Counting independent sets up to the tree threshold

Dror Weitz *

March 16, 2006

Abstract

We consider the problem of approximately counting/sampling weighted independent sets of a graph G with activity λ, i.e., where the weight of an independent set I is $\lambda^{|I|}$. We present a novel analysis yielding a deterministic approximation scheme which runs in polynomial time for *any* graph of maximum degree Δ and $\lambda < \lambda_c = (\Delta-1)^{\Delta-1}/(\Delta-2)^{\Delta}$. This improves on the previously known general bound of $\lambda \leq \frac{2}{\Delta-2}$. The new regime includes the interesting case of $\lambda = 1$ (uniform weights) and $\Delta \leq 5$. The previous bound required $\Delta \leq 4$ for uniform approximate counting and there is evidence that for $\Delta \geq 6$ the problem is hard. Note that λ_c is the critical activity for uniqueness of the *Gibbs* measure on the regular tree of degree Δ, i.e., for $\lambda \leq \lambda_c$ the probability that the root is in the independent set is asymptotically independent of the configuration on the leaves far below. Indeed, our analysis is focused on establishing decay of correlations with distance in the above weighted distribution. We show that on any graph of maximum degree Δ correlations decay with distance at least as fast as they do on the regular tree of the same degree. This resolves an open conjecture in statistical physics. Our comparison of a general graph with the tree uses an algorithmic argument yielding the approximation scheme mentioned above. Also, by existing arguments, establishing decay of correlations for all graphs and $\lambda < \lambda_c$ gives that the *Glauber* dynamics is rapidly mixing in this regime. However, the implication from decay of correlations to rapid mixing of the dynamics is only known to hold for graphs of subexponential growth, and hence, our result regarding the Glauber dynamics is limited to this class of graphs.

*DIMACS Center, Rutgers university, Piscataway, NJ 08854 U.S.A. Email: dror@dimacs.rutgers.edu. Most of this work was done while the author was a postdoctoral fellow at the School of Mathematics, Institute for Advanced Study, Princeton, NJ, supported by a grant from the state of New Jersey and NSF grants DMS-0111298 and CCR-0324906.

1 Introduction

Counting (or sampling) independent sets of a graph G is a problem that has been widely studied both in computer science and statistical physics. The problem usually includes an additional parameter λ, and the goal is to count/sample weighted independent sets such that the weight of an independent set I is proportional to $\lambda^{|I|}$. In statistical physics this model is referred to as the hard-core model with activity λ. Intuitively, we expect the problem to become harder as λ increases (i.e., as the weight shifts to the larger independent sets). Indeed, an algorithm for sampling independent sets with λ arbitrarily large can be used to find a maximum independent set, an NP-hard problem. Based on this intuition, it was shown [15] that for any $\Delta \geq 4$, it is NP-hard to approximate the above sum, even to within a polynomial factor, for graphs of maximum degree Δ and $\lambda > \frac{c}{\Delta}$, where c is a (large enough) absolute constant.

In this paper we are interested in the regime for which an efficient approximation scheme does exist. The best known [7, 23] general bound on λ for which there exists an FPRAS (Fully Polynomial Randomized Approximation Scheme) for counting independent sets is $\lambda \leq \frac{2}{\Delta-2}$, where Δ is the maximum degree of the graph G. This algorithm is based on approximately sampling from the desired distribution using the *Glauber* dynamics, a well-known Markov chain where in each step a vertex is chosen uniformly at random and its value (occupied or unoccupied) is updated conditioned on the current values of its neighbors. The argument in [7, 23] establishes a certain local contraction of the dynamics' operator, one that implies fast convergence of the chain. The same local contraction also implies that the stationary distribution exhibits decay of correlations with distance, i.e., the probability that a vertex v is occupied is asymptotically independent from the configuration of vertices far away from v. This is in line with a general correspondence between rapid mixing of this chain and decay of correlations with distance in the stationary distribution [22, 16, 4, 8, 24]. We note that establishing decay of correlations in distance is of independent importance (in fact, one of the main interests) for statistical physicists because it means that the *Gibbs* measure associated with the model is unique, i.e., there is a *unique* macroscopic equilibrium.

Arguments of the kind used in [7, 23] have a known limit. Specifically, it is known that any argument that works for general graphs and which establishes decay of correlations with distance as a byproduct is bound to fail for $\lambda > \lambda_c \equiv \lambda_c(\Delta) = (\Delta-1)^{\Delta-1}/(\Delta-2)^{\Delta}$. The reason for this is that λ_c is the critical activity for decay of correlations on the infinite regular tree in which every vertex has degree Δ, i.e., for $\lambda > \lambda_c$ the probability that the root of this tree is occupied has non-negligible dependence on the configuration ℓ levels below for arbitrarily large ℓ. (Notice that counting independent sets of the tree is easy — can be done in linear time — for arbitrary λ. Nevertheless, the fact that correlations persist over arbitrarily long distances in the tree for $\lambda > \lambda_c$ forbids any argument that works for general graphs and implies decay of correlation with distance.)

It has been conjectured [20] that the above limit imposed by the regular tree can be matched, i.e., that the regular tree is the worst-case graph in terms of decay of correlations and that for any graph of maximum degree Δ correlations decay with distance throughout the regime $\lambda \leq \lambda_c$. The algorithmic version of this conjecture states that there exists an FPRAS for counting independent sets for any graph of maximum degree Δ and $\lambda \leq \lambda_c$. In this paper we prove both versions of the above conjecture. (For the algorithmic result we require a strict inequality.) Specifically, we present a novel tree representation for a general graph G, one that allows us to show that correlations on G decay with distance at least as fast as they do on the regular tree. Our analysis is algorithmic in nature, yielding a novel, tree-like, deterministic algorithm for counting independent sets. We show

that if correlations on the regular tree decay exponentially fast with distance then for *any* graph of the same maximum degree, the above algorithm gives a $(1+\epsilon)$ approximation of the number of independent sets in time polynomial in $\frac{n}{\epsilon}$, where n is the size of G. The fact that we establish decay of correlations for any graph and $\lambda < \lambda_c$ also implies [22, 16, 4, 8, 24] rapid mixing of the Glauber dynamics in this regime. However, the latter implication is only known to hold in graphs whose growth is subexponential. (The volume of balls of radius r around a vertex v grow subexponentially in r.) Thus, we establish rapid mixing of the Glauber dynamics for any graph in this class and any $\lambda < \lambda_c$. It is worth mentioning here that rapid mixing of the Glauber dynamics for $\lambda < \lambda_c$ was already established in [12] for graphs of girth at least 6 and large degree ($\Delta = \Omega(\log n)$). Thus, our algorithmic results can also be viewed as eliminating the girth and large degree requirements imposed in [12].

While our results extend the regime of λ for which we can approximately count independent sets in general graphs and in which decay of correlations in known to hold, the new regime also includes improvements for the following two interesting specific cases. The first is when $\lambda = 1$, i.e., when all the independent sets are equally weighted. The previously known bound [7, 23] gives efficient approximate counting of uniformly weighted independent sets for graphs of maximum degree 4. Our bound extends this to graphs of maximum degree 5. Notice that in some respect this is tight since there are graphs [6] of maximum degree 6 for which any "local" algorithm for approximately sampling a uniform independent set must take exponential time. ([6] also establishes NP-hardness of approximation for graphs of maximum degree 25.) Notice also that counting independent sets of graphs of maximum degree 4 (exactly) is #P-complete, and to the best of our knowledge, our algorithm is the first to give a *deterministic* fully polynomial approximation scheme (FPTAS) for a "natural" #P-complete problem. (See [11] for another recent example, where a deterministic FPTAS is established for a different #P-hard problem).

The second important example affected by our results is the integer lattice $\mathbb{Z}^2$, which is probably the most interesting graph to statistical physicists. Simulations suggest that the critical activity for decay of correlations on $\mathbb{Z}^2$ is around 3.79. However, the rigorously known bounds are far from that. The previous general bound [7, 23] gives decay of correlations and efficient approximate counting (by counting independent sets of $\mathbb{Z}^2$ we mean counting independent sets of finite sub-squares of $\mathbb{Z}^2$) for $\lambda \leq 1$. A better bound (for $\mathbb{Z}^2$) is achieved by a percolation argument [2] yielding decay of correlations and hence rapid mixing of the Glauber dynamics for $\lambda < \frac{p_c}{1-p_c}$, where p_c is the critical probability for site percolation. Plugging in the best rigorous bound [1] on p_c for $\mathbb{Z}^2$ gives the regime $\lambda < 1.255$. A different approach specific to $\mathbb{Z}^2$ is based on calculating whether the *Dobrushin-Shlosman condition* holds, a condition on the distribution in a finite subset that implies decay of correlations on the infinite lattice. A series of results [5, 14, 18] improved the bound on λ by calculating the Dobrushin-Shlosman condition for larger and larger rectangles, yielding the previously best known bound of $\lambda < 1.508$, resulting from a calculation of the distribution on a 6×6 rectangle. Our results significantly improve on all the above bounds for $\mathbb{Z}^2$ and extend the regime to $\lambda < 27/16 = 1.6875$.

Another important implication of our results is that since they match the limit imposed by the tree, any further improvement on the regime of λ for which approximate counting is known to be contractible must either not imply decay of correlations (i.e., cannot only rely on arguments of "local" nature), or be specific to a certain graph (or class of graphs) making use of some structural properties other than just the maximum degree of the graph. Finally, we stress that the main argument in which we compare a general graph to a tree is based on an elegant (and rather simple) recursive procedure, i.e., it is an argument natural to computer scientists. Thus, this work serves

as another example where techniques from theoretical computer science are used to solve open problems in other mathematical fields (in this case, statistical physics).

2 Preliminaries and statement of results

Let $G = (V, E)$ be a graph and $\lambda > 0$ an activity parameter. We are interested in counting (or sampling) independent sets of G where the weight of each independent set $I \subset V$ is $\lambda^{|I|}$. Specifically, letting $Z \equiv Z_G^\lambda = \sum_I \lambda^{|I|}$, where the summation is over independent sets I of G, we are interested in calculating Z or sampling from the distribution in which the probability of outputting the independent set I is $\frac{\lambda^{|I|}}{Z}$. In statistical physics the above is referred to as the hard-core model with activity λ and Z is called the partition function. Notice that if we set $\lambda = 1$ then Z simply counts the number of independent sets of G (the weight of each independent set is 1).

Most of our analysis will be concerned with the marginals of the above distribution at a single vertex, i.e., the probability that a vertex v is in the chosen independent set. We refer to the latter event as v being occupied and denote its probability by

$$p_v \equiv p_{G,v}^\lambda = \frac{\sum_{I \ni v} \lambda^{|I|}}{Z_G^\lambda} .$$

As mentioned in the introduction and as we shall further see below, efficiently approximating Z is closely related to the distribution over independent sets exhibiting decay of correlations with distance, where by the latter we mean that conditioning on whether a subset of vertices Λ is in the independent set or not has little influence on p_v when v and Λ are far from each other. Let $\Lambda \subset V$ be a subset of V and $\sigma_\Lambda \in \{0,1\}^\Lambda$ a configuration over the vertices in Λ, i.e., for $u \in \Lambda$, $\sigma_\Lambda(u) = 1$ indicates that u is occupied and $\sigma_\Lambda(u) = 0$ indicates that it is not. We write $p_v^{\sigma_\Lambda}$ for the probability that v is occupied conditioned on the configuration in Λ being fixed as specified by σ_Λ. We will only consider configurations σ_Λ which specify independent sets of Λ so the above conditional probability is well defined. Notice that since the distribution is over independent sets then the above conditional probability is the same as the unconditional probability on a smaller graph, the one resulting from deleting all the vertices in Λ as well as the neighbors of u for every u such that $\sigma_\Lambda(u) = 1$. We are now ready to define the notions of decay of correlations we will use.

Definition 2.1 *Let $\delta : \mathbb{N} \to \mathbb{R}^+$. We say that the distribution over independent sets of $G = (V, E)$ with activity parameter λ exhibits* weak spatial mixing *with rate $\delta(\cdot)$ if and only if for every $v \in V$, $\Lambda \subset V$, and any two configurations $\sigma_\Lambda, \tau_\Lambda$ specifying independent sets of Λ,*

$$|p_v^{\sigma_\Lambda} - p_v^{\tau_\Lambda}| \leq \delta(\mathrm{dist}(v, \Lambda)),$$

where $\mathrm{dist}(v, \Lambda)$ stands for the graph distance (the length of the shortest path) between the vertex v and the subset Λ.

In statistical physics the graph G is usually an infinite graph (such as the square integer lattice $\mathbb{Z}^2$) and weak mixing with rate δ that goes to zero is equivalent to the uniqueness of the *Gibbs* measure, i.e., to the existence of a unique macroscopic equilibrium.

Definition 2.2 *Let $\delta : \mathbb{N} \to \mathbb{R}^+$. We say that the distribution over independent sets of $G = (V, E)$ with activity parameter λ exhibits* strong spatial mixing *with rate $\delta(\cdot)$ if and only if for every $v \in V$, $\Lambda \subset V$ and any two configurations $\sigma_\Lambda, \tau_\Lambda$ specifying independent sets of Λ,*

$$|p_v^{\sigma_\Lambda} - p_v^{\tau_\Lambda}| \leq \delta(\mathrm{dist}(v, \Delta)),$$

where $\Delta \subseteq \Lambda$ stands for the subset on which σ_Λ and τ_Λ differ.

Let us briefly discuss the difference between the two kinds of spatial mixing. First, notice that by definition strong spatial mixing is indeed stronger, i.e., strong spatial mixing with rate δ implies weak spatial mixing with rate δ. The difference between the two is that in the strong mixing definition we are allowed to fix vertices that are close to v as long as we fix them to the same value in both σ and τ. One might be tempted to conclude that weak spatial mixing implies strong spatial mixing since fixing additional vertices to the same value in both σ and τ should only decrease the influence of vertices in Δ. However, this intuition is not generally true and the prime counter example is the *ferromagnetic Ising model* at appropriate temperature and positive external field on appropriate graphs. Roughly speaking, the reason for these counter examples is that fixing vertices close to v may shift p_v to a regime where it is more sensitive to the configuration in Δ.

Remark: In the literature, strong spatial mixing is usually referred to as what we call strong spatial mixing with exponential decay, i.e., with $\delta(\ell) = C\exp(-\alpha\ell)$ for some positive constants C and α. Also, it is often the case in those definitions that Δ is required to consist of a single vertex. The single-vertex definition with exponential decay is equivalent to the one with Δ of arbitrary size when the graph grows sub-exponentially (as integer lattices, for example), but is less meaningful otherwise (the single-vertex definition of strong spatial mixing does not even imply weak spatial mixing in exponentially-growing graphs). We take the same approach as in the author's thesis [24] by allowing Δ to be of arbitrary size, an approach which defines a stronger yet more useful condition when considering general graphs.

Our analysis is concentrated on comparing the sensitivity of the value at a vertex v in a general graph G (to conditions on other vertices in G, as in the spatial-mixing definitions above) with the sensitivity of the value at the root of the regular tree (to conditions on other vertices in the tree). Let $\widehat{\mathbb{T}}^b$ be the infinite regular tree where each vertex has degree $b+1$. Notice that if we root this tree at any particular vertex then the root has $b+1$ children, while the rest of the vertices have b children each. This graph is usually referred to as the *Bethe lattice* or *Cayley tree*. (Later on we will also need to refer to the rooted infinite regular tree in which the root has b children as do the rest of the vertices. The latter tree will be denoted as $\mathbb{T}^b$.) Our two main technical results are the following.

Theorem 2.3 *For every positive integer b and any λ, if $\widehat{\mathbb{T}}^b$ with activity λ exhibits strong spatial mixing with rate δ then, with the same activity λ, every graph of maximum degree $b+1$ exhibits strong spatial mixing with rate δ.*

Theorem 2.4 *For every positive integer b and any λ, if $\widehat{\mathbb{T}}^b$ with activity λ exhibits weak spatial mixing with rate $\delta(\cdot)$ then it also exhibits strong spatial mixing with rate $\frac{(1+\lambda)(\lambda+(1+\lambda)^{b+1})}{\lambda}\delta(\cdot)$.*

Remark: Theorem 2.3 is not specific to the independent-sets model. As will be apparent from its proof below, it applies to any model in which each vertex is assigned one of two values (or spins), e.g., Ising models. On the other hand, Theorem 2.4 is not as general since, for example, the ferromagnetic Ising model

on the regular tree at appropriate temperature and positive external field exhibits weak spatial mixing with a decaying rate but not strong spatial mixing. However, the crucial property for Theorem 2.4 seems to be the antiferromagnetic nature of the interaction between neighboring vertices. Indeed, in the full version of the paper we show that the analogous version of Theorem 2.4 also holds for the antiferromagnetic Ising model (with any external field).

The regime of λ for which $\widehat{\mathbb{T}}^b$ exhibits weak spatial mixing with a rate that goes to zero (uniqueness of the Gibbs measure) is well known [13, 21]:

Proposition 2.5 *For every positive integer b, $\widehat{\mathbb{T}}^b$ with activity parameter λ exhibits weak spatial mixing with a rate δ that goes to zero if and only if $\lambda \leq \lambda_c(b) = \frac{b^b}{(b-1)^{b+1}}$. If the inequality is strict then the rate δ can be taken to go to zero exponentially fast (with constants that depend on b and λ).*

Corollary 2.6 *For every positive integer b and any $\lambda \leq \lambda_c(b)$ there exists a decaying rate δ such that for every graph G of maximum degree $b+1$, G with activity λ exhibits strong spatial mixing with rate δ. (In particular, for any graph of maximum degree $b+1$, the Gibbs measure is unique for $\lambda \leq \lambda_c(b)$.) Furthermore, the rate δ can be taken to decay exponentially fast (with constants that depend on b and λ) if $\lambda < \lambda_c(b)$.*

Notice that the threshold for decay of correlations given in Corollary 2.6 is tight since, as mentioned in Proposition 2.5, $\widehat{\mathbb{T}}^b$ is an example of a graph that does not exhibit spatial mixing with any decaying rate when $\lambda > \lambda_c(b)$. Also, the corollary proves Conjecture 2.1 in [20]. (This conjecture considers a model in which each vertex is associated with its own activity parameter and the regime asserted is the one in which the maximum activity is $\leq \lambda_c(b)$. It can easily be seen from our proofs below that Corollary 2.6 holds in this scenario as well.)

As already mentioned, establishing strong spatial mixing has algorithmic implications. The reason for this is that then the distribution over independent sets is very local in nature and this suggests that local algorithms that approximate the partition function should work well. In fact, the proof of Theorem 2.3 implicitly includes an algorithm of this kind.

Theorem 2.7 *There exists a deterministic algorithm such that for every integer b and any λ, if $\widehat{\mathbb{T}}^b$ with activity λ exhibits strong spatial mixing with rate $\delta(\ell) = O(\exp(-\alpha\ell))$ for some $\alpha > 0$ then on input of any graph G of maximum degree $b+1$ the algorithm approximates the partition function Z_G^λ to within a factor $(1+\epsilon)$ in time $O((\frac{(1+\epsilon)n}{\epsilon})^{1+(\ln b)/\alpha})$, where $n = |V|$. (The algorithm outputs two numbers Z_1, Z_2 such that $Z_1 \leq Z_G^\lambda \leq Z_2$ and $Z_2 \leq (1+\epsilon)Z_1$.) Similarly, there is a randomized algorithm that under the same condition (and with the same running time) generates independent sets of G where for every independent set I, the probability that the algorithm outputs I is within a factor $(1 \pm \epsilon)$ from $\frac{\lambda^{|I|}}{Z_G^\lambda}$.*

Notice that Theorem 2.7 requires that $\widehat{\mathbb{T}}^b$ exhibits strong spatial mixing rather than the graph G, the independent sets of which the algorithm counts. (Strong spatial mixing on $\widehat{\mathbb{T}}^b$ may be a stronger condition than strong spatial mixing on G.)

Corollary 2.8 *There exists a deterministic algorithm such that for every positive integer b and any $\lambda < \lambda_c(b)$, on input of any graph G of maximum degree $b+1$, the algorithm approximates the partition function Z_G^λ to within a factor $(1+\epsilon)$ in time polynomial in $\frac{(1+\epsilon)n}{\epsilon}$, where the exponent*

depends on b and λ. Similarly, there is a randomized algorithm that for the same choice of parameters (and with the same running time) generates independent sets of G where for every independent set I, the probability that the algorithm outputs I is within a factor $(1 \pm \epsilon)$ from $\frac{\lambda^{|I|}}{Z_G^\lambda}$.

An existing and well-known algorithm for sampling independent sets (and by a standard reduction, for approximating the partition function Z) is the *Glauber* dynamics. This dynamics is a local Markov chain over independent sets of G where in each time step a vertex v of G is chosen uniformly at random and its value it updated conditioned on its neighbors, i.e., if one or more of its neighbors is occupied then v is deterministically set to be unoccupied while if all the neighbors are unoccupied then v is set to be occupied with probability $\frac{\lambda}{1+\lambda}$ and unoccupied otherwise. It is straightforward to see that this chain indeed converges to the desired distribution over independent sets of G, but for some graphs and values of λ the mixing time (the number of steps required before the chain is within a "small" variation distance from the stationary distribution) is exponential [6] in the size of G. However, it is generally believed that if strong spatial mixing with exponential decay holds for G and λ then the Glauber dynamics should mix in polynomial time. Indeed, the latter implication is known to hold when the graph G grows "slowly enough", with the prime example of such graphs being the integer lattices $\mathbb{Z}^d$. Connections as above between strong spatial mixing and the rate of convergence of the Glauber dynamics were the subject of a number of papers in Statistical Physics [22, 16, 4] as well as in Computer Science [8, 24]. The following is a partial summary that is sufficient for the discussion here.

Theorem 2.9 *If G is a graph that grows subexponentially and if G with activity parameter λ exhibits strong spatial mixing with exponential decay then the mixing time of the Glauber dynamics for sampling independent sets of G with activity λ is $O(n^2)$.*

Corollary 2.10 *If G is a graph that grows subexponentially and has maximum degree $b+1$ then the Glauber dynamics on G mixes in time $O(n^2)$ for any $\lambda < \lambda_c(b)$.*

Remark: In stating Theorem 2.9 we allowed ourselves a minor inaccuracy. Strictly speaking, the arguments in [8] and [24] require that the graph grows polynomially. To get rapid mixing for graphs that grow with any subexponential rate we need to establish a slightly stronger property than strong spatial mixing involving the existence of a coupling with certain properties (see [10] for an example). However, an extension of our results above shows that this stronger property holds throughout the regime $\lambda < \lambda_c(b)$. We do not delve on these details because our main focus in this paper is establishing an efficient approximation scheme for counting independent sets for general graphs and $\lambda < \lambda_c(b)$, which we did in Corollary 2.8, rather than an exact analysis of the performance of the Glauber dynamics. Nevertheless, since the performance of the Glauber dynamics is interesting in its own right we do mention the implications of our results on it.

As mentioned in the introduction, the previously best bound [7, 23] on λ for which approximate counting is known to be contractible is $\lambda \leq \frac{2}{b-1}$, where rapid mixing of the Glauber dynamics is established for arbitrary graphs of maximum degree $b+1$ and any λ in this regime. Thus, Corollary 2.8 extends the regime for which efficient approximate counting is known to exist for arbitrary graphs. In addition, Corollary 2.10 extends the previous regime for which the Glauber dynamics is known to mix in polynomial time, but only for graphs with subexponential growth.

3 A (self-avoiding-walk) tree representation: proof of Theorem 2.3

In this section we present a novel procedure for calculating the probability p_v that a vertex $v \in G$ is occupied. As we show below, the calculation carried out by this procedure is exactly the same as the one carried out when calculating the probability that the root of an appropriate tree is occupied. Thus, the probability that v is occupied equals the probability that the root of this tree is occupied, and this fact will immediately prove Theorem 2.3.

In order the describe the tree corresponding to $G = (V, E)$ and $v \in V$ we need to refer to an ordering on the neighbors of each vertex in G. Thus, for every vertex $u \in V$ we fix an (arbitrary) enumeration of the edges incident to u. From here onwards, whenever we say that an edge $\{u, w\}$ (or a neighbor w of u) is larger than $\{u, x\}$ (or a neighbor x of u) we interpret this according to the enumeration of the edges incident to u that we fixed here. The description of the tree corresponding to (G, v) includes the tree (as a graph) together with a specification of a subset of its leaves that are fixed to specific values (occupied or unoccupied). As we mentioned before, fixing the values of certain vertices is equivalent to considering a smaller tree with no fixed vertices, but the first description will be more convenient to work with in the proofs below.

We denote the tree corresponding to (G, v) as $T_{\text{saw}}(G, v)$. This tree is essentially the tree of self-avoiding walks originating at v except that the vertices closing a cycle are also included in the tree and are fixed to be either occupied or unoccupied. Specifically, $T_{\text{saw}}(G, v)$ is defined as the tree of all paths originating at v, except that whenever a path closes a cycle the copy (in the tree) of the vertex closing the cycle (in G) is fixed to occupied if the edge closing the cycle is larger than the edge starting the cycle and unoccupied otherwise, with the rest of the path ignored. See Figure 1 for an example of a small graph and its corresponding tree $T_{\text{saw}}(G, v)$. We note that trees of self-avoiding walks have appeared before (explicitly or implicitly) in arguments for establishing decay of correlations (see, e.g., [9] and [10]), and our construction was partially inspired by these references. Also, after completing this manuscript the author has learned that a tree equivalent to $T_{\text{saw}}(G, v)$ has already appeared in [19], though in a somewhat different context and without suggesting the kind of result established in Theorem 2.3.

The crucial point of the correspondence we establish below between the probability p_v that v is occupied and the probability that the root of $T_{\text{saw}}(G, v)$ is occupied is that it continues to hold when we impose an arbitrary condition on any subset of the vertices of G (and the corresponding condition on the tree). Notice that there is a natural way to correspond a condition on G with a condition on $T_{\text{saw}}(G, v)$. Specifically, since every vertex in the tree $T_{\text{saw}}(G, v)$ corresponds to a vertex in G in a natural way, if a condition on G fixes the vertex u to a certain value, the corresponding condition on $T_{\text{saw}}(G, v)$ fixes all the copies of u to the same value. Whenever we fix a condition on $T_{\text{saw}}(G, v)$ that corresponds to a condition on G as above then for each fixed vertex x in $T_{\text{saw}}(G, v)$ we also erase the subtree underneath x in order to make sure the resulting condition is well defined. (Notice that in any case, conditioned on the value at x, the value at the root is independent of the configuration on the subtree underneath x.) An alternative description of the resulting tree is to first fix the condition on G and then construct the corresponding tree of paths $T_{\text{saw}}(G, v)$ so that whenever a path visits a fixed vertex, the value of that vertex is copied to the tree and the path is not continued further, i.e., the corresponding fixed vertex in the tree is a leaf.

Theorem 3.1 *For every graph $G = (V, E)$, any λ, every $\Lambda \subset V$ and any configuration σ_Λ,*

$$p_v^{\sigma_\Lambda} \;=\; \mathbb{P}_v^{\sigma_\Lambda},$$

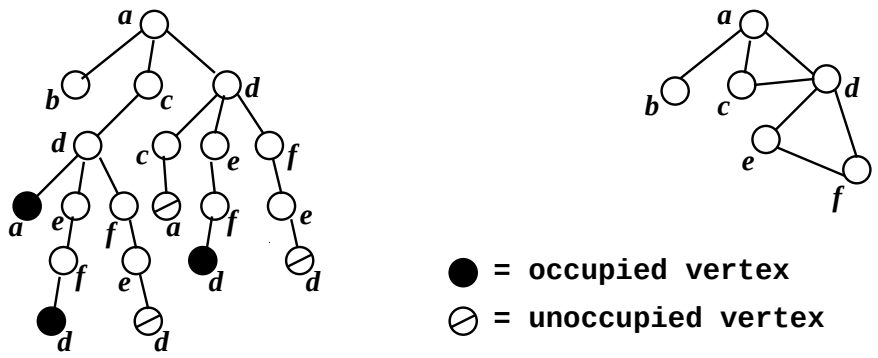

Figure 1: The construction of T_{saw}. The tree on the left is $T_{\text{saw}}(G, a)$, where G is the graph on the right and where the order on the neighbors of each vertex in G is lexicographic. In order to better illustrate the construction we labeled each vertex in the tree with the name of its corresponding vertex in G. Notice that vertices that close cycles are fixed to be either occupied or unoccupied. For example, the bottom-left copy of d is fixed to be occupied because the edge $\{d, f\}$ that closes the cycle is larger than the edge $\{d, e\}$ that starts it.

where $\mathbb{P}_v^{\sigma_\Lambda} \equiv \mathbb{P}_{G,v}^{\sigma_\Lambda}(\lambda)$ stands for the probability that the root of $T_{\text{saw}}(G, v)$ is occupied when imposing the condition corresponding to σ_Λ as described above.

Remark: Notice that $T_{\text{saw}}(G, v)$ has two types of fixed vertices. The first type is a "structural" one: these fixed vertices arise from the cycle structure of the graph G (can be thought of as expressing the influence a vertex has on itself through the cycle), and both their composition and values are independent of the condition imposed on G. Fixed vertices of the second type are those that correspond to fixed vertices in G and the values they are fixed to are simply copied from their corresponding vertices in G. Also, although it may seem that a fixed vertex in $T_{\text{saw}}(G, v)$ may be assigned two conflicting values (if it is of both types), this is cannot happen since a structural fixed vertex that corresponds to a vertex u in G always has an ancestor that also corresponds to u, and therefore, if u is fixed in G then the ancestor is fixed in $T_{\text{saw}}(G, v)$, and thus the subtree underneath the ancestor is erased. (An alternative way to see this is that a fixed vertex in G can never be part of a cycle since in the construction of the tree of paths, the path ends whenever visiting a fixed vertex.)

Before going on to prove Theorem 3.1, we note that Theorem 2.3 follows almost immediately from it. To see this, we simply observe that Theorem 3.1 gives that $|p_v^{\sigma_\Lambda} - p_v^{\tau_\Lambda}| = |\mathbb{P}_v^{\sigma_\Lambda} - \mathbb{P}_v^{\tau_\Lambda}|$, and that for any subset Δ of vertices of G, $\text{dist}(v, \Delta)$ is exactly the same as the distance between the root of $T_{\text{saw}}(G, v)$ and the subset of vertices of the tree composed of the copies of vertices in Δ. (This is because paths in the tree correspond to paths in G.) Thus, when we impose the two conditions corresponding to σ_Λ and τ_Λ respectively, we in fact impose two conditions that differ on a subset of the vertices of $T_{\text{saw}}(G, v)$ whose distance from the root is exactly $\text{dist}(v, \Delta)$, where Δ is the subset of vertices of G on which σ_Λ and τ_Λ differ, and where we used the fact that the values of the structural fixed vertices of $T_{\text{saw}}(G, v)$ do not depend on the condition we impose of G. The only remaining gap from the statement of Theorem 2.3 is that the latter considers the regular infinite tree $\widehat{\mathbb{T}}^b$, while we consider $T_{\text{saw}}(G, v)$. Notice, however, that $T_{\text{saw}}(G, v)$ is a subtree of $\widehat{\mathbb{T}}^b$ since the degree of every vertex in $T_{\text{saw}}(G, v)$ is at most the degree of the corresponding vertex in G. Furthermore, since fixing a vertex in the tree to be unoccupied has the same effect (on the probability of occupation at the root) as erasing the subtree rooted at this vertex, $T_{\text{saw}}(G, v)$ can be considered as $\widehat{\mathbb{T}}^b$ with additional vertices being fixed to be unoccupied. This completes the proof of Theorem 2.3 assuming Theorem 3.1, and we thus continue with the proof of the latter.

Proof of Theorem 3.1: As we already mentioned, the proof of the theorem is based on a novel procedure for calculating p_v, one that is essentially the same as a procedure to calculate the probability of occupation at the root of $T_{\mathrm{saw}}(G,v)$. In order to describe these procedures, it will be convenient to make a change of variable and work with ratios of probabilities rather than the probability of occupation itself. We thus write $R_v \equiv R_{G,v}(\lambda) = p_v/(1-p_v)$ for the ratio between the probabilities that v is occupied and unoccupied, respectively. (This notation also applies when we impose conditions, i.e., $R_v^{\sigma_\Lambda}$ stands for the ratio of the two probabilities under the condition σ_Λ.) Notice that we allow $p_v = 1$, in which case we set $R_v = \infty$.

We now describe a standard procedure for calculating the probability of occupation at the root of any given tree. Let T be a rooted tree, $\Lambda \subset T$ a subset of its vertices, and σ_Λ a configuration of Λ. Write $R_T^{\sigma_\Lambda}$ for the ratios of the probabilities that the root is occupied and unoccupied, respectively, when imposing the condition σ_Λ. The lack of cycles on a tree means that once we fix the value at the root, the configurations of the subtrees rooted at the children of the original root are all independent of each other. A trivial calculation then gives that

$$R_T^{\sigma_\Lambda} \;=\; \lambda \prod_{i=1}^{d} \frac{1}{1+R_{T_i}^{\sigma_{\Lambda_i}}}, \tag{1}$$

where d is the number of children of the root, T_i is the subtree rooted at the i-th child, $\Lambda_i = \Lambda \cap T_i$, and σ_{Λ_i} is the restriction of σ to Λ_i. Notice that (1) defines a recursive procedure for calculating $R_T^{\sigma_\Lambda}$ once we observe that the base cases are either when $v \in \Lambda$, in which case $R_v = \infty$ or $R_v = 0$ (depending on whether v is fixed to be occupied or unoccupied), or when v has no children (and is not fixed), in which case $R_v = \lambda$. Indeed, this simple recursive calculation has been widely used for analyzing the distribution of independent sets of trees (see, e.g., [13, 17]), and in particular, Proposition 2.5 is obtained [13] by analyzing the fixed points of (1).

We now go on to describe our novel procedure for calculating the probability of occupation at v in general graphs. Our goal is to mimic (1), i.e., to write R_v in terms of ratios R_{u_i}, where u_i varies over the neighbors of v. This will give us a recursive procedure to calculate R_v similar to the one described for trees. The problem is that now the values at the different neighbors u_i may depend on each other even when we fix the value at v, and hence we cannot get a clean product as in (1). However, by imposing appropriate conditions (in fact, a different condition for the contribution of each neighbor), we do get a similar product expression. In order to describe the generalization of (1) to general graphs we need additional notation. With v being the vertex for which we wish to calculate R_v, let G' be the same as G except that v is replaced by d vertices $v_1,\ldots,v_d$, where d is the degree of v. Each v_i has a single edge connecting it to u_i, where u_i is the i-th neighbor of v in G, and the order on the neighbors of v is the same as the one used in the definition of $T_{\mathrm{saw}}(G,v)$. In addition, we associate with each of the vertices v_i the activity $\lambda^{1/d}$ rather than λ (i.e., when v_i is included in the independent set it contributes a factor $\lambda^{1/d}$ to the weight rather than λ). Now, it is easy to see that an independent set in G' in which all the v_i are unoccupied has the same weight as the corresponding independent set of G (with v unoccupied), and similarly when all the v_i are occupied (with v occupied in G). In particular, $R_{G,v}^{\sigma_\Lambda}$ is exactly the ratio between the probabilities in G' that all the v_i are occupied and all are unoccupied, respectively, conditioned on σ_Λ. Writing the latter ratio as a telescopic product gives that

$$R_{G,v}^{\sigma_\Lambda} \;=\; \prod_{i=1}^{d} R_{G',v_i}^{\sigma_\Lambda \tau_i},$$

where $\sigma_\Lambda \tau_i$ stands for the concatenation of the two configurations σ_Λ and τ_i, and where τ_i is the configuration of the v_j for $j \neq i$ in which the values are fixed to occupied for $j < i$ and to unoccupied for $j > i$. Intuitively, we can think of the above splitting of v to d one-degree copies as a step that cancels cycles, with the conditions τ_i expressing the influence of v on itself through cycles in G. Indeed, if v is not on any cycle then the different copies v_i are on different clusters and thus the condition τ_i has no influence on the probability that v_i is occupied.

Now, since v_i is connected only to u_i in G' then it is easy to see that

$$R^{\sigma_\Lambda \tau_i}_{G',v_i} = \frac{\lambda^{1/d}}{1 + R^{\sigma_\Lambda \tau_i}_{(G' \setminus v_i),u_i}} ,$$

and hence,

$$R^{\sigma_\Lambda}_{G,v} \;=\; \lambda \prod_{i=1}^{d} \frac{1}{1 + R^{\sigma_\Lambda \tau_i}_{(G' \setminus v_i),u_i}} \;. \tag{2}$$

Notice that (2) defines a recursive procedure for calculating $R_{G,v}$ in the same manner that (1) defines such a procedure for the tree (the base cases are the same as in the tree). To see that the procedure for G terminates, observe that although the number of vertices may increase down the recursion, the number of unfixed vertices reduces by one in each step since in the calculation of $R_{(G' \setminus v_i),u_i}$ all the v_j are either fixed (if $j \neq i$) or erased from the graph (if $j = i$).

We go on to show that the above procedure for calculating $R^{\sigma_\Lambda}_{G,v}$ gives exactly the same result as running the tree procedure for $T_{\mathrm{saw}}(G, v)$ with the condition corresponding to σ_Λ imposed on it. Notice that the calculation carried out (as a function of the values returned by the recursive calls) is exactly the same in the two recursive equations (1) and (2). Now, since the stopping rules are the same for both procedures, in order to complete the proof by induction, it is enough to show that the tree $T_{\mathrm{saw}}(G' \setminus v_i, u_i)$ with the condition corresponding to $\sigma_\Lambda \tau_i$ imposed on it is exactly the same as the subtree of $T_{\mathrm{saw}}(G, v)$ rooted at the i-th child of the original root with the condition corresponding to σ_Λ imposed on it. Establishing the latter is enough because then, by induction, the values returned by the recursive calls are the same for both procedures. It is easy to observe that the two trees are indeed the same since both are in fact the tree of paths in G starting at u_i, except that whenever v is visited, the corresponding vertex in the tree is fixed to either occupied or unoccupied depending on whether the path reached v from a neighbor smaller or greater than i. This completes the proof of Theorem 3.1. □

4 Monotonicity on trees: proof of Theorem 2.4

In this section we prove that on a regular tree weak spatial mixing implies strong spatial mixing. In fact, we prove a stronger statement regarding a certain monotonicity in the activity λ. In order to state this result we need to extend our model and consider the regular tree $\widehat{\mathbb{T}}^b$ with a vector of activities $\vec{\lambda}$, where each vertex v of $\widehat{\mathbb{T}}^b$ is associated with its own activity $\vec{\lambda}(v)$ (so the weight of an independent set I is $\prod_{v \in I} \vec{\lambda}(v)$). Notice that strong spatial mixing on the regular tree with activity λ is equivalent to weak spatial mixing on the same tree for all assignments of activities $\vec{\lambda}$ in which for every v, $\vec{\lambda}(v) = \lambda$ or $\vec{\lambda}(v) = 0$. To see this, notice that w.l.o.g. we can assume that the two configurations σ_Λ and τ_Λ in the definition of strong spatial mixing set the configuration in $\Lambda \setminus \Delta$ (the subset on which they agree) to all unoccupied (since fixing a vertex to be occupied is

the same as fixing its neighbors to be unoccupied). Now, notice that setting $\vec{\lambda}(v) = 0$ has the same effect as fixing v to be unoccupied, so comparing the two conditions σ_Λ and τ_Λ is exactly the same as comparing the two restrictions of these conditions to Δ when the activities in $\Lambda \setminus \Delta$ are set to zero. Thus, in order to prove Theorem 2.4 is it is enough to show that setting an arbitrary subset of the activities to zero only decreases the sensitivity to conditioning on Δ. We will in fact show that *any* decrease in the activities reduces this sensitivity.

As before, it will be more convenient to work with ratios of probabilities $R = \frac{p}{1-p}$. Recall that we are looking to bound the sensitivity of the ratio R_v to conditions set at distance ℓ from v. First, w.l.o.g. we will consider $\widehat{\mathbb{T}}^b$ to be rooted at v and analyze the sensitivity of the root to various conditions, omitting v from the notation. Notice that, since the tree is a bipartite graph, in order to bound the sensitivity of the value at the root to conditions on a subset at distance ℓ below it is enough to consider the two conditions in which all the vertices at level $\ell + 1$ are set to all occupied and all unoccupied, respectively. This is because these are the conditions that minimize and maximize the probability that the root is occupied, respectively. Let $R^E_\ell \equiv R^E_\ell(\vec{\lambda})$ stand for the ratio at the root conditioned on the configuration in which all the vertices at distance ℓ from the root are set to occupied if ℓ is even and to all unoccupied if ℓ is odd (so R^E_ℓ maximizes this ratio among conditions at distance ℓ from the root). Let R^O_ℓ stand for the ratio at the root conditioned on the negation of the configuration above (so R^O_ℓ minimizes this ratio among conditions at distance ℓ from the root). The main result of this section reads as follows.

Theorem 4.1 *Fix an arbitrary $\lambda \geq 0$. Let $\vec{\lambda}$ be an assignment of activities to the vertices of $\widehat{\mathbb{T}}^b$ such that $0 \leq \vec{\lambda}(v) \leq \lambda$ for every $v \in \widehat{\mathbb{T}}^b$. Then, for every ℓ,*

$$\frac{R^E_\ell(\vec{\lambda})}{R^O_\ell(\vec{\lambda})} \;\leq\; \frac{R^E_\ell(\lambda)}{R^O_\ell(\lambda)} \;. \tag{3}$$

Theorem 2.4 follows from Theorem 4.1 as explained above and once we notice that $\frac{R^E}{R^O} - 1 = \frac{p^E - p^O}{p^O(1-p^E)}$ and that for $\ell \geq 2$, $p^E_\ell(\lambda) \leq \frac{\lambda}{1+\lambda}$ and $p^O_\ell(\lambda) \geq \frac{\lambda}{\lambda+(1+\lambda)^{b+1}}$.

Theorem 4.1 may come as a bit of a surprise since the monotonicity in $\vec{\lambda}$ does not hold in general, i.e., (3) does not necessarily hold if we replace the uniform assignment λ with a general assignment $\vec{\lambda}'$ that dominates $\vec{\lambda}$. See [3] for an example of such non-monotone behavior. Indeed, the proof of Theorem 4.1 requires somewhat delicate arguments that make use of certain monotone properties of the distribution under a uniform assignment λ.

In order to prove Theorem 4.1 it will be convenient to consider the slightly modified tree $\mathbb{T}^b$, where the root has b children rather than $b + 1$. (The rest of the vertices have b children each, as in $\widehat{\mathbb{T}}^b$.) At the end we will establish the theorem for $\widehat{\mathbb{T}}^b$ as well. Let R^E_ℓ and R^O_ℓ be defined as above, except that now they describe the ratios at the root of $\mathbb{T}^b$. We will show the following claim, which adds on what is claimed in Theorem 4.1. (The addition is needed in order to carry the induction).

Lemma 4.2 *For every integer $\ell \geq 1$ and any assignment of activities $0 \leq \vec{\lambda} \leq \lambda$ to the vertices of $\mathbb{T}^b$, the following two inequalities hold:*

$$\frac{R^E_\ell(\vec{\lambda})}{R^O_\ell(\vec{\lambda})} \;\leq\; \frac{R^E_\ell(\lambda)}{R^O_\ell(\lambda)}\,; \tag{4}$$

$$\frac{1 + R^E_\ell(\vec{\lambda})}{1 + R^O_\ell(\vec{\lambda})} \;\leq\; \frac{1 + R^E_\ell(\lambda)}{1 + R^O_\ell(\lambda)}\,. \tag{5}$$

Proof: We first give some context by noticing that (4) implies (5) when the ratios w.r.t. $\vec{\lambda}$ are smaller than those w.r.t. λ and (5) implies (4) when the ratios w.r.t. $\vec{\lambda}$ are the larger ones. Let us also clarify two special cases in which one or more of the ratios is zero. If $R^E_\ell(\vec{\lambda}) = R^O_\ell(\vec{\lambda}) = 0$ (this can only happen if the activity assigned to the root is zero) then we let the ratio $\frac{R^E_\ell(\vec{\lambda})}{R^O_\ell(\vec{\lambda})} = 1$. If $R^O_\ell(\lambda) = 0$ (this happens only when $\ell = 1$, i.e., the fixed vertices are the children of the root) then $R^O_\ell(\vec{\lambda}) = 0$ as well. In this case we set $\frac{R^E_\ell(\vec{\lambda})}{R^O_\ell(\vec{\lambda})} = \frac{R^E_\ell(\vec{\lambda})}{R^O_\ell(\vec{\lambda})} = \infty$. We note that for both cases our choices are valid since the only way in which we will use the fact that $\frac{R^E_\ell(\vec{\lambda})}{R^O_\ell(\vec{\lambda})} \leq \frac{R^E_\ell(\vec{\lambda})}{R^O_\ell(\vec{\lambda})}$ (as an induction hypothesis) is in that there exists $\alpha \geq 0$ such that $R^O_\ell(\vec{\lambda}) = \alpha R^O_\ell(\lambda)$ and $R^E_\ell(\vec{\lambda}) \leq \alpha R^E_\ell(\lambda)$. Indeed, this holds in the two special cases mentioned above.

The proof of the lemma goes by induction on ℓ. For the base case of $\ell = 1$, as we already noticed, $R^O_\ell(\vec{\lambda}) = R^O_\ell(\lambda) = 0$. On the other hand, $R^E_\ell(\lambda) = \lambda$ and in the same manner $R^E_\ell(\vec{\lambda}) = \lambda'$, where λ' is the activity assigned to the root. Since $\lambda' \leq \lambda$ then the statement of the lemma clearly holds.

Assume by induction that the lemma holds for ℓ and all assignments $0 \leq \vec{\lambda} \leq \lambda$, and we will show it holds for $\ell + 1$ and an arbitrary assignment $0 \leq \vec{\lambda} \leq \lambda$. We first mention an elementary fact that we use throughout the proof. For $1 \leq i \leq b$, let $\vec{\lambda}_i$ stand for the restriction of the assignment $\vec{\lambda}$ to the subtree rooted at the i-th child of the root of $\mathbb{T}^b$, which is again an assignment to the vertices of $\mathbb{T}^b$. (The subtree is isomorphic to $\mathbb{T}^b$.) Writing (1) with the notation used here gives that $R^E_{\ell+1}(\vec{\lambda}) = \lambda' \prod_{i=1}^b \frac{1}{1+R^O_\ell(\vec{\lambda}_i)}$, where λ' is the activity that $\vec{\lambda}$ assigns to the root. Similarly, $R^O_{\ell+1}(\vec{\lambda}) = \lambda' \prod_{i=1}^b \frac{1}{1+R^E_\ell(\vec{\lambda}_i)}$.

Now, it is immediate from the second inequality of the induction hypothesis that

$$\begin{aligned}
\frac{R^E_{\ell+1}(\vec{\lambda})}{R^O_{\ell+1}(\vec{\lambda})} &= \prod_{i=1}^b \left(\frac{1 + R^E_\ell(\vec{\lambda}_i)}{1 + R^O_\ell(\vec{\lambda}_i)} \right) \\
&\leq \left(\frac{1 + R^E_\ell(\lambda)}{1 + R^O_\ell(\lambda)} \right)^b \\
&= \frac{R^E_{\ell+1}(\lambda)}{R^O_{\ell+1}(\lambda)} .
\end{aligned}$$

The remaining (main) part of the proof is to show that $\frac{1+R^E_{\ell+1}(\vec{\lambda})}{1+R^O_{\ell+1}(\vec{\lambda})} \leq \frac{1+R^E_{\ell+1}(\lambda)}{1+R^O_{\ell+1}(\lambda)}$. An important fact to keep in mind is that for any non-negative x, x', y, y', if $x' \leq x$ and $1 \leq \frac{x'}{y'} \leq \frac{x}{y}$ then $\frac{1+x'}{1+y'} \leq \frac{1+x}{1+y}$. A first consequence of this fact is that w.l.o.g. we can assume that the activity at the root $\lambda' = \lambda$. This is because the effect of reducing the activity at the root is that the ratios $R^E_{\ell+1}(\vec{\lambda})$ and $R^O_{\ell+1}(\vec{\lambda})$ reduce both by the same factor, and therefore, if the claim holds for some assignment $\vec{\lambda}$ in which $\lambda' = \lambda$ then the claim also holds for the same assignment with a reduced activity at the root. A second consequence is that if $R^E_{\ell+1}(\vec{\lambda}) \leq R^E_{\ell+1}(\lambda)$ then we immediately get that $\frac{1+R^E_{\ell+1}(\vec{\lambda})}{1+R^O_{\ell+1}(\vec{\lambda})} \leq \frac{1+R^E_{\ell+1}(\lambda)}{1+R^O_{\ell+1}(\lambda)}$ because we already showed that $\frac{R^E_{\ell+1}(\vec{\lambda})}{R^O_{\ell+1}(\vec{\lambda})} \leq \frac{R^E_{\ell+1}(\lambda)}{R^O_{\ell+1}(\lambda)}$.

From the above we learn that the only remaining case is when the ratio at the root is larger under $\vec{\lambda}$ than under the uniform assignment λ, or equivalently, when the ratios at the children

are smaller (in some average way) under $\vec{\lambda}$ than under λ. In fact, we now show that w.l.o.g. we can assume that the ratios at the children given the *ODD* condition under $\vec{\lambda}$ are *all* at most $R^O_\ell(\lambda)$. Let $\vec{\lambda}$ be an assignment such that $R^O_\ell(\vec{\lambda}_i) \geq R^O_\ell(\lambda)$ for some $1 \leq i \leq b$, and let $\vec{\lambda}'$ be the same as $\vec{\lambda}$ except that $\vec{\lambda}'_i$ is uniformly equal to λ (i.e., we change the activities at the subtree rooted at the i-th child to be all λ). Notice that by the induction hypothesis, $\frac{1+R^E_\ell(\vec{\lambda}_i)}{1+R^O_\ell(\vec{\lambda}_i)} \leq \frac{1+R^E_\ell(\lambda)}{1+R^O_\ell(\lambda)}$, and therefore $\frac{R^E_{\ell+1}(\vec{\lambda})}{R^E_{\ell+1}(\vec{\lambda})} \leq \frac{R^E_{\ell+1}(\vec{\lambda}')}{R^E_{\ell+1}(\vec{\lambda}')}$ because the ratios at the rest of the children are the same under both assignments. Furthermore, since $R^O_\ell(\vec{\lambda}_i) \geq R^O_\ell(\lambda)$ then $R^E_{\ell+1}(\vec{\lambda}) \leq R^E_{\ell+1}(\vec{\lambda}')$, where again we used the fact that the ratios at the rest of the children are the same under both assignments. As discussed above, the fact that both $\frac{R^E_{\ell+1}(\vec{\lambda})}{R^E_{\ell+1}(\vec{\lambda})} \leq \frac{R^E_{\ell+1}(\vec{\lambda}')}{R^E_{\ell+1}(\vec{\lambda}')}$ and $R^E_{\ell+1}(\vec{\lambda}) \leq R^E_{\ell+1}(\vec{\lambda}')$ implies that $\frac{1+R^E_{\ell+1}(\vec{\lambda})}{1+R^E_{\ell+1}(\vec{\lambda})} \leq \frac{1+R^E_{\ell+1}(\vec{\lambda}')}{1+R^E_{\ell+1}(\vec{\lambda}')}$. Therefore, we can indeed assume w.l.o.g. that $R^O_\ell(\vec{\lambda}_i) \leq R^O_\ell(\lambda)$ for all $1 \leq i \leq b$.

We now continue with establishing that $\frac{1+R^E_{\ell+1}(\vec{\lambda})}{1+R^O_{\ell+1}(\vec{\lambda})} \leq \frac{1+R^E_{\ell+1}(\lambda)}{1+R^O_{\ell+1}(\lambda)}$. For $1 \leq i \leq b$, suppose that $R^O_\ell(\vec{\lambda}_i) = \alpha_i R^O_\ell(\lambda)$ for some $\alpha_i \geq 0$. From the previous paragraph, we know it is enough to consider the case where $\alpha_i \leq 1$ for every i. By the induction hypothesis we know that $R^E_\ell(\vec{\lambda}_i) \leq \alpha_i R^E_\ell(\lambda)$. W.l.o.g. we can assume that $R^E_\ell(\vec{\lambda}_i) = \alpha_i R^E_\ell(\lambda)$ because the only effect of increasing R^E_ℓ while holding everything else fixed is decreasing $R^O_{\ell+1}$, i.e., increasing the ratio $\frac{1+R^E_{\ell+1}}{1+R^O_{\ell+1}}$. Recall that $R^E_{\ell+1}(\vec{\lambda}) = \frac{\lambda}{\prod_{i=1}^b(1+R^O_\ell(\vec{\lambda}_i))} = \frac{\lambda}{\prod_{i=1}^b(1+\alpha_i R^O_\ell(\lambda))}$ (recalling also that we are under the assumption that the activity at the root is λ), with a similar expression for $R^O_{\ell+1}(\vec{\lambda})$. (Since from now on all mentions of R will be of the ratio under of the uniform assignment λ, we drop it from the notation and simply write R^O_ℓ and R^E_ℓ to stand for $R^O_\ell(\lambda)$ and $R^E_\ell(\lambda)$, respectively.) In order to complete the proof of the lemma it is now enough to show that for every array of values α_i such that $0 \leq \alpha_i \leq 1$ for every $1 \leq i \leq b$,

$$\frac{1+\frac{\lambda}{\prod_{i=1}^b(1+\alpha_i R^O_\ell)}}{1+\frac{\lambda}{\prod_{i=1}^b(1+\alpha_i R^E_\ell)}} \;\leq\; \frac{1+R^E_{\ell+1}}{1+R^O_{\ell+1}},$$

i.e., that

$$1+\frac{\lambda}{\prod_{i=1}^b(1+\alpha_i R^O_\ell)} - \frac{1+R^E_{\ell+1}}{1+R^O_{\ell+1}} - \frac{(1+R^E_{\ell+1})\lambda}{(1+R^O_{\ell+1})\prod_{i=1}^b(1+\alpha_i R^E_\ell)} \;\leq\; 0. \tag{6}$$

If $\alpha_i = 1$ for every i then (6) clearly holds (as an equality). It is therefore enough to show that for every i, the derivative of the l.h.s. of (6) w.r.t. α_i is positive for all $(\alpha_1, \ldots, \alpha_b) \in [0,1]^b$. By symmetry, it is enough to show this for $i = 1$. Now, the derivative w.r.t. α_1 is:

$$-\frac{\lambda R^O_\ell}{(1+\alpha_1 R^O_\ell)\prod_{i=1}^b(1+\alpha_i R^O_\ell)} + \frac{\lambda(1+R^E_{\ell+1})R^E_\ell}{(1+R^O_{\ell+1})(1+\alpha_1 R^E_\ell)\prod_{i=1}^b(1+\alpha_i R^E_\ell)}.$$

Since we wish to show that the last expression is positive, it is enough to show that

$$\left(\frac{1+\alpha_1 R^E_\ell}{1+\alpha_1 R^O_\ell}\right)\prod_{i=1}^b\left(\frac{1+\alpha_i R^E_\ell}{1+\alpha_i R^O_\ell}\right) \;\leq\; \frac{(1+R^E_{\ell+1})R^E_\ell}{(1+R^O_{\ell+1})R^O_\ell}.$$

Now, since R_ℓ^O is increasing in ℓ and R_ℓ^E is decreasing (because placing a condition at level $\ell+1$ is the same as placing a convex combination of conditions at level ℓ), then

$$\begin{aligned}\frac{(1+R_{\ell+1}^E)R_\ell^E}{(1+R_{\ell+1}^O)R_\ell^O} &\geq \frac{(1+R_\ell^E)R_{\ell+1}^E}{(1+R_\ell^O)R_{\ell+1}^O} \\ &= \left(\frac{1+R_\ell^E}{1+R_\ell^O}\right)^{b+1} \\ &\geq \left(\frac{1+\alpha_1 R_\ell^E}{1+\alpha_1 R_\ell^O}\right)\prod_{i=1}^{b}\left(\frac{1+\alpha_i R_\ell^E}{1+\alpha_i R_\ell^O}\right),\end{aligned} \tag{7}$$

as required. The equality above follows from the fact that $R_{\ell+1}^E = \frac{\lambda}{(1+R_\ell^O)^b}$, while for the last inequality we used the fact that $R_\ell^E \geq R_\ell^O$, and therefore, $\frac{1+\alpha R_\ell^E}{1+\alpha R_\ell^O} \leq \frac{1+R_\ell^E}{1+R_\ell^O}$ for every $0 \leq \alpha \leq 1$. This completes the proof of Lemma 4.2. □

Remark: As we already mentioned following the statement of Theorem 4.1, the theorem does not hold if we replace the uniform assignment λ with a general assignment $\vec{\lambda}'$ that dominates $\vec{\lambda}$. The crucial property of the uniform assignment used in the above proof is that R_ℓ^O is increasing in ℓ and R_ℓ^E is decreasing, which allowed us to establish (7). Indeed, the theorem still holds if we replace the uniform assignment λ with $\vec{\lambda}'$ that dominates $\vec{\lambda}$ as long as under $\vec{\lambda}'$ the conditional ratios at the root (given the *ODD* and *EVEN* conditions) are sandwiched between the same conditional ratios at the children (for every child).

We end this section by showing that Theorem 4.1 holds for $\widehat{\mathbb{T}}^b$ (the tree in which the root has $b+1$ children), as claimed. Maintaining the notation R for the ratio at the root of $\mathbb{T}^b$ and writing $\mathbb{R}$ for the ratio of probabilities at the root of $\widehat{\mathbb{T}}^b$, we have

$$\begin{aligned}\frac{\mathbb{R}_\ell^E(\vec{\lambda})}{\mathbb{R}_\ell^O(\vec{\lambda})} &= \prod_{i=1}^{b+1}\left(\frac{1+R_{\ell-1}^E(\vec{\lambda}_i)}{1+R_{\ell-1}^O(\vec{\lambda}_i)}\right) \\ &\leq \left(\frac{1+R_{\ell-1}^E(\lambda)}{1+R_{\ell-1}^O(\lambda)}\right)^{b+1} \\ &= \frac{\mathbb{R}_\ell^E(\lambda)}{\mathbb{R}_\ell^O(\lambda)},\end{aligned}$$

where the two equalities follow from the fact that the subtree of $\widehat{\mathbb{T}}^b$ rooted at any given child of the original root is isomorphic to $\mathbb{T}^b$, while the inequality is an application of the second part of Lemma 4.2.

5 Algorithmic implications

In this section we describe and analyze the algorithm for approximating the partition function Z_G^λ (or for approximately sampling independent sets of G) claimed in Theorem 2.7. This algorithm is based on the recursive procedure for calculating p_v described in Section 3. Before giving the details of the algorithm, we note that a standard argument reduces the calculation of Z to the calculation of p_v. To see this, notice that in order to calculate Z it is enough to calculate the probability of

the empty set since this probability is exactly $1/Z$. Now, in order to calculate the probability of the empty set we first calculate the probability that v is unoccupied (i.e., $1 - p_v$) for some vertex v and multiply this by the probability of the empty set in $G \setminus v$ (i.e., conditioned on v being unoccupied), where the latter is computed recursively. Similarly, in order to generate a random independent set we can choose v to be occupied with probability p_v and unoccupied otherwise, and continue to generate the rest of the configuration conditioned on the chosen value at v. We thus concentrate on calculating p_v.

In Section 3 we described a recursive procedure for calculating p_v that was based on (2). It is clear from the analysis there that the time complexity of this procedure is $O(|T_{\text{saw}}(G, v)|)$, which may be exponential in $|G|$. (A trivial upper bound on $|T_{\text{saw}}(G, v)|$ is b^ℓ, where ℓ is the length of the longest path in G.) This is in line with the fact that it is NP-hard [6] to calculate (and even approximate) Z_G^λ (and therefore p_v) for general G and λ. Nevertheless, when the tree exhibits strong spatial mixing with exponential decay we can use a slight modification of this algorithm to approximate p_v in polynomial time. Notice that if we can output two numbers p_1, p_2 such that $p_1 \leq p_v \leq p_2$ and $p_2 \leq (1 + \frac{\epsilon}{(1+\epsilon)n})p_1$, then, by the same reduction as above, we can output two numbers Z_1, Z_2 such that $Z_1 \leq Z \leq Z_2$ and $Z_2 \leq (1 + \epsilon)Z_1$, or generate a random independent set such that the probability of outputting I is within a factor $(1 \pm \epsilon)$ of the required probability $\frac{\lambda^{|I|}}{Z}$.

Now, notice that if we replace the ratio $R^{\sigma_\Lambda, \tau_i}_{(G' \setminus v_i), u_i}$ in (2) with an upper bound on this ratio we get a lower bound on $R^{\sigma_\Lambda}_{G,v}$, and similarly, plugging in a lower bound on the first ratio will result in an upper bound on the latter. We can thus have the following recursive procedure for calculating lower and upper bounds on p_v. The recursive calls return a lower and an upper bound on $R^{\sigma_\Lambda, \tau_i}_{(G' \setminus v_i), u_i}$ for each i. The lower bounds are then used to compute an upper bound on $R^{\sigma_\Lambda}_{G,v}$ and vice versa. The procedure has three stopping rules. The first two are as in the procedure presented in Section 3, namely, if v is fixed by σ_Λ or if v has no neighbors then both the lower and upper bounds are set to the same value as described in Section 3. The third stopping rule is the following. If none of the first two rules apply and if the stack of the recursion is ℓ levels deep, where ℓ is a parameter of the algorithm, set the lower and upper bounds on R to 0 and ∞, respectively.

We go on to analyze the above algorithm. First, a trivial induction (already described above) verifies that the algorithm outputs two numbers p_1, p_2 such that $p_1 \leq p_v \leq p_2$. What remains to be analyzed is the size of the interval $[p_1, p_2]$ and the running time of the algorithm. Notice that if we run this algorithm for calculating p_v with the levels parameter set to ℓ then the upper bound p_2 that we get is exactly the probability that the root of $T_{\text{saw}}(G, v)$ is occupied conditioned on all the vertices that are not already fixed at level ℓ below the root being occupied (respectively unoccupied) if ℓ is even (respectively odd). The lower bound p_1 is exactly the probability of the root being occupied under the negated condition. Now, since we are assuming the tree exhibits strong spatial mixing with rate $\delta(\ell) = O(\exp(-\alpha\ell))$ for some $\alpha > 0$, if we run the algorithm with $\ell = [\ln(\frac{(1+\epsilon)n}{\epsilon})]/\alpha + O(1)$ then $p_2 \leq (1 + \frac{\epsilon}{(1+\epsilon)n})p_1$, as required. On the other hand, the running time of the algorithm with parameter ℓ is in the order of the size of $T_{\text{saw}}(G, v)$ restricted to its first ℓ levels. A trivial upper bound on the running time is thus $O(b^\ell) = O((\frac{(1+\epsilon)n}{\epsilon})^{(\ln b)/\alpha})$, which yields the bound stated in Theorem 2.7.

Remark: On a first look it seems that one needs to know the rate of decay in the definition of strong spatial mixing in order to know how to set the parameter ℓ. This can be tackled in two different ways. The first is that the rate of decay on the regular tree can be calculated efficiently (or at least sufficiently approximated). An alternative approach is to simply try all settings of ℓ (increasing its value by one in each iteration) until the interval we get is sufficiently small. This is a valid approach since the bound on the running time that

we used is exponential in ℓ, and hence trying all settings of $\ell = 1, \ldots, m$ has running time in the same order as running it only with $\ell = m$.

6 Future Research

Notice that the recursive procedure (and hence the tree representation) described in Section 3 is not specific to the independent-sets model. Indeed, it applies to any model of nearest-neighbor interaction in which each vertex is assigned a binary value (e.g., Ising models). Thus, as we already mentioned, Theorem 2.3 can be generalized to any model with a binary *spin* space. An interesting open question is whether it holds in models with larger spin spaces. Of particular interest is the model of proper colorings, where the goal is to show that $b+2$ colors (the threshold for weak spatial mixing on the regular tree) are enough for spatial mixing on any graph of maximum degree $b+1$. (Notice that it is not even known that the regular tree exhibits *strong* spatial mixing throughout this regime.)

Another direction for future research is lowering the bound on λ for which the computational problem is NP-hard. A particularly interesting question here is whether the problem is hard already for $\lambda > \lambda_c$. A positive answer would establish the first rigorous correspondence between computational complexity and phase transitions in statistical physics.

Acknowledgments

The author would like to thank Elchanan Mossel, Alistair Sinclair and Fabio Martinelli for numerous fruitful discussions, and for the inspiration to revisit this problem. Thanks also go to Leslie-Ann Goldberg and Nir Halman for pointing out that the algorithm presented here is the first to give a deterministic FPTAS for a #P-complete problem.

References

[1] J. VAN DEN BERG and A. ERMAKOV, "A new lower bound for the critical probability of site percolation on the square lattice," *Random Structures and Algorithms* **8** (1996), pp. 199–212.

[2] J. VAN DEN BERG and J. STEIF, "Percolation and the hard-core lattice gas model," *Stochastic Processes and their Applications* **49** (1994), pp. 179–197.

[3] G. BRIGHTWELL, O. HÄGGSTRÖM and P. WINKLER, "Nonmonotonic behavior in hard-core and Widom-Rowlinson models," *J. Stat. Physics* **94** (1999), pp. 415–435.

[4] F. CESI, "Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields," *Probability Theory and Related Fields* **120** (2001), pp. 569–584.

[5] R.L. DOBRUSHIN, J. KOLAFA, S.B. SHLOSMAN, "Phase diagram of a two-dimensional Ising antiferromagnet (computer assisted proof)," *Comm. Math. Phys.* **102** (1985), pp. 89–103.

[6] M. DYER, A. FRIEZE and M. JERRUM, "On counting independent sets in sparse graphs," *Proc. 40th IEEE Symp. on Foundations of Computer Science*, 1999, pp. 210–217.

[7] M. DYER and C. GREENHILL, "On Markov chains for independent sets," *J. Algorithms* **35** (2000), pp. 17–49.

[8] M. DYER, A. SINCLAIR, E. VIGODA and D. WEITZ, "Mixing in time and space for lattice spin systems: A combinatorial view," *Random Structures and Algorithms* **24** (2004), pp. 461–479.

[9] M. FISHER and M. SYKES, "Excluded volume problem and the Ising model of ferromagnetism," *Physical Review* **114** (1959), pp. 45–58.

[10] L.A. GOLDBERG, R. MARTIN and M. PATERSON, "Strong spatial mixing with fewer colours for lattice graphs," *Proc. 45th IEEE Symp. on Foundations of Computer Science*, 2004, pp. 562–571.

[11] N. HALMAN, D. KLABJAN, M. MOSTAGIR, J. ORLIN and D. SIMCHI-LEVI, "A fully polynomial time approximation scheme for single-item stochastic lot-sizing problems with discrete demand," submitted for publication, 2006.

[12] T. HAYES and E. VIGODA, "Coupling with the stationary distribution and improved sampling for colorings and independent sets," *Proc. 16th ACM-SIAM Symp. on Discrete Algorithms*, 2005, pp. 971–979.

[13] F.P. KELLY, "Stochastic models of computer communication systems," *Journal of the Royal Statistical Society B* **47** (1985), pp. 379–395.

[14] A.B. KIRILLOV, D.C. RADULESCU and D.F. STYER, "Vassertein distances in two-state systems," *J. Stat. Phys.* **56** (1989), pp. 931–937.

[15] M. LUBY and E. VIGODA, "Approximately counting up to four," *Proc. 29th ACM Symp. on Theory of Computing*, 1997, pp. 682–687.

[16] F. MARTINELLI and E. OLIVIERI, "Approach to equilibrium of Glauber dynamics in the one phase region I: The attractive case," *Comm. Math. Phys.* **161** (1994), pp. 447–486.

[17] F. MARTINELLI, A. SINCLAIR and D. WEITZ, "Fast mixing for independent sets, colorings and other models on trees," *Proc. 15th ACM-SIAM Symp. on Discrete Algorithms*, 2004, pp. 456–465.

[18] D.C. RADULESCU, "A computer-assisted proof of uniqueness of phase for the hard-square lattice gas model in two dimensions," Ph.D. dissertation, Rutgers, The State University of New Jersey, October 1997.

[19] A. D. SCOTT and A. D. SOKAL, "The repulsive lattice gas, the independent-set polynomial, and the Lovasz local lemma," *J. Stat. Phys.* **118** (2005), pp. 1151–1261.

[20] A. D. SOKAL, "A personal list of unsolved problems concerning lattice gases and antiferromagnetic Potts models," *Markov Process. Related Fields* **7** (2001), pp. 21–38.

[21] F. SPITZER, "Markov random fields on an infinite tree," *Annals of Probability* **3** (1975), pp. 387–398.

[22] D.W. STROOCK and B. ZEGARLINSKI, "The logarithmic Sobolev inequality for discrete spin systems on a lattice," *Comm. Math. Phys.* **149** (1992), pp. 175–194.

[23] E. VIGODA, "A note on the Glauber dynamics for sampling independent sets," *Electronic Journal of Combinatorics* **8**(1) (2001).

[24] D. WEITZ "Mixing in time and space for discrete spin systems," Ph.D. dissertation, University of California at Berkeley, May 2004.