\mathbb{P}\big(|B_{\text{int}}(r)| > \lambda G(r)^2\big) \le 2e^{-c\lambda},$$

where c is an absolute constant.

PROOF. This is a standard corollary of Aizenman and Newman's diagrammatic bounds. Refer to [6], Section 6.3, for a complete treatment. The classic picture is as follows. We wish to calculate the nth moment of $|B_{\text{int}}(r)|$. For this, we note that if $0 \leftrightarrow x_i$ for $i = 1, \dots, n$ then there exist $y_1, \dots, y_{n-1}$ and a tree describing the connection scheme U with $n+1$ leaves (corresponding to the vertices $0, x_1, \dots, x_n$) and $n-1$ inner points (corresponding to the vertices $y_1, \dots, y_{n-1}$) all of which are of degree 3 such that for every edge of U, the corresponding vertices are connected by an open path, and all such paths are edge-disjoint. For convenience, denote $(z_1, \dots, z_{2n}) = (0, x_1, \dots, x_n, y_1, \dots, y_{n-1})$ and let the vertices of the tree U be the numbers $\{1, \dots, 2n\}$ correspondingly.

Now, in our case the paths from 0 to x_i are constrained to be of length $\le r$ and this constraint is carried over to all internal paths of our tree. Hence,

$$\begin{aligned}
&\mathbb{E}\big|\{(x_1, \dots, x_n) : 0 \overset{r}{\leftrightarrow} x_i \ \forall i\}\big| \\
&\qquad \le \sum_U \sum_{z_2,\dots,z_{2n}} \mathbb{P}\big(z_i \overset{r}{\leftrightarrow} z_j \text{ for all edges } (i,j) \text{ of } U, \text{ disjointly}\big) \\
\text{By BK} \quad &\qquad \le \sum_U \sum_{z_2,\dots,z_{2n}} \prod_{(i,j)\in U} \mathbb{P}\big(z_i \overset{r}{\leftrightarrow} z_j\big) \\
&\qquad = \sum_U G(r)^{2n-1} = 1 \cdot 3 \cdot 5 \cdots (2n-3) \cdot G(r)^{2n-1} \\
&\qquad \le (2nG(r)^2)^n.
\end{aligned}$$

The proof is now complete as

$$\begin{aligned}
\mathbb{P}\big(|B_{\text{int}}(r)| > \lambda G(r)^2\big) &= \mathbb{P}\big(|B_{\text{int}}(r)|^n > \lambda^n G(r)^{2n}\big) \\
&\le \frac{\mathbb{E}(|B_{\text{int}}(r)|^n)}{\lambda^n G(r)^{2n}} \le \left(\frac{2n}{\lambda}\right)^n
\end{aligned}$$

and setting $n = \lfloor \lambda/4 \rfloor$ we are done. $\square$

The last general claim we wish to demonstrate before moving on to the product of two trees is the invariance step in Schramm's lemma. Recall the sketch of the lemma and also the discussion in Section 1.3. The claim there is that for any branching process with a time-reversible step, the resulting $\mathbf{W}$ is invariant to the automorphisms of the tree $\mathbf{T}$. We will now prove this fact, and we formulate it in sufficient generality so that it can be used both for the original version of Schramm's lemma and for our purposes.

This is the only place in the paper where it matters which Cayley graph we are talking about, so let us fix that all Cayley graphs are *right* Cayley graphs, that is, if Γ is a finitely generated group and S a set of generators then the edges of the Cayley graphs are $\{(g, gs) : g \in \Gamma, s \in S\}$.

DEFINITION 1. Let Γ be a finitely generated group and let S be a set of generators. Let μ be some discrete measure on Γ with $\mu(x) = \mu(x^{-1})$. Let $m \geq 1$ be some integer. Let $0 < p < 1$.

- Define $\mathbf{T}$ to be a regular tree of degree $m + 1$. We will use $\mathbf{T}$ also to denote the set of vertices of $\mathbf{T}$, and the edges will be denoted by $E(\mathbf{T})$. Fix one element of $\mathbf{T}$, call it the root and denote it by $\mathbf{r}$.
- Define $\pi : \mathbf{T} \to \Gamma$ which is a random map ("the locations of the particles"). For $\mathbf{r}$ the root of the tree $\mathbf{T}$, we define $\pi(\mathbf{r}) = 0$ where 0 is the identity element of Γ. We continue inductively. Assume $\pi(\mathbf{v})$ is already defined. For every child $\mathbf{w}$ of $\mathbf{v}$, we define $\pi(\mathbf{w}) = \pi(\mathbf{v})X_{\mathbf{v},\mathbf{w}}$ where $X_{\mathbf{v},\mathbf{w}}$ are i.i.d. random variables distributed like μ.
- Finally, define Schramm's process $\mathbf{W} = \mathbf{W}(\Gamma, S, \mu, m, p)$ to be a random subset of the edges of $\mathbf{T}$ defined by

$$(\mathbf{u}, \mathbf{v}) \in \mathbf{W} \quad \iff \quad \pi(\mathbf{u}) \leftrightarrow \pi(\mathbf{v}) \qquad \forall (\mathbf{u}, \mathbf{v}) \in E(\mathbf{T}),$$

  where $g \leftrightarrow h$ denotes that g and h are connected in a p-percolation process independent of the $X_{\mathbf{u},\mathbf{v}}$ on the (right) Cayley graph of Γ with respect to the set of generators S.

LEMMA 3. *For any Γ, S, μ, m and p as above, Schramm's process is invariant to the automorphisms of* $\mathbf{T}$.

PROOF. Denote the root of $\mathbf{T}$ by $\mathbf{r}$ and its children by $\mathbf{v}_1, \ldots, \mathbf{v}_{m+1}$. Let $\varphi : \mathbf{T} \to \mathbf{T}$ be the automorphism that takes $\mathbf{v}_1 \mapsto \mathbf{r}$ and $\mathbf{r} \mapsto \mathbf{v}_{m+1}$ but otherwise preserves the order among children: the m children of $\mathbf{v}_1$ are mapped to $\mathbf{v}_1, \ldots, \mathbf{v}_m$ in order, $\mathbf{v}_2, \ldots, \mathbf{v}_{m+1}$ are mapped to the m children of $\mathbf{v}_{m+1}$ in order etc. For every permutation of $m + 1$ elements σ, let $\psi_\sigma : \mathbf{T} \to \mathbf{T}$ be the automorphism permuting the children of $\mathbf{r}$ according to σ but otherwise preserving the order. Let H be the group of automorphisms of $\mathbf{T}$ generated by φ and all ψ_σ.

It is straightforward to verify that for any automorphism θ of $\mathbf{T}$ and every r there exists an $\eta \in H$ which is identical to θ on the entire ball of radius r in $\mathbf{T}$. In other words, H is dense in the compact-open topology. This implies that it is enough to show that the distribution of $\mathbf{W}$ is invariant to the action of H and hence it is enough to show that it is invariant to the action of φ and ψ_σ. Verifying ψ_σ is immediate, so we are left with showing that $\varphi\mathbf{W} \sim \mathbf{W}$ where we define $(\varphi\mathbf{W})(e) = \mathbf{W}(\varphi^{-1}(e))$. It will be more convenient to verify that $\varphi^{-1}\mathbf{W} \sim \mathbf{W}$ and we will do so.

Write our probability space as $\Omega_1 \times \Omega_2$ where Ω_1 is the probability space of the branching random walk and Ω_2 is the probability space of the percolation. Further, write Ω_1 as $\Gamma^{E(\mathbf{T})}$ where $E(\mathbf{T})$ is the set of edges of $\mathbf{T}$, with the measure being the product measure $\mu^{E(\mathbf{T})}$. Any automorphism φ of $\mathbf{T}$ induces a measure preserving map $\alpha : \Omega_1 \to \Omega_1$ by

$$\alpha(\omega)(\mathbf{x}, \mathbf{y}) = \begin{cases} \omega(\varphi(\mathbf{x}), \varphi(\mathbf{y})), & \varphi \text{ preserves the orientation of } (\mathbf{x}, \mathbf{y}), \\ \omega(\varphi(\mathbf{y}), \varphi(\mathbf{x}))^{-1}, & \text{otherwise.} \end{cases}$$

The $^{-1}$ on the bottom clause stands for inversion in the group Γ. Also we need to explain what does it mean that "φ preserves the orientation of $(\mathbf{x}, \mathbf{y})$"—this means, when $\mathbf{x}$ is the father of $\mathbf{y}$, that $\varphi(\mathbf{x})$ is the father of $\varphi(\mathbf{y})$. Of course, our φ only reverses the orientation of one edge, $(\mathbf{r}, \mathbf{v}_1)$. α is measure preserving because the measure μ is invariant to the operation $^{-1}$. Slightly abusing notations we consider α also as a map $\Omega_1 \times \Omega_2 \to \Omega_1 \times \Omega_2$ acting only on the first coordinate.

We now define a second measure preserving map $\beta : \Omega_1 \times \Omega_2 \to \Omega_1 \times \Omega_2$ as follows: for any $\omega \in \Omega_1$ we let $f(\omega)$ be an automorphism of the Cayley graph of Γ given by

$$f(\omega)v = \omega(\mathbf{r}, \mathbf{v}_{m+1})^{-1} v,$$

where the product is in the group Γ—again we consider ω as an element of $\Gamma^{E(\mathbf{T})}$ so $\omega(\mathbf{r}, \mathbf{v}_{m+1})$ is simply the position of the $(m+1)$st child of the original particle. We also consider f as acting on Ω_2 (which is just the product space $\{0, 1\}^{E(\Gamma)}$) by $f\omega_2(v, w) = \omega_2(f^{-1}(v), f^{-1}(w))$. We now define $\beta(\omega_1, \omega_2) = (\omega_1, f(\omega_1)\omega_2)$. Since $f(\omega_1)$ is measure preserving for any ω_1, we get that β is measure preserving by Fubini's theorem.

The lemma is now finished because applying the measure preserving transformation $\alpha \circ \beta$ to the probability space is the same as applying φ^{-1} to $\mathbf{W}$. Let us verify this formally. We consider $\mathbf{W}$ as a function $\Omega_1 \times \Omega_2 \to \{0, 1\}^{E(\mathbf{T})}$ defined by "$\mathbf{W}(\omega_1, \omega_2)(\mathbf{x}, \mathbf{y}) = 1$ if $\prod \omega_1(\mathbf{x}_i, \mathbf{x}_{i+1})$ is connected to $\prod \omega_1(\mathbf{y}_i, \mathbf{y}_{i+1})$ in the configuration ω_2," where for an element $\mathbf{x} \in \mathbf{T}$ we define $\mathbf{x}_0, \mathbf{x}_1, \ldots$ to be the elements of the tree on the branch from $\mathbf{r} = \mathbf{x}_0$ to $\mathbf{x}$, and where the $\prod$ is in the group Γ and is taken left-to-right, that is, $\omega(\mathbf{x}_0, \mathbf{x}_1)\omega(\mathbf{x}_1, \mathbf{x}_2) \cdots$. We wish to show that

$\mathbf{W}(\alpha(\beta(\omega))) = \varphi^{-1}\mathbf{W}(\omega)$. But $\mathbf{W}(\alpha(\beta(\omega))(\mathbf{x},\mathbf{y}) = 1$ if $\prod \alpha(\beta(\omega))_1(\mathbf{x}_i,\mathbf{x}_{i+1})$ is connected to $\prod \alpha(\beta(\omega))_1(\mathbf{y}_i,\mathbf{y}_{i+1})$ in $\alpha(\beta(\omega))_2$. Now,

$$\begin{aligned}
&\alpha(\beta(\omega))_1(\mathbf{x},\mathbf{y})\\
&\quad = \begin{cases} \beta(\omega)_1(\varphi(\mathbf{x}),\varphi(\mathbf{y})), & \varphi \text{ preserves the orientation of } (\mathbf{x},\mathbf{y}),\\ \beta(\omega)_1(\varphi(\mathbf{y}),\varphi(\mathbf{x}))^{-1}, & \text{otherwise,}\end{cases}\\
&\quad = \begin{cases} \omega_1(\varphi(\mathbf{x}),\varphi(\mathbf{y})), & \varphi \text{ preserves the orientation of } (\mathbf{x},\mathbf{y})\\ \omega_1(\varphi(\mathbf{y}),\varphi(\mathbf{x}))^{-1}, & \text{otherwise,}\end{cases}
\end{aligned}$$

so

$$\begin{aligned}
&\prod \alpha(\beta(\omega))_1(\mathbf{x}_i,\mathbf{x}_{i+1})\\
&\qquad = \begin{cases} \omega_1(\mathbf{r},\mathbf{v}_{m+1})^{-1}\displaystyle\prod_{i=1}\omega_1(\varphi(\mathbf{x}_i),\varphi(\mathbf{x}_{i+1})), & \mathbf{x}_1 = \mathbf{v}_1,\\ \displaystyle\prod_{i=0}\omega_1(\varphi(\mathbf{x}_i),\varphi(\mathbf{x}_{i+1})), & \text{otherwise.}\end{cases}
\end{aligned}$$

Comparing to the branch from $\mathbf{r}$ to $\varphi(\mathbf{x})$, we get

$$\prod \alpha(\beta(\omega))_1(\mathbf{x}_i,\mathbf{x}_{i+1}) = \omega_1(\mathbf{r},\mathbf{v}_{m+1})^{-1}\prod \omega_1(\varphi(\mathbf{x})_i,\varphi(\mathbf{x})_{i+1})$$

in both cases.

For the percolation configuration we have a similar calculation,

$$\begin{aligned}
\alpha(\beta(\omega))_2(v,w) &= \beta(\omega)_2(v,w) = (f(\omega_1)\omega_2)(v,w)\\
&= \omega_2(f(\omega_1)^{-1}v, f(\omega_1)^{-1}w)\\
&= \omega_2(\omega_1(\mathbf{r},\mathbf{v}_{m+1})v, \omega_1(\mathbf{r},\mathbf{v}_{m+1})w).
\end{aligned}$$

We put the formulas for Ω_1 and Ω_2 together and get that $\mathbf{W}(\alpha(\beta(\omega)))(\mathbf{x},\mathbf{y}) = 1$ if and only if $\omega_1(\mathbf{r},\mathbf{v}_{m+1})^{-1}\prod \omega_1(\varphi(\mathbf{x})_i,\varphi(\mathbf{x})_{i+1})$ and $\omega_1(\mathbf{r},\mathbf{v}_{m+1})^{-1}\prod \omega_1(\varphi(\mathbf{y})_i, \varphi(\mathbf{y})_{i+1})$ are connected in the configuration $\omega_2(\omega_1(\mathbf{r},\mathbf{v}_{m+1})v, \omega_1(\mathbf{r},\mathbf{v}_{m+1})w)$. The terms $\omega_1(\mathbf{r},\mathbf{v}_{m+1})$ now cancel (recall the a right Cayley graph is invariant to left translations) and we get that this happens if and only if $\prod \omega_1(\varphi(\mathbf{x})_i,\varphi(\mathbf{x})_{i+1})$ is connected to $\prod \omega_1(\varphi(\mathbf{y})_i,\varphi(\mathbf{y})_{i+1})$ in ω_2 which is exactly $\mathbf{W}(\omega)(\varphi(\mathbf{x}),\varphi(\mathbf{y}))$. In other words

$$\mathbf{W}(\alpha(\beta(\omega))(\mathbf{x},\mathbf{y}) = \mathbf{W}(\omega)(\varphi(\mathbf{x}),\varphi(\mathbf{y})) \qquad \text{for all } \mathbf{x} \text{ and } \mathbf{y},$$

which is exactly $\mathbf{W}(\alpha(\beta(\omega))) = \varphi^{-1}\mathbf{W}(\omega)$ which shows that the distribution of $\mathbf{W}$ is invariant to φ^{-1} and hence to φ. As explained, this shows that $\mathbf{W}$ is invariant to a group of automorphisms dense in the compact-open topology, hence to all automorphisms, proving the lemma. $\square$

3. The product of two trees.

LEMMA 4. *At p_c there is no infinite cluster for $T \times T$.*

PROOF. By [5], every nonamenable Cayley graph satisfies this property. Hence, we need only show that $T \times T$ is a nonamenable Cayley graph. This however is easy. Any tree of degree d is the Cayley graph of the free product

$$G_d := \underbrace{\mathbb{Z}/2\mathbb{Z} * \cdots * \mathbb{Z}/2\mathbb{Z}}_{d \text{ times}}$$

with the natural generators (namely one from every copy of $\mathbb{Z}/2\mathbb{Z}$). Of course if d is even then one can simply take a free group with $\frac{1}{2}d$ generators. The product is thus the Cayley graph of the group $G_d \times G_d$. The claim of nonamenability is just as easy. We follow [5] and say that a graph is nonamenable if there exists some c such that for all finite A, $|\partial A| \ge c|A|$. Let therefore $A \subset T \times T$ be any finite set. For any $x \in T$, denote by A_x the slice $\{y : (y, x) \in A\}$. Then, when $A_x \ne \varnothing$,

$$|\partial A_x| = (d-2)|A_x| + 2, \tag{5}$$

where ∂ is the edge boundary in the tree T. Equation (5) is a well-known property of regular trees and may be readily proved by induction on $|A_x|$. Summing over x, we get

$$|\partial A| \ge \sum_x |\partial A_x| > \sum_x (d-2)|A_x| = (d-2)|A|$$

as needed. □

DEFINITION 2. We denote by Γ the group whose Cayley graph is $T \times T$, namely $G_d \times G_d$ from the previous lemma.

The next lemma is the adaptation of Schramm's argument to our setting.

LEMMA 5. *For any $x \in T \times T$,*

$$\mathbb{P}(0 \leftrightarrow x) \le C|x|^2(d-1)^{-|x|/2}.$$

PROOF. Let $x = (x_1, x_2)$ and let $k_i = |x_i|$ be the distances of the coordinates of x from the root in the two trees. For any $y = (y_1, y_2) \in T \times T$ denote

$$L(y) = \{(z_1, z_2) : d(z_1, y_1) = k_1, d(z_2, y_2) = k_2\}.$$

Clearly $|L(y)| = (d-1)^{|x|}$. Recall the definition of Schramm's process (Definition 1 on page 1875). First, we need a group and we take the group to be our Γ, the group whose Cayley graph is $T \times T$. The next element in Schramm's process is a branching process with a time-reversible step μ. We take

$$\mu(y) = \begin{cases} \dfrac{1}{|L(0)|}, & y \in L(0), \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, $\mu(y) = \mu(y^{-1})$, that is, is time-reversible. Finally, we need two parameters, m the branching number (which we leave unspecified for a while) and p, which we take to be $p_c(T \times T)$. Schramm's process is now a random subset $\mathbf{W}$ of the edges of a regular tree of degree $m+1$ which we denote by $\mathbf{T}$. The statement of Lemma 3 now says

$$\mathbf{W} \text{ is invariant to the automorphisms of } \mathbf{T}.$$

Examine now the parameter m—recall that Schramm's process involves a branching random walk on $T \times T$ where each particle splits into m children and then each child makes one step of μ. For every m for which the branching process is transient (i.e., with probability 1 only a finite number of particles return to any given point), then the configuration $\mathbf{W}$ contains no infinite component. This is due to Lemma 4, since the cluster of 0 is finite (with probability 1) and only a finite number of particles return to each of its points (again with probability 1). Let $\mathbf{W}$ be a subset of edges of $\mathbf{T}$, and denote by $\mathcal{C}(\mathbf{x}) = \mathcal{C}(\mathbf{x}; \mathbf{W})$ the cluster of some $\mathbf{x} \in \mathbf{T}$ in $\mathbf{W}$, that is, all $\mathbf{y}$ connected to $\mathbf{x}$ by a path of edges in $\mathbf{W}$. Denote also $\deg \mathbf{x} = |\{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in \mathbf{W}\}|$, the degree of $\mathbf{x}$ in $\mathbf{W}$. We now define

$$M(\mathbf{x}, \mathbf{y}; \mathbf{W}) = \begin{cases} \dfrac{\deg \mathbf{x}}{|\mathcal{C}(\mathbf{x})|}, & \mathbf{y} \in \mathcal{C}(\mathbf{x}) \text{ and } |\mathcal{C}(\mathbf{x})| < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, M is invariant to the automorphisms of $\mathbf{T}$ in the sense that $M(\mathbf{x}, \mathbf{y}; \mathbf{W}) = M(\varphi\mathbf{x}, \varphi\mathbf{y}; \varphi\mathbf{W})$ for any automorphism φ of $\mathbf{T}$. A simple change of variables known as the "mass transport principle," see [5], equation (2.1), shows that

$$\sum_{\mathbf{y} \in \mathbf{T}} \mathbb{E} M(\mathbf{x}, \mathbf{y}; \mathbf{W}) = \sum_{\mathbf{y} \in \mathbf{T}} \mathbb{E} M(\mathbf{y}, \mathbf{x}; \mathbf{W}) \qquad \forall \mathbf{x} \in \mathbf{T}$$

($\mathbb{E}$ here is with respect to $\mathbf{W}$ and we use the invariance of $\mathbf{W}$ too). Now, the left-hand side is obviously just $\mathbb{E}(\deg \mathbf{x} \cdot \mathbf{1}_{|\mathcal{C}(\mathbf{x})| < \infty})$ (the sum and the expectation may be exchanged since M is positive). Since our clusters are finite a.s., the left-hand side is simply $\mathbb{E}(\deg \mathbf{x})$. The right-hand side, on the other hand, is the expected average degree of $\mathcal{C}(\mathbf{x})$. However, $\mathbf{T}$ is a tree, and therefore $\mathcal{C}(\mathbf{x})$ is a finite tree and it is well known (and easy to prove inductively) that the average degree of any finite tree is <2. Hence, we get

$$\mathbb{E}(\deg \mathbf{x}) < 2.$$

What is the meaning of $\mathbb{E}\deg\mathbf{x}$? Since our branching process consists of sending $m+1$ particles to random points in $L(0)$, and all particles are identical, it means that for a random point $y \in L(0)$

$$\mathbb{P}(0 \leftrightarrow y) \le \frac{2}{m+1}. \tag{6}$$

However, the symmetries of the two trees show that for the $y \in L(0)$, the probabilities $\mathbb{P}(0 \leftrightarrow y)$ are all equal. Hence, we get (6) for *any* $y \in L(0)$. This is the crux of Schramm's argument, and we need only to calculate an m for which the branching process is still transient (ideally one might want to find the maximal such m, but the following argument is not precise).

CLAIM. *With the definitions above, for $m = \frac{c}{|x|^2}(d-1)^{|x|/2}$ the branching process is transient.*

PROOF. Examine some $y = (y_1, y_2)$. Let $z = (z_1, z_2)$ be a random point in $L(y)$. We wish to understand the distribution of $|z_i|$. Clearly, z_1 and z_2 are independent. Let us therefore examine z_1. Going to distance k_1 from y_1 is equivalent to making a nonbacktracking walk of distance k_1. With probability $\frac{1}{d}$, the first step is in the direction of 0. If this happens, then at each step we have probability $\frac{1}{d-1}$ to step in the direction of 0. Once we did one step away from 0, we can never go back. Therefore, the number of steps taken in the direction of 0 is dominated by a geometric variable with expectation $\frac{1}{d-2}$. We get that for a given point $\mathbf{u} \in \mathbf{T}$ with $|\mathbf{u}| = l$, the probability that the corresponding particle is in 0 can be bounded by

$$\begin{aligned}\mathbb{P}\big(\varphi(\mathbf{u})_1 = 0\big) &\le \mathbb{P}\left(\sum_{i=1}^{l} \mathrm{Geom}_i \ge \frac{lk_1}{2}\right) = \mathbb{P}\left(\sum_{i=1}^{\lceil lk_1/2\rceil + l-1} \mathrm{Bin}_i < l\right)\\ &\le (d-1)^{-\lceil lk_1/2\rceil} \sum_{i=0}^{l-1} \binom{\lceil lk_1/2\rceil + l - 1}{i} \le \big(Ck_1(d-1)^{-k_1/2}\big)^l,\end{aligned}$$

where Geom_i are independent geometric variables with expectation $\frac{1}{d-2}$ and Bin_i are independent Bernoulli trials which give 0 with probability $\frac{1}{d-1}$ and 1 with probability $1 - \frac{1}{d-1}$. The same calculation for the other tree gives

$$\mathbb{P}\big(\varphi(\mathbf{u})_2 = 0\big) \le \big(Ck_2(d-1)^{-k_2/2}\big)^l$$

and since the two trees are independent we get

$$\mathbb{P}\big(\varphi(\mathbf{u}) = 0\big) \le \big(Ck_1k_2(d-1)^{-|x|/2}\big)^l \le \big(C_1|x|^2(d-1)^{-|x|/2}\big)^l.$$

Taking m with $m \cdot C_1|x|^2(d-1)^{-|x|/2} < 1$, we see that

$$\mathbb{E}\big(|\{\mathbf{u} \in \mathbf{T} \text{ s.t. } \varphi(\mathbf{u}) = 0\}|\big) \le \sum_{l=0}^{\infty} \big(m \cdot C_1|x|^2(d-1)^{-|x|/2}\big)^l < \infty.$$

Thus, the process is transient, proving the claim. With (6), this also finishes the proof of Lemma 5. □

As explained in the Introduction, Lemma 5 shows that the triangle sum grows only polynomially, and the next step is to show an "open triangle condition" [recall (4), page 1870] with "logarithmic opening." Here is the precise formulation.

LEMMA 6. *There exists some* C_1 *such that for any* $r \geq 2$ *and any* $w \in T \times T$ *with* $|w| > C_1 \log r$, $\nabla(w; r) \leq \frac{1}{2}$.

Recall that $\nabla(w;r) = \sum_{u,v} \mathbb{P}(0 \overset{r}{\leftrightarrow} u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(v \overset{r}{\leftrightarrow} w)$.

PROOF OF LEMMA 6. It will be convenient to replace the restriction that the path between 0 and u is of length $\leq r$ with simply $|u_i| \leq r$ (as before, u_1 and u_2 are the two coordinates and $|u_i|$ is their distance from the root of T), and similarly $|v_i| \leq r + |w|$. Denote $s = r + |w|$. Applying Lemma 5, we get

$$
\begin{aligned}
\nabla(w;r) &= \sum_{u,v} \mathbb{P}(0 \overset{r}{\leftrightarrow} u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(v \overset{r}{\leftrightarrow} w) \\
&\leq C \sum_{|u|,|v| \leq s} |u|^2 (d-1)^{-|u|/2} |u-v|^2 (d-1)^{-d(u,v)/2} \\
&\qquad \times |v-w|^2 (d-1)^{-d(v,w)/2} \\
&\leq C s^6 \sum_{|u_1|,|u_2|,|v_1|,|v_2| \leq s} (d-1)^{-(|u|+d(u,v)+d(v,w))/2}
\end{aligned}
\tag{7}
$$

and at this point the sum decomposes into a product of one term for each tree. Totally we get

$$
\nabla(w;r) = C s^6 \biggl(\sum_{|u_1|,|v_1| \leq s} (d-1)^{-(|u_1|+d(u_1,v_1)+d(v_1,w_1))/2} \biggr)^2 .
$$

Estimating the terms in the last expression is straightforward. The subtree generated by 0, u_1, v_1 and w_1 may take one of the 3 shapes in Figure 2. Take the first (leftmost) case as an example. Denote the two branch points by a and

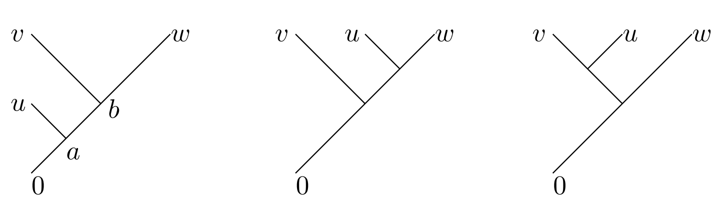

FIG. 2. 3 *ways* 0, u, v *and* w *may be connected in the tree.*

b so that $|u_1| = |a| + d(u_1,a)$, $d(u_1,v_1) = d(u_1,a) + d(a,b) + d(b,v_1)$ and $d(v_1,w_1) = d(v_1,b) + d(b,w_1)$. We get

$$\tfrac{1}{2}\big(|u_1| + d(u_1,v_1) + d(v_1,w_1)\big) = d(a,u_1) + d(b,v_1) + \tfrac{1}{2}|w_1|$$

(in the second case of Figure 2 you get a $\geq$ rather than an $=$). Fixing a and summing over all u_1 gives

$$\sum_{\substack{u_1 \text{ in the subtree}\\ \text{of } a, |u_1 - a| \leq s}} (d-1)^{-d(a,u_1)} = s.$$

Similarly, fixing b and summing over v_1 gives another factor of s. Finally, a and b have $\leq s$ possibilities each. Hence, we get

$$\sum_{\substack{u_1, v_1 \text{ connected as}\\ \text{in the first diagram}}} (d-1)^{-(|u_1| + d(u_1,v_1) + d(v_1,w_1))/2} \leq s^4 (d-1)^{-|w_1|/2}.$$

A similar calculation works for the other 2 diagrams and we get

$$\sum_{|u_1|, |v_1| \leq s} (d-1)^{-(|u_1| + d(u_1,v_1) + d(v_1,w_1))/2} \leq 3s^4 (d-1)^{-|w_1|/2}.$$

The sum over u_2, v_2 and w_2 is the same, and multiplying we get

$$\sum_{|u|, |v| \leq s} (d-1)^{-(|u| + d(u,v) + d(v,w))/2} \leq 9s^8 (d-1)^{-|w|/2} \tag{8}$$

(we remark this partial step in the computation as it will be needed below in Lemma 10. Note that it works for an arbitrary s and not just for $s = r + |w|$). Inserting into (7), we end up with

$$\nabla(w; r) \leq C(r + |w|)^{14} (d-1)^{-|w|/2}$$

and the lemma is finished: with a choice of C_1 sufficiently large we get $\nabla(w; r) \leq \frac{1}{2}$. □

LEMMA 7. *Let C_1 be as in Lemma 6. Then for any $r \geq 2$ and any w with $|w| \geq C_1 \log r$,*

$$\mathbb{E}\big(\big|\{(x,y) : 0 \overset{r}{\leftrightarrow} x, xw \overset{r}{\leftrightarrow} y, 0 \not\leftrightarrow y\}\big|\big) \geq \tfrac{1}{2} G(r)^2,$$

where xw stands for the product in the group Γ whose Cayley graph is $T \times T$.

Note that in the restriction $0 \not\leftrightarrow y$ we do not require anything from the length of the path. The restrictions that the path is $\leq r$ apply only to the paths from 0 to x and from xw to y.

The proof is identical to that of [11], Lemma 3.2, and we include it mainly for completeness.

PROOF OF LEMMA 7. By multiplying with x^{-1} from the left and then doing the change of variables $x^{-1}y \mapsto y$, $x^{-1} \mapsto x$, we see that it is enough to show

$$\mathbb{E}(|\{(x,y): 0 \overset{r}{\leftrightarrow} x, w \overset{r}{\leftrightarrow} y, 0 \not\leftrightarrow w\}|) \geq \tfrac{1}{2}G(r)^2.$$

Fix some x and y. Now let us condition on the cluster of 0, $\mathcal{C}(0)$. We get

$$\begin{aligned}&\mathbb{P}(0 \overset{r}{\leftrightarrow} x, w \overset{r}{\leftrightarrow} y, 0 \not\leftrightarrow w)\\ &\quad= \sum_{\text{admissable } A, w\notin A} \mathbb{P}(\mathcal{C}(0) = A)\mathbb{P}(w \overset{r}{\leftrightarrow} y \mid \mathcal{C}(0) = A),\end{aligned}$$

where "A admissible" means that A is a connected subgraph of $T \times T$ containing 0, x, and a path of length $\leq r$ between 0 and x. Note that for admissible A with $w \notin A$ we have $\mathbb{P}(w \overset{r}{\leftrightarrow} y \mid \mathcal{C}(0) = A) = \mathbb{P}(w \overset{r}{\leftrightarrow} y \text{ off } A)$ where the event $\{w \overset{r}{\leftrightarrow} y$ off $A\}$ means that there exists an open path of length at most r connecting w to y which avoids the vertices of A. At this point, we can remove the condition $w \notin A$ since in this case the event $\{w \overset{r}{\leftrightarrow} y$ off $A\}$ is empty. We get

$$\mathbb{P}(0 \overset{r}{\leftrightarrow} x, w \overset{r}{\leftrightarrow} y, 0 \not\leftrightarrow w) = \sum_{\text{admissable } A} \mathbb{P}(\mathcal{C}(0) = A)\mathbb{P}(w \overset{r}{\leftrightarrow} y \text{ off } A).$$

Now, obviously

$$\mathbb{P}(0 \overset{r}{\leftrightarrow} x)\mathbb{P}(w \overset{r}{\leftrightarrow} y) = \sum_{\text{admissable } A} \mathbb{P}(\mathcal{C}(0) = A)\mathbb{P}(w \overset{r}{\leftrightarrow} y),$$

and we subtract these two equalities and get

$$\begin{aligned}&\mathbb{P}(0 \overset{r}{\leftrightarrow} x, w \overset{r}{\leftrightarrow} y, 0 \not\leftrightarrow w)\\ &\quad= \mathbb{P}(0 \overset{r}{\leftrightarrow} x)\mathbb{P}(w \overset{r}{\leftrightarrow} y)\\ &\qquad- \sum_{\text{admissable } A} \mathbb{P}(\mathcal{C}(0) = A)\mathbb{P}(w \overset{r}{\leftrightarrow} y \text{ only on } A),\end{aligned} \tag{9}$$

where the event $\{w \overset{r}{\leftrightarrow} y$ only on $A\}$ means that there exists an open path between w and y of length at most r and any such path must have a vertex in A. Denoting such a vertex by v we get that $\{w \overset{r}{\leftrightarrow} v\} \circ \{v \overset{r}{\leftrightarrow} y\}$. Hence, for any subgraph A of $T \times T$ we have

$$\mathbb{P}(w \overset{r}{\leftrightarrow} y \text{ only on } A) \leq \sum_{v\in A} \mathbb{P}(\{w \overset{r}{\leftrightarrow} v\} \circ \{v \overset{r}{\leftrightarrow} y\}).$$

Putting this into the second term of the right-hand side of (9) and changing the order of summation gives that we can bound this term from above by

$$\begin{aligned}\text{The 2nd term in (9)} &\leq \sum_{v\in T\times T} \mathbb{P}(\{v \overset{r}{\leftrightarrow} w\} \circ \{v \overset{r}{\leftrightarrow} y\}) \sum_{A \text{ admissible}, v\in A} \mathbb{P}(\mathcal{C}(0) = A)\\ &= \sum_{v\in T\times T} \mathbb{P}(\{v \overset{r}{\leftrightarrow} w\} \circ \{v \overset{r}{\leftrightarrow} y\})\mathbb{P}(0 \overset{r}{\leftrightarrow} x, 0 \leftrightarrow v).\end{aligned}$$

Now, if $0 \overset{r}{\leftrightarrow} x$ and $0 \leftrightarrow v$ then there exists u such that the events $0 \overset{r}{\leftrightarrow} u$, $u \leftrightarrow v$ and $u \overset{r}{\leftrightarrow} x$ occur disjointly. We use the BK inequality and get

$$\leq \sum_{u,v \in T \times T} \mathbb{P}(v \overset{r}{\leftrightarrow} w)\mathbb{P}(v \overset{r}{\leftrightarrow} y)\mathbb{P}(0 \overset{r}{\leftrightarrow} u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(u \overset{r}{\leftrightarrow} x).$$

Let us now sum (9) over x and y and use the estimate above for the second term on its right-hand side. We get

$$\begin{aligned} &\sum_{x,y \in T \times T} \mathbb{P}(0 \overset{r}{\leftrightarrow} x, w \overset{r}{\leftrightarrow} y, 0 \not\leftrightarrow w) \\ &\qquad \geq G(r)^2 - G(r)^2 \sum_{u,v} \mathbb{P}(0 \overset{r}{\leftrightarrow} u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(v \overset{r}{\leftrightarrow} w) \end{aligned} \tag{10}$$

(please record this inequality in this form as we will need it later in Lemma 10). With Lemma 6, we are done. □

LEMMA 8. *There exists some constant C_1 such that $G(2r) \geq r^{-C_1} G(r)^2$ for all $r \geq 2$.*

PROOF. Fix w as in Lemma 7—to be more precise, let w be of minimal distance from 0 satisfying the conclusion of Lemma 7. The proof uses a modification argument, namely we show that by a modification that "costs" no more than r^C, one gets from the event of Lemma 7 to the event $\{0 \overset{2r+|w|}{\longleftrightarrow} y\}$, summing over the probabilities of which would allow to lower bound $G(2r + |w|)$. This is enough because, clearly,

$$B_{\mathrm{int}}(2r + |w|) \subset \bigcup_{x \in B_{\mathrm{int}}(2r)} B(x, |w|)$$

[note that the ball on the right, $B(x, |w|)$, is in the original graph and not in the intrinsic metric] and hence $G(2r + |w|) \leq G(2r) \cdot |B(|w|)| \leq G(2r) \cdot 2d^{|w|}$ and since $|w| \leq C \log r$,

$$G(2r) \geq r^{-C} G(2r + |w|)$$

so it is enough to lower bound $G(2r + |w|)$. Returning to the modification, the process would be to take the clusters containing the path from 0 to x and from xw to y and connect them by the shortest possible path. Formally, we do as follows. Let A be the collection of all triplets (π, x, y) such that $x, y \in T \times T$ and π is some configuration on $T \times T$ such that in π we have $0 \overset{r}{\leftrightarrow} x$, $xw \overset{r}{\leftrightarrow} y$ and $0 \not\leftrightarrow y$. Let B be the collection of all couples (π, y) where π is some configuration such that $0 \overset{2r+|w|}{\longleftrightarrow} y$. We shall construct a $\varphi : A \to B$ with the following two properties:

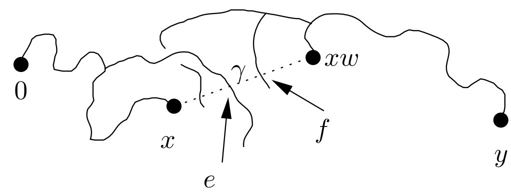

FIG. 3. *The construction of the modification φ.*

- φ is no more than r^C to 1.
- The Radon–Nikodym derivative of φ is bounded below by r^{-C} (we consider A and B as measure spaces with the counting measure for x and y and the usual product measure for π).

Clearly, once φ is constructed, we would get

$$G(2r+|w|)=|B|\geq r^{-C}|A|\geq r^{-C}\cdot\tfrac{1}{2}G(r)^2,$$

where the last inequality is by Lemma 7 and where $|A|$ and $|B|$ stand for the total measure of A and B, respectively. So we need only construct φ.

The construction is as follows. Let $(\pi,x,y)\in A$. Let γ be a shortest path from x to xw (choose γ arbitrarily, e.g., first walk on the first tree and then on the second). Let e be the last point on γ which is in $\mathcal{C}(x)$. Let f be the first point on γ after e which is in $\mathcal{C}(xw)$. Let π' be the configuration one gets by opening every edge of π on the piece of γ between e and f. See Figure 3. Define $\varphi(\pi,x,y)=(\pi',y)$. Clearly, the Radon–Nikodym derivative is equal to

$$\left(\frac{p}{1-p}\right)^{\#\text{closed edges in }(e,f)}\geq\left(\frac{p}{1-p}\right)^{|\gamma|}\geq\left(\frac{p}{1-p}\right)^{C\log r}=r^{-C},$$

where $p=p_c(T\times T)$. Recall that $|\gamma|\leq C\log r$ by Lemma 7. To show that φ is no more than r^C to 1, examine one couple $(\pi',y)\in B$. If $\varphi(x,y,\pi)=(\pi',y)$, then all edges between e and f must be pivotal for the connection $0\leftrightarrow y$. Since there can be no more than $2r+|w|$ edges which are pivotal for the connection $0\xleftrightarrow{2r+|w|}y$, we see that e has no more than $2r+|w|$ possibilities. Since $|e-x|\leq|w|$, we see that x has no more than $(d-1)^{|w|}\leq(d-1)^{C\log r}=r^C$ possibilities. Once x is fixed so is γ. The original configuration on γ has $2^{|w|}\leq r^C$ possibilities. The shows that φ is no more than r^C to 1 and finishes the lemma. $\square$

LEMMA 9. $G(r)\leq Cr^C$.

PROOF. This is a more-or-less direct corollary of Lemmas 1 and 8. Let C_1 be the constant from Lemma 8. Assume by contradiction that for some r, $G(r)>$

$(4r)^{C_1}$. Then by applying Lemma 8 repeatedly,

$$\begin{aligned}
G(2r) &\ge r^{-C_1} G(r)^2 > (2^4 r)^{C_1},\\
G(4r) &\ge (2r)^{-C_1} G(2r)^2 > (2^{8-1} r)^{C_1},\\
&\vdots\\
G(2^{k+1}r) &\ge (2^k r)^{-C_1} G(2^k r)^2 > (2^{2^{k+2}-\sum_{l=1}^{k} l 2^{k-l}} r)^{C_1} > (2^{2^{k+1}} r)^{C_1},
\end{aligned}$$

which means that $G(s)$ increases exponentially in s, contradicting Lemma 1. □

We now repeat the arguments of Lemmas 7–9, but use Lemma 9 as an input to get better results.

LEMMA 10. *There exists some constant C_1 such that for any $r \ge 3$ and any w with $|w| \ge C_1 \log\log r$,*

$$\mathbb{E}\big(\big|\{(x,y) : 0 \overset{r}{\leftrightarrow} x, xw \overset{r}{\leftrightarrow} y, 0 \nleftrightarrow y\}\big|\big) \ge \tfrac{1}{2} G(r)^2.$$

PROOF. We start the calculation from (10), which, we recall, stated that

$$\begin{aligned}
&\sum_{x,y \in T\times T} \mathbb{P}(0 \overset{r}{\leftrightarrow} x, w \overset{r}{\leftrightarrow} y, 0 \nleftrightarrow w)\\
&\qquad \ge G(r)^2 - G(r)^2 \sum_{u,v} \mathbb{P}(0 \overset{r}{\leftrightarrow} u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(v \overset{r}{\leftrightarrow} w).
\end{aligned}$$

Recall also that the sum on the right-hand side is denoted by $\nabla(w; r)$. We now separate the sum into two parts, according to whether $\max\{|u|, |v|\} \le \log^2 r$ or not. The first case is calculated exactly as in Lemma 7, as follows

$$\begin{aligned}
&\sum_{|u|,|v| \le \log^2 r} \mathbb{P}(0 \overset{r}{\leftrightarrow} u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(v \overset{r}{\leftrightarrow} w)\\
\text{By Lemma 5}\quad &\le \sum_{|u|,|v| \le \log^2 r} |u|^2 d(u,v)^2 d(v,w)^2 (d-1)^{-(|u|+d(u,v)+d(v,w))/2}\\
&\le C|w|^2 \log^{12} r \sum_{|u|,|v| \le \log^2 r} (d-1)^{-(|u|+d(u,v)+d(v,w))/2}\\
\text{By (8)}\quad &\le C|w|^2 \log^{12} r \cdot 9(\log^2 r)^8 (d-1)^{-|w|/2}\\
&= C|w|^2 (d-1)^{-|w|/2} \log^{28} r
\end{aligned}$$

and this is $\le \frac{1}{4}$ if only C_1 is sufficiently large. Now assume $\max\{|u|, |v|\} > \log^2 r$. As u and v are symmetric, we may assume $|u| > \log^2 r$. Let L be the level of $T \times T$

which contains u, namely $L = \{z : z_1 = u_1, z_2 = u_2\}$. Clearly, $|L| = (d-1)^{|u|}$. By Lemma 9,

$$\sum_{z\in L} \mathbb{P}(0 \overset{r}{\leftrightarrow} z) \le G(r) \le Cr^C.$$

But L is completely symmetric, so for any $z \in L$, $\mathbb{P}(0 \overset{r}{\leftrightarrow} z) = \mathbb{P}(0 \overset{r}{\leftrightarrow} u)$. We get

$$\mathbb{P}(0 \overset{r}{\leftrightarrow} u) = \frac{1}{|L|} \sum_{z\in L} \mathbb{P}(0 \overset{r}{\leftrightarrow} z) \le \frac{Cr^C}{(d-1)^{|u|}}.$$

We need to compare this estimate to the estimate of (8) from Lemma 7 which we also used above, namely to $(d-1)^{-|u|/2}$. So we write this as

$$\mathbb{P}(0 \overset{r}{\leftrightarrow} u) \le (d-1)^{-|u|/2} \cdot \frac{Cr^C}{(d-1)^{(\log^2 r)/2}} \le (d-1)^{-|u|/2} \cdot Cr^{-15}.$$

This allows us to write

$$\begin{aligned}
&\sum_{\substack{u,v\\ |u|>\log^2 r}} \mathbb{P}(0 \overset{r}{\leftrightarrow} u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(v \overset{r}{\leftrightarrow} w)\\
&\quad \le Cr^{-15} \cdot Cr^4 \sum_{|u|,|v|\le r} (d-1)^{-(|u|+d(u,v)+d(v,w))/2}\\
&\quad \overset{(8)}{\le} Cr^{-11} \cdot Cr^8 (d-1)^{-|w|/2}
\end{aligned}$$

and we see that this part of the sum is in fact negligible (if r is sufficiently large or if C_1 is chosen sufficiently large). This shows that $\nabla(w; r) \le \frac{1}{2}$ and concludes the lemma. □

LEMMA 11. $G(r) \le Cr(\log r)^C$.

PROOF. This is nothing more than repeating the arguments of Lemmas 8 and 9. Let us verify some of the details. We first show

$$G(2r) \ge \frac{G(r)^2}{r \log^C r}, \tag{11}$$

that is, the analog of Lemma 8. We again construct a $\varphi : A \to B$ (A and B being exactly as in Lemma 8) which is no more than $r \log^C r$ to 1, and with the Radon–Nikodym derivative bounded below by $\log^{-C} r$. The construction of φ is identical, that is, we take the shortest path γ from x to xw, let e be the last point of $\mathcal{C}(x)$ on γ and f be the first point of $\mathcal{C}(xw)$ on γ after e. This time, though, because $|w| \approx \log\log r$, the Radon–Nikodym derivative of φ will be $\ge (p/(1-p))^{|w|} \ge \log^{-C} r$. To invert φ, we need to find e, x and ω. e still has $2r + |w|$ possibilities, but given e, x has only $(d-1)^{|w|} \le \log^C r$ possibilities, and ω has only $2^{|w|} \le \log^C r$ possibilities. This shows (11).

Concluding from (11) the lemma is identical to the proof of Lemma 9—if $G(r) > r(4\log r)^C$ for some r then it starts growing exponentially—and we will omit it. $\square$

LEMMA 12. *In the extrinsic metric*,

$$\mathbb{E}\big(\mathcal{C}(0)\cap B(r)\big) \le Cr^3.$$

The idea is quite simple. We consider $\mathcal{C}(0)$ as a branching process embedded into $T \times T$ and show that it escapes from 0 with "positive speed." Our "time" for the branching process is the intrinsic distance $d_{\text{int}}(0,x)$ so escaping in positive speed mean simply that $d_{\text{int}}(0,x) \approx |x|$ (this is step 5 of the sketch on page 1871). Here, are the details. It is enough to prove:

LEMMA 13. $\mathbb{E}((B_{\text{int}}(4r^2)\setminus B_{\text{int}}(r^2)) \cap B(r)) \le Cr^{-2}$.

PROOF OF LEMMA 12 GIVEN LEMMA 13. We apply Lemma 13 with the parameter $r_{\text{Lemma 13}} = r2^k$ and get

$$\begin{aligned}
&\mathbb{E}\big(\big(B_{\text{int}}(4^{k+1}r^2)\setminus B_{\text{int}}(4^k r^2)\big)\cap B(r)\big)\\
&\qquad \le \mathbb{E}\big(\big(B_{\text{int}}(4^{k+1}r^2)\setminus B_{\text{int}}(4^k r^2)\big)\cap B(r2^k)\big) \le \frac{C}{r^2 4^k},
\end{aligned}$$

which we sum over k to get

$$\mathbb{E}\big(\big(\mathcal{C}(0)\setminus B_{\text{int}}(r^2)\big)\cap B(r)\big) \le \frac{C}{r^2}.$$

For the interior part, we just use Lemma 11 and get

$$\mathbb{E}\big(B_{\text{int}}(r^2)\cap B(r)\big) \le \mathbb{E}(B_{\text{int}}(r^2)) \le Cr^2\log^C r$$

proving the claim. $\square$

PROOF OF LEMMA 13. Clearly, we may assume r is sufficiently large. We shall define a sequence of subsets of $\mathcal{C}(0)$, $\{\partial_m\}_{m=1}^\infty$ (here ∂_m is just a notation). Intuitively, you should consider ∂_m simply as $\partial B_{\text{int}}(mr)$—I could not make the proof work with this definition of ∂_m so a somewhat more complicated, inductive definition will be used [formally also ∂_m is a set of vertices while $\partial B_{\text{int}}(mr)$ is a set of edges]. It will preserve some of the feeling of $\partial B_{\text{int}}(mr)$ as every $v \in \partial_m$ will satisfy $\frac{1}{2}mr \le d_{\text{int}}(0,v) \le mr$, and since for any $x \in \mathcal{C}(0)$ there will be a path from 0 to x visiting each ∂_m in turn and spending $\approx r$ steps between any two levels. This path will not be a geodesic in $\mathcal{C}(0)$ (i.e., a shortest possible path), but this is not important.

During the induction process, we shall expose parts of $\mathcal{C}(0)$. We shall denote the exposed part by E_m [which is formally a collection of edges of $\mathcal{C}(0)$ and of $\partial\mathcal{C}(0)$, though we shall say about a vertex v that it is "in E_m" if there exists some path of open edges in E_m from 0 to v]. ∂_m will be the "boundary of E_m" in the sense that any vertex of ∂_m is in E_m, and any vertex in $E_m \setminus \partial_m$ is fully exposed, that is, all edges coming out of it (whether open or closed) are in E_m. We start with $\partial_0 = \{0\}$ and E_0 having no edges. Note that saying about E_m that it is "exposed" is not just a name, it carries some meaning, namely that for any m and any set of edges A, one can infer whether $E_m = A$ or not merely by examining the states of the edges of A. We will keep this property through the induction.

1. *The construction of* ∂_m. Assume E_m has already been calculated. Let $y \in \partial_m$. We will now construct the "children of y" which will belong to ∂_{m+1}. For this purpose, let $Q = Q(y)$ be the set of all vertices q satisfying that q is connected to y by an open path of length $\le r$ off E_m (we are making the exception that y is in E_m, but no other vertex of the path can be in E_m, this is the precise meaning of "off" here). Clearly, $\mathbb{E}(Q) \le G(r)$ and by Lemma 11 we get

$$\mathbb{E}(Q) \le C_1 r(\log r)^C. \tag{12}$$

We make at this point the convention that every formula involving Q is in fact conditioned over E_m. For example, (12) should be read as $\mathbb{E}(Q|E_m) \le C_1 r(\log r)^C$. We need to examine two special parts of Q—the vertices "close to y" and the vertices "beyond y." For the first part, let ℓ be defined by

$$|B(\ell)| \le \sqrt{r}$$

and ℓ being the maximal with this property (note that this is the usual ball in our original graph $T \times T$, and we have $\ell \approx \log r$). Clearly,

$$\mathbb{E}\big(|Q \cap B(y,\ell)|\big) \le |B(y,\ell)| \le \sqrt{r}. \tag{13}$$

For the second part, fix some k_1 and k_2, and let L be the corresponding level, that is,

$$L = \{z : d(z_i, y_i) = k_i, i = 1, 2\}.$$

A straightforward calculation shows that

$$\big|\{z \in L : |z| < |y|\}\big| \le |L|e^{-c|k|}.$$

However, L is completely symmetric with respect to y. Therefore,

$$\mathbb{E}\big|\{z \in L : z \overset{r}{\leftrightarrow} y \text{ and } |z| < |y|\}\big| \le G(r)\frac{|\{z \in L : |z| < |y|\}|}{|L|} \le G(r)e^{-c|k|}.$$

Requiring that the connection is off E_m only makes things worse, so

$$\mathbb{E}\big|\{z \in Q \cap L : |z| < |y|\}\big| \le G(r)e^{-c|k|}. \tag{14}$$

Summing this over all $|k| > \ell$ gives

$$\mathbb{E}\big|\{z \in Q \setminus B(y,\ell) : |z| < |y|\}\big| \le \sum_{|k|>\ell} G(r)e^{-c|k|} \le Ce^{-c\ell}G(r) \le Cr^{1-c}. \tag{15}$$

We may sum up (13) and (15) as

$$\mathbb{E}\big|\{z \in Q : |z| < |y| + \ell\}\big| \le C_2 r^{1-c}. \tag{16}$$

Equipped with the estimates (12) and (16) we may now proceed to define an important element of the construction, the parameter s. Define $Q_s = Q_s(y)$ similarly to Q but with the requirement that the shortest open path (off E_m) has length exactly s [below we will denote this by "$d_{\mathrm{off}\ E_m}(y,z) = s$" for short]. We get $Q = \dot\bigcup_{s=1}^{r} Q_s$. This means that for at least $\frac{2}{3}$ of the s between $\frac{1}{2}r$ and r we must have

$$\mathbb{E}|Q_s| \le 6C_1(\log r)^C \tag{17}$$

and similarly,

$$\mathbb{E}\big|\{z \in Q_s : |z| < |y| + \ell\}\big| \le 6C_2 r^{-c}. \tag{18}$$

Of course, the set of s which satisfy (17) may be different than the set of s which satisfy (18), but they intersect and we choose one s which satisfies both [sometimes we will denote it by $s(y)$ for clarity]. We note that the set of such "good" s is a random set which depends on E_m and on y, and this is exactly why we cannot define $\partial_m = \partial B_{\mathrm{int}}(0, mr)$.

We may now complete the description of the construction. We define E_{m+1} to be the set of all edges (z,x) for all z satisfying that there exists a $y \in \partial_m$ such that $d_{\mathrm{off}\ E_m}(y,z) < s(y)$; and all x which are a neighbor of z. We define ∂_{m+1} to be the boundary of E_{m+1} in the sense above. We get that for any $z \in \partial_{m+1}$ there exists some $y \in \partial_m$ such that the $d_{\mathrm{off}\ E_m}(y,z) = s(y)$ (but not vice versa—see Figure 4).

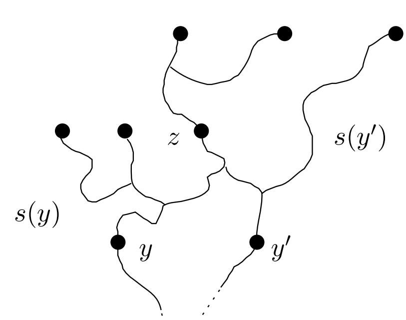

FIG. 4. *The point z has distance exactly $s(y)$ from y but is not included in ∂_{m+1} because its distance from y' is less than $s(y')$.*

It will be convenient to note at this point that

$$B_{\text{int}}\left(\tfrac{1}{2}mr\right) \subset E_m \subset B_{\text{int}}(mr) \tag{19}$$

(here E_m is considered as a set of vertices). This is a simple consequence of the fact that all $s(y)$ for all y are in $[\frac{1}{2}r, r]$. Therefore the direction $E_m \subset B_{\text{int}}(mr)$ is immediate. For the other direction let x satisfy that $d_{\text{int}}(x, E_m) \le \frac{1}{2}r$. Let $y \in \partial_m$ be the closest point in E_m to x. Then there exists an open path of length $\le \frac{1}{2}r < s(y)$ off E_m from y to x. Hence, $x \in E_{m+1}$ by definition. Hence,

$$\begin{aligned} & B_{\text{int}}\left(\tfrac{1}{2}(m+1)r\right) = \left\{x : d_{\text{int}}\left(x, B_{\text{int}}\left(\tfrac{1}{2}mr\right)\right) \le \tfrac{1}{2}r\right\} \\ \text{inductively} \qquad & \subset \left\{x : d_{\text{int}}(x, E_m) \le \tfrac{1}{2}r\right\} \\ \text{by the argument above} \qquad & \subset E_{m+1} \end{aligned}$$

showing (19).

2. *The number of bad paths.* Let $x \in \partial_m$ be some vertex. Then there exists a sequence $0 = x_0, \dots, x_m = x$ with $x_i \in \partial_i$ and such that for each $i = 0, \dots, m-1$ we have that $d_{\text{off } E_i}(x_i, x_{i+1}) = s(x_i)$. This sequence might not be unique but this is not important. We call such sequences "∂-paths."

DEFINITION 3. We say that a given ∂-path $x_0, \dots, x_m$ is bad if there are $> m/\sqrt{\log r}$ values of $i \in \{0, \dots, m-1\}$ such that $|x_{i+1}| < |x_i| + \ell$.

Let us estimate the expected number of bad ∂-paths. Let $I \subset \{0, \dots, m-1\}$ be some set of indices with $|I| > m/\sqrt{\log r}$. Let $E(I; m)$ be the expected number of ∂-paths such that $|x_{i+1}| < |x_i| + \ell$ for every $i \in I$. We get, directly from the choice of s above,

$$E(I; m) \le \begin{cases} E(I; m-1) \cdot C(\log r)^C, & m-1 \notin I, \\ E(I \setminus \{m-1\}; m-1) \cdot Cr^{-c}, & m-1 \in I, \end{cases} \tag{20}$$

[seeing (20) is a standard exercise in the "off method"—one conditions on E_{m-1} and examines each path individually. Because being connected off E_{m-1} does not examine the edges of E_{m-1} at all, the estimates (17) and (18) hold after the conditioning. Here is where we need the property that E_m is "exposed," i.e., that conditioning over $E_m = A$ gives no information on edges not in A]. We apply (20) recursively and get

$$E(I; m) \le (C(\log r)^C)^{m-|I|}(Cr^{-c})^{|I|}$$

and since $|I| > m/\sqrt{\log r}$,

$$\le \exp\left(m\left(C \log\log r - c\sqrt{\log r}\right)\right).$$

Summing over all I can be bounded roughly by multiplying this by 2^m, and we end with

$$\mathbb{E}|\{\text{bad } \partial\text{-paths}\}| \le \exp\bigl(m\bigl(C - c\sqrt{\log r}\bigr)\bigr) \tag{21}$$

and in particular we get the same estimate for the probability that even one bad ∂-path exists.

3. *The contribution of the bad part.* Examine a good path $x_0, \dots, x_m$. We know it contains $< m/\sqrt{\log r}$ "bad" i for which $|x_{i+1}| < |x_i| + \ell$. We still need to estimate how much "damage" can each such bad i cause, that is, upper bound $|x_i| - |x_{i+1}|$. For this purpose, we return to (14), which shows that

$$\mathbb{E}\bigl|\{z \in Q : |z| < |y| - j\}\bigr| \le \sum_{|k|>j} G(r)e^{-c|k|} \le G(r)e^{-cj} \qquad \forall y \in \partial_i \ \forall i$$

and in particular there exists some constant C_3 such that

$$\mathbb{E}\bigl|\{z \in Q : |z| < |y| - C_3 \log r\}\bigr| \le r^{-10} \qquad \forall y \in \partial_i \ \forall i. \tag{22}$$

We now wish to sum over all v, so denote by **BE** (for "bad edges") the set of all (y, z) with $y \in \partial_i$, $i = 0, \dots, m-1$, and $z \in Q_{s(y)}(y)$ such that $|z| < |y| - C_3 \log r$. So we want to estimate $\mathbb{P}(\mathbf{BE} \neq \varnothing)$. Recall that $\partial_i \subset B_{\mathrm{int}}(ir)$ (19). Recall also our convention about Q which stated that (22) in fact holds also after conditioning over E_i. We write

$$\begin{aligned}
\mathbb{P}(\mathbf{BE} \neq \varnothing) \le \mathbb{E}(|\mathbf{BE}|) &\le \sum_{i=0}^{m-1} \sum_{y,z} \mathbb{E}\mathbb{P}\bigl((y,z) \in \mathbf{BE} | E_i\bigr) \\
&\le \sum_{i=0}^{m-1} \sum_{y} \mathbb{E}\bigl(r^{-10}\mathbb{P}(y \in \partial_i | E_i)\bigr) \\
&= r^{-10} \sum_{i=0}^{m-1} \mathbb{E}|\partial_i| \le r^{-10} G(mr) \\
&\le mr^{-9}(\log mr)^C.
\end{aligned}$$

4. *Wrapping it all up.* Let $m \in \{r, \dots, 8r\}$. Our analysis of bad ∂-paths concluded with (21) which states that

$$\begin{aligned}
&\mathbb{P}(\exists \text{ bad } \partial\text{-path for some } m \in \{r, \dots, 8r\}) \\
&\qquad \le (7r+1)\exp\bigl(r\bigl(C - c\sqrt{\log r}\bigr)\bigr) \le Cr^{-7}.
\end{aligned}$$

Denote this event by $\mathcal{B}_1$. From part 3 of the proof, we have that, with probability $\le Cr^{-7}$, there exists some ∂-path and some i such that $|x_{i+1}| \le |x_i| - C_3 \log r$.

Denote this event by $\mathcal{B}_2$ and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$. If $\neg\mathcal{B}$ happened (the "good" case), then for every $x \in \partial_m$ we have

$$|x| \geq \ell m - \frac{m}{\sqrt{\log r}} \cdot C_3 \log r \geq cr \log r$$

for r sufficiently large.

We now move from $x \in \partial_m$ to some arbitrary $x \in B_{\mathrm{int}}(4r^2) \setminus B_{\mathrm{int}}(r^2)$. Let m_0 be the first m such that $x \in E_{m+1}$, and let $y \in \partial_m$ be the closest point to x, so that $d_{\mathrm{int}}(y, x) \leq r$. Recall (19) which stated that $B_{\mathrm{int}}(\frac{1}{2}mr) \subset E_m \subset B_{\mathrm{int}}(mr)$. We get that $m_0 \in \{r, \dots, 8r\}$, and therefore (still assuming $\neg\mathcal{B}$) $|y| \geq cr \log r$ so

$$|x| > |y| - d(y, x) \geq cr \log r - r.$$

This finishes the lemma, as we see that $(B_{\mathrm{int}}(4r^2) \setminus B_{\mathrm{int}}(r^2)) \cap B(r)$ can be nonempty only if the bad event $\mathcal{B}$ happened. On the one hand, by Lemma 2 we see that

$$\mathbb{E}\big(|B_{\mathrm{int}}(4r^2)| \cdot \mathbf{1}\{|B_{\mathrm{int}}(4r^2)| > C(\log r)G(4r^2)^2\}\big) \leq Cr^{-2}$$

and by Lemma 11 $C(\log r)G(4r^2)^2 \leq C(\log r)^C r^4 \leq Cr^5$. We get

$$\begin{aligned}
&\mathbb{E}\big(\big(B_{\mathrm{int}}(4r^2) \setminus B_{\mathrm{int}}(r^2)\big) \cap B(r)\big) \\
&\quad \leq \mathbb{E}\big(|B_{\mathrm{int}}(4r^2) \cdot \mathbf{1}\{|B_{\mathrm{int}}(4r^2)| > Cr^5\}\big) + Cr^5 \cdot \mathbb{P}(\mathcal{B}) \\
&\quad \leq Cr^{-2},
\end{aligned}$$

which proves Lemma 13 and hence Lemma 12. □

PROOF OF THEOREM 1. By Lemma 12 and the fact that every u has $(d-1)^{|u|}$ clones,

$$\mathbb{P}(0 \leftrightarrow u) \leq C|u|^3(d-1)^{|u|}. \tag{23}$$

With this we write

$$\begin{aligned}
\nabla &= \sum_{u,v} \mathbb{P}(0 \leftrightarrow u)\mathbb{P}(u \leftrightarrow v)\mathbb{P}(v \leftrightarrow 0) \\
\text{by (23)} \qquad &\leq \sum_{u,v} |u|^3 d(u,v)^3 |v|^3 (d-1)^{-|u|-d(u,v)-|v|} \\
\text{as in Lemma 6} \qquad &\leq \left(C \sum_{u_1,v_1} |u_1|^3 d(u_1,v_1)^3 |v_1|^3 (d-1)^{-|u_1|-d(u_1,v_1)-|v_1|}\right)^2,
\end{aligned}$$

where the last sum is over u_1 and v_1 in the tree T (and not in $T \times T$). Denote by a the point where the paths from 0 to u_1 and from 0 to v_1 split. Then $|u_1| =$

$|a| + d(a, u_1)$, $d(u_1, v_1) = d(u_1, a) + d(a, v_1)$ and $|v_1| = |a| + d(a, v_1)$ so all in all

$$\begin{aligned}
&\sum_{u_1,v_1} (|u_1| d(u_1, v_1)|v_1|)^3 (d-1)^{-|u_1|-d(u_1,v_1)-|v_1|} \\
&\qquad \le \sum_{u_1,v_1} (|u_1| d(u_1, v_1)|v_1|)^3 (d-1)^{-2|a|-2d(a,v_1)-2d(a,v_1)} \\
&\qquad \le C \sum_{u_1,v_1} (|a| d(a, u_1) d(a, v_1))^6 (d-1)^{-2|a|-2d(a,v_1)-2d(a,v_1)}.
\end{aligned}$$

Fixing a we may sum over v_1 in the subtree of a and get $\sum d(a, v_1)^6 (d-1)^{-2d(a,v_1)}$ which is finite and independent of a. The sum over the u_1 in the subtree of a gives an identical contribution. Finally, we sum over a and get a third contribution identical to the previous two. Hence, this sum is finite and so is ∇. $\square$

Acknowledgments. I was fortunate enough to hear the proof of Schramm's lemma in a talk given by Yuval Peres in an AIM workshop on "Percolation on transitive graphs," May 2008. I wish to thank Yuval Peres for the inspiring talk and, of course, the hospitality of the American Institute of Mathematics. Also, I wish to thank Remco van der Hofstad for answering all my questions on lace expansion, Markus Heydenreich for pointing out some errors in a draft version and Johan Tykesson for pointing me to [3].

REFERENCES

[1] AIZENMAN, M. and BARSKY, D. J. (1987). Sharpness of the phase transition in percolation models. *Comm. Math. Phys.* **108** 489–526. MR0874906

[2] AIZENMAN, M. and NEWMAN, C. M. (1984). Tree graph inequalities and critical behavior in percolation models. *J. Stat. Phys.* **36** 107–143. MR0762034

[3] ANTUNOVIĆ, T. and VESELIĆ, I. (2008). Sharpness of the phase transition and exponential decay of the subcritical cluster size for percolation and quasi-transitive graphs. *J. Stat. Phys.* **130** 983–1009. MR2384072

[4] BARSKY, D. J. and AIZENMAN, M. (1991). Percolation critical exponents under the triangle condition. *Ann. Probab.* **19** 1520–1536. MR1127713

[5] BENJAMINI, I., LYONS, R., PERES, Y. and SCHRAMM, O. (1999). Critical percolation on any nonamenable group has no infinite clusters. *Ann. Probab.* **27** 1347–1356. MR1733151

[6] GRIMMETT, G. (1999). *Percolation*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **321**. Springer, Berlin. MR1707339

[7] HARA, T. and SLADE, G. (1990). Mean-field critical behaviour for percolation in high dimensions. *Comm. Math. Phys.* **128** 333–391. MR1043524

[8] HEYDENREICH, M., VAN DER HOFSTAD, R. and SAKAI, A. (2008). Mean-field behavior for long- and finite range Ising model, percolation and self-avoiding walk. *J. Stat. Phys.* **132** 1001–1049. MR2430773

[9] KOZMA, G. Percolation on a product of two trees, 1st version. Available at http://arxiv.org/abs/1003.5240v1.

[10] KOZMA, G. (2011). The triangle and the open triangle. *Ann. Inst. H. Poincaré Probab. Statist.* To appear. Available at http://arxiv.org/abs/0907.1959.

[11] KOZMA, G. and NACHMIAS, A. (2009). The Alexander–Orbach conjecture holds in high dimensions. *Invent. Math.* **178** 635–654. MR2551766

[12] KOZMA, G. and NACHMIAS, A. (2011). Arm exponents in high dimensional percolation. *J. Amer. Math. Soc.* **24** 375–409.

[13] LYONS, R. and PERES, Y. (2011). *Probability on Trees and Networks*. Cambridge Univ. Press, Cambridge. To appear. Current version available at http://mypage.iu.edu/~rdlyons/.

[14] NGUYEN, B. G. (1987). Gap exponents for percolation processes with triangle condition. *J. Stat. Phys.* **49** 235–243. MR0923855

[15] SCHONMANN, R. H. (2001). Multiplicity of phase transitions and mean-field criticality on highly non-amenable graphs. *Comm. Math. Phys.* **219** 271–322. MR1833805

[16] SCHONMANN, R. H. (2002). Mean-field criticality for percolation on planar non-amenable graphs. *Comm. Math. Phys.* **225** 453–463. MR1888869

[17] WOESS, W. (2000). *Random Walks on Infinite Graphs and Groups*. *Cambridge Tracts in Mathematics* **138**. Cambridge Univ. Press, Cambridge. MR1743100

[18] WU, C. C. (1993). Critical behavior or percolation and Markov fields on branching planes. *J. Appl. Probab.* **30** 538–547. MR1232733

DEPARTMENT OF MATHEMATICS
WEIZMANN INSTITUTE OF SCIENCE
REHOVOT 76100
ISRAEL
E-MAIL: gadyk@weizmann.ac.il