Counting Without Sampling: Asymptotics of the Log-Partition Function for Certain Statistical Physics Models*†‡

Antar Bandyopadhyay,[1] David Gamarnik[2]

[1]*Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, New Delhi Centre, New Delhi 110016, India; e-mail: antar@isid.ac.in*

[2]*Operations Research Center and Sloan School of Management, MIT, Cambridge, Massachusetts 02139; e-mail: gamarnik@mit.edu*

ABSTRACT: In this article we propose new methods for computing the asymptotic value for the logarithm of the partition function (*free energy*) for certain statistical physics models on certain type of finite graphs, as the size of the underlying graph goes to infinity. The two models considered are the *hard-core* (independent set) model when the *activity* parameter λ is *small*, and also the Potts (*q*-coloring) model. We only consider the graphs with *large girth*. In particular, we prove that asymptotically the logarithm of the number of *independent sets* of any r-regular graph with large girth when rescaled is approximately constant if $r \leq 5$. For example, we show that every 4-regular n-node graph with large girth has approximately $(1.494\cdots)^n$-many independent sets, for large n. Further, we prove that for every r-regular graph with $r \geq 2$, with n nodes and large girth, the number of proper $q \geq r+1$ colorings is approximately $[q(1 - \frac{1}{q})^{\frac{r}{2}}]^n$, for large n. We also show that these results hold for random regular graphs with high probability (w.h.p.) as well.

As a byproduct of our method we obtain simple algorithms for the problem of computing approximately the logarithm of the number of independent sets and proper colorings, in low degree

---

Correspondence to: David Gamarnik

graphs with large girth. These algorithms are deterministic and use certain correlation decay properties for the corresponding Gibbs measures, and its implications to uniqueness of the Gibbs measures on the infinite trees, as well as some simple *cavity trick* which is well known in the physics and the Markov chain sampling literature. © 2008 Wiley Periodicals, Inc. Random Struct. Alg., 33, 452–479, 2008

Keywords: graph counting; Gibbs measures; correlation decay; approximation algorithms

1. INTRODUCTION

Counting is a natural counterpart to a combinatorial optimization problem. The typical set up involves counting the number of feasible solutions to some combinatorially constrained problem. The most widely studied such problems involve counting the number of solutions to a bin packing problem [22], counting the number of independent sets (also known as hard-core model in statistical physics) [13, 26], matchings [22], proper colorings in graphs (Potts model in statistical physics) [12, 13], computing volume of a convex body [11, 25, 27], permanent of a matrix (counting the number of full matchings of a bipartite graph) [7, 21–23, 38], etc.

A natural extension of the counting problem is (exponentially) weighted counting, that is computing the partition function. Partition function is a fundamental object in statistical physics, and the connection between the counting and the partition function is well known. There are many results in the physics literature on computing partition functions for various statistical physics models. Unfortunately, most of these results are not rigorous, and involve what is known as the replica-symmetry and the replica symmetry breaking cavity method, also known as, the replica symmetry breaking Ansatz [33]. The process of getting a rigorous argument for these spectacular but often unproven results of physics was undertaken relatively recently in mathematics. For example, Talgrand [37] proved the validity of the Parisi formula for the partition function limit of a Sherrington-Kirpatrick's model. Also Talagrand [36] proved the existence, and showed a method for computing the partition function limit of a random K-SAT problem in an appropriately defined high temperature regime. However, the process of building a full mathematical picture of the cavity method is still largely under way.

In this article we propose new methods for computing the asymptotic for the log-partition function for certain statistical physics models on graphs with low degree and *large girth*. To be more specific, we obtain explicit asymptotic for the log of the partition function for the *hard-core* (independent set) model on r-regular graphs with large girth, when the activity parameter λ is *small*, and we show that such asymptotic does not depend on the choice of the graph sequence. For example, we show that in every 4-regular graph with n nodes and large girth, the number of independent sets is approximately $(1.494\ldots)^n$, *irrespectively of the graph!* Precisely, we show that the logarithm of the number of independent sets divided by n approaches $\log(1.494\ldots)$ as girth increases. The class of regular graphs with large girth is very rich and the fact that the number of independent sets is the same in all of them is an interesting result on its own right. The value $1.494\ldots$ is a numeric approximation of a solution to a certain fixed-point equation. We obtain similar limiting numeric values for the case of r-regular graphs when $r = 2, 3, 4, 5$. We also obtain explicit asymptotic for the log of the number of proper coloring of r-regular graphs with large girth. For example, we show that for every constant $q \geq r + 1$, the number of q colorings in every r-regular graphs with large girth is approximately $(q(1 - \frac{1}{q})^{\frac{r}{2}})^n$, when n is large. We note, that our results

allow both q and r to be arbitrarily small. All of the known results for counting which are based on Markov chain method require q and r to be at least a large positive constant [12].

Moreover, we show similar asymptotic results when $n \to \infty$ holds, for both the hard-core model and counting the number of proper q-colorings for random r-regular graphs on n vertices.

As a by product of our analysis, we also obtain simple algorithms for computing approximately the logarithm of the number of independent sets and colorings (computing the log-partition function) in low degree graphs with large girth. In particular, we propose a simple polynomial time algorithm for computing approximately the logarithm of the number of independent sets in graphs with maximum degree ≤ 4 and large girth. Similarly, for every q we propose a simple computable expression for the logarithm of the number of proper q-colorings of any graph with maximum degree $r \leq q - 1$ and large girth.

It is worthwhile to mention that typically for such problems, the set of feasible solutions is exponentially large and exhaustive search is computationally prohibited. This complexity appears to be fundamentally unavoidable, [38]. Modulo a complexity theoretic conjecture, the problems in #P do not admit polynomial time algorithms, and thus research focused on approximation algorithms. Here the most powerful method comes from the theory of rapidly mixing Markov chains. The typical setup involves relating counting problem to a sampling problem via certain telescoping trick (see for example identity (1) below) and then computing some marginal probabilities using sampling technique. The main technical challenge is establishing that the underlying Markov chain mixes in polynomial time (rapid mixing). The scope of Markov chains for which rapid mixing has been established includes such notable breakthrough results as Jerrum and Sinclair's [21], and Jerrum, Sinclair and Vigoda's [23] proof of rapid mixing of a Markov chain related to permanent, and Dyer, Frieze and Kannan [11] proof of rapid mixing of a Markov chain related to computing the volume of a convex body. Subsequent improvements in running time for computing volumes have been established by Kannan, Lovasz, and Simonovits [25] and Lovasz and Vempala [27]. Somewhat closer to the topic of this article, Luby and Vigoda [26] showed that a Markov chain related to counting independent sets is rapidly mixing, when the underlying graph has degree at most 4. The principal strength of the Markov chain-based randomized algorithms as opposed to the type of algorithms considered in this article, is the fact that randomized algorithms provide an arbitrary level of approximation of the *partition function itself*, rather than the *logarithm of the partition function*—clearly a far stronger level of approximation. Also our computation time, while polynomial in the size of the graph, is not polynomial in ϵ—the approximation parameter. Thus, our algorithm is PAS (polynomial time approximation scheme) as opposed to FPRAS (fully polynomial time randomized approximation scheme) as is typically established using Markov chains method. But contrast to that, our algorithms are deterministic, so they do not suffer from sampling error.

The main technical approach underlying our results is the progress in understanding properties of the Gibbs distributions on the infinite regular trees for independent sets, coloring, Ising and some other related models in the context of *correlation decay*, and the connection of thereof to the uniqueness of the Gibbs measure. We use this stream of work to propose a different method for computing marginal probability featuring in cavity equation (1) below. In one of the earliest results in this area, Spitzer [35], Zachary [42], Kelly [24], the following phase transition property for independent set on the infinite r-regular trees was established: the probability that a root of the tree belongs to an independent set

selected according to the Gibbs measure is asymptotically independent from the finite depth boundary of a tree, provided that the activity λ is sufficiently small. The "counting" case $\lambda = 1$ satisfies this condition for $r \leq 5$ but breaks down for larger r. A recent extension of this result to general Galton-Watson type random trees and Erdös-Rényi type random graphs was done by Bandyopadhyay [5]. Similar uniqueness property is also known for Ising model [16] and recently was established for coloring in the case of $q \geq r + 1$ colors by Jonasson [20], closing an open problem posed earlier by Brightwell and Winkler [8]. The correlation decay property (long-range independence) featured lately very prominently in a variety of contexts including Aldous' proof of the ζ_2-limit for the random assignment problem [2], bivariate uniqueness and endogeny of recursive distributional equations by Aldous and Bandyopadhyay [1], Bandyopadhyay [6], Bandyopadhyay [5], Warren [39], the local weak convergence properties [4, 15, 17, 18], and the problems of reconstruction on a tree, [30]. Yet, the importance of the correlation decay property for the uniqueness of the Gibbs distribution was well recognized long time ago in the fundamental work of Dobrushin [14], dating back to 70's. Although Dobrushin's work was conducted primarily for lattices, there is a recent extension of this work by Weitz [40] to more general graphs.

In this article we establish the correlation decay property for independent sets, similar to the one considered by Kelly [24] but for an arbitrary (not necessarily regular) tree with maximum degree at most 4. This property coupled with the cavity trick (1) almost immediately leads to a simple algorithm for computing approximately the partition function for independent sets. The corresponding algorithm for colorings is obtained by a simple extension of the Jonasson's [20] uniqueness theorem for colorings. Methodologically, our approach consists of implementations of the following three steps. First computing appropriate marginal probabilities on a tree. This step typically involves a very simple recursive type computation. Then showing that the boundary of the tree has a vanishing impact on this marginally probability (correlation decay). Finally, the correlation decay is used to project the results of computation of marginal probabilities to non-tree graphs with locally tree-like structure. It is important to note that we establish a form of the correlation decay—the marginal probability is asymptotically independent from an *arbitrary* boundary condition in the supremum sense. Although this allows us to bypass the necessity of knowing the "correct" boundary condition implied by the rest of the graph, this is probably too pessimistic. Weaker forms of correlation decays have been studied actively in the context of reconstruction on trees [28–31], and might prove to be useful for our problem of counting as well.

Our explicit results for regular graphs are obtained by explicit computations of marginal probabilities for regular trees. An additional technical difficulty is the fact that the cavity step "destroys" the regularity of the graph. A simple trick introduced by Mezard and Parisi [32], (see also [34]) fixes this problem via some "rewiring" step. The regime corresponding to the correlation-decay property in our sense, is called a *liquid phase*. Our results then can be viewed as a rigorous treatment of liquid phase solution for independent sets model. Thus, our work strengthens further an interesting and intriguing connection between the statistical physics and the theory of algorithms.

The rest of the article is organized as follows. In the following section we provide the necessary background and definitions. Main results and their extensions, including the extensions to random regular graphs are presented in Section 3. Proofs are derived in Sections 4.5. Some conclusions and direction to further research are presented in Section 6.

2. NOTATIONS AND BASICS

Throughout the article we consider a simple graph G with the node set $V = \{v_1, \ldots, v_n\}$ and edge set $E = \{e_1, \ldots, e_m\}$. We will always denote by $n = n(G)$ the number of nodes in the graph. With some abuse of notation we will be writing $v \in G$, if v belongs to the node set V of the graph G. For every $v \in G$, $r(v) = r(v, G)$ denotes the degree of v in G. $N(v, G)$ denotes the set of neighbors of v in G. The maximum degree and the girth (size of the smallest cycle) of G are denoted by $r = r(G) = \max_{1 \leq k \leq n} r(v_k)$ and $g = g(G)$, respectively. Let $\mathcal{G}_0(n, g, r)$ be the set of all degree-r graphs G with n nodes and girth at least g. Also let $\mathcal{G}(n, g, r)$ be the set of all r-regular graphs G with n nodes and girth at least g. Typically, we will be considering graphs with a fixed maximum degree r, while the girth diverging to infinity as a function of n. For every positive integer t and every node v_i, we denote by $T(v_i, t)$ the depth-t neighborhood of v_i – the set of nodes reachable from v_i by paths of lengths at most t. Clearly $g > 2t$ implies that $T(v_i, t)$ is a tree for every node v_i. A subset $I \subset V$ is independent, if no two nodes of I share an edge. $\mathcal{I} = \mathcal{I}(G)$ denotes the set of all independent sets in G. A proper coloring $C \in \mathcal{C}(q)$ is an assignment $C : V \to \{1, \ldots, q\}$ of nodes V to colors $\{1, 2, \ldots, q\}$ such that no two nodes which share an edge are assigned to the same color. For every $q \in \mathbb{N}$, $\mathcal{C}(q, G) = \mathcal{C}(q)$ denotes the set of all proper colorings of the nodes of G by colors $\{1, 2, \ldots, q\}$. Throughout the paper we will only consider the case $q \geq r + 1$, then as it is well-known, the set $\mathcal{C}(q)$ is non-empty. There is a way of defining a general model which simultaneously includes the model for independent sets and colorings by means of graph homomorphisms. This formalism has been used in a variety of papers [9, 13]. Here, for simplicity we do not resort to this formalism.

A classical object in statistical physics is the Gibbs probability distribution on the sets $\mathcal{I}, \mathcal{C}(q)$. Fix $\lambda > 0$ and $\lambda_j > 0$ for all $1 \leq j \leq q$, these will be called the activity (fugacity) parameters.

The Gibbs distribution on the set $\mathcal{I}$ assigns a probability proportional to $\lambda^{|I|}$ to each independent set I. More precisely,

$$\mathbb{P}(\boldsymbol{I} = I) = \frac{\lambda^{|I|}}{Z(\lambda)},$$

where $\boldsymbol{I}$ is a random independent set with respect to Gibbs measure, and $Z(\lambda) = Z(\lambda, G) = \sum_{I \in \mathcal{I}} \lambda^{|I|}$, is the normalizing constant, also known as, the partition function. The quantity $\log Z(\lambda)$ is known as the *free energy*. To emphasize the underlying graph, sometimes we will denote the Gibbs measure by $\mathbb{P}_G(\cdot)$. When $\lambda = 1$, $Z(\lambda, G) = Z(1, G) = |\mathcal{I}|$ and the Gibbs distribution is simply the uniform distribution on the set of all independent sets.

For the case of coloring, the Gibbs distribution on the set $\mathcal{C}(q)$ of proper colorings is introduced similarly as

$$\mathbb{P}(\boldsymbol{C} = C) = \frac{\prod_{1 \leq j \leq q} \lambda_j^{|C_j|}}{Z(\lambda)},$$

where $\boldsymbol{C}$ is a random coloring distributed according to the Gibbs distribution, and $\lambda = (\lambda_1, \ldots, \lambda_q)$ is a fixed vector of activity parameters, $C_j = \{v \in V : C(v) = j\}$, and $Z(\lambda) = Z(\lambda, G) = \sum_{C \in \mathcal{C}(q)} \prod_{1 \leq j \leq q} \lambda_j^{|C_j|}$ is again the normalizing constant, which is the partition function. Again the special case $\lambda_j = 1, 1 \leq j \leq q$ corresponds to the uniform distribution on the set $\mathcal{C}(q)$ of proper q-colorings. In this article we focus exclusively on this special case and use notation $Z(q, G)$ or $Z(G)$ instead.

2.1. Simple Cavity Trick

There exists a way to represent the partition functions $Z(\lambda, G)$ and $Z(q, G)$ in terms of the marginals of the Gibbs measure in the following sense, which is known as the *cavity trick*. Let $G_0 = G$ and $G_k = G \setminus \{v_1, \ldots, v_k\}, k = 1, 2, \ldots, n$. The following two propositions give expressions for the partition functions $Z(\lambda, G)$ and $Z(q, G)$, in terms of the measures on smaller graphs.

Proposition 1. *The following relation holds*

$$\frac{Z(\lambda, G_k)}{Z(\lambda, G_{k-1})} = \mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I}). \tag{1}$$

As a result,

$$Z(\lambda, G) = \prod_{k=1}^{n} \mathbb{P}_{G_{k-1}}^{-1}(v_k \notin \boldsymbol{I}). \tag{2}$$

This proposition is well known and is used for Markov chain based approximation algorithms for counting. We provide the proof for completeness. For convenience we assume that a partition function of an empty graph is equal to the unity.

Proof. The proof is obtained by considering a telescoping product

$$Z(\lambda, G) = \prod_{k=1}^{n} \frac{Z(\lambda, G_{k-1})}{Z(\lambda, G_k)}$$

and observing

$$\mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I}) = \frac{\sum_{I \in \mathcal{I}(G_{k-1}) : v_k \notin I} \lambda^{|I|}}{Z(\lambda, G_{k-1})} = \frac{Z(\lambda, G_k)}{Z(\lambda, G_{k-1})}. \qquad \blacksquare$$

The analog of Proposition 1 is somewhat more complicated. For a random coloring $\boldsymbol{C}$ selected according to the Gibbs distribution and for any subset of nodes A, denote by $\boldsymbol{C}(A)$ the set of colors assigned to A. In particular, $\boldsymbol{C}(N(v_k, G_{k-1}))$ is the set of colors used by coloring $\boldsymbol{C}$ for the neighbors of the node v_k in the graph G_{k-1}. We will also write $\boldsymbol{C}(v)$ instead of $\boldsymbol{C}(\{v\})$ for every node $v \in G$. Again for convenience we assume that the number of proper q-colorings of an empty graph is equal to unity.

Proposition 2. *The following relation holds*

$$\frac{Z(q, G_{k-1})}{Z(q, G_k)} = q - \mathbb{E}_{G_k}[|\boldsymbol{C}(N(v_k, G_{k-1}))|]. \tag{3}$$

As a result,

$$Z(q, G) = \prod_{k=1}^{n} [q - \mathbb{E}_{G_k}[|\boldsymbol{C}(N(v_k, G_{k-1}))|]]. \tag{4}$$

Proof. The second part is obtained again by considering a telescoping product $Z(q, G) = \prod_{1 \leq k \leq n} \frac{Z(q, G_{k-1})}{Z(q, G_k)}$. To prove the first part we observe that

$$Z(q, G_{k-1}) = \sum_{1 \leq m \leq r(v_k, G_{k-1})} (q - m) |\{C \in \mathcal{C}(G_k) : C(N(v_k, G_{k-1})) = m\}|$$

where we simply observe that if the coloring C uses m colors for the neighbors of v_k in G_{k-1} then there are $q - m$ colors left for v_k itself. Then we divide both parts by $Z(q, G_k)$ and observe that

$$\sum_{1 \le m \le r(v_k, G_{k-1})} m \frac{|\{C \in \mathcal{C}(G_k) : C(N(v_k, G_{k-1})) = m\}|}{Z(q, G_k)} = \mathbb{E}_{G_k}[|\boldsymbol{C}(N(v_k, G_{k-1}))|].$$

∎

3. PROBLEM FORMULATION AND RESULTS

We start by giving a precise mathematical definition of the *large girth* concept used in this paper.

Definition 1. An (infinite) family of graphs $\mathcal{G}$ is defined to have *large girth* if there exists an increasing function $f : \mathbb{N} \to \mathbb{N}$ such that $\lim_{s \to \infty} f(s) = \infty$ and for every $G \in \mathcal{G}$ with n nodes

$$g(G) \ge f(n).$$

3.1. Log-Partition Function Asymptotics for the Regular Graphs

Our first two results provide explicit estimates for the cardinality of the number of independent sets $\mathcal{I}$ and colorings $\mathcal{C}(q)$ in the special case of regular graphs with large girth.

Theorem 1. *Suppose* $\lambda < (r - 1)^{r-1}/(r - 2)^r$. *Then the partition function* $Z(\lambda, G)$ *corresponding to independent sets satisfies*

$$\lim_{g \to \infty} \sup_{G \in \mathcal{G}(n,g,r)} \left| \frac{\log Z(\lambda, G)}{n} - \log \left(x^{-\frac{r}{2}} (2 - x)^{-\frac{r-2}{2}} \right) \right| = 0,$$

where x is the unique positive solution of $x = 1/(1 + \lambda x^{r-1})$. When $r = 2, 3, 4, 5$ and $\lambda = 1$, the corresponding limits for $n^{-1} \log |\mathcal{I}(G)|$ are respectively, $\log 1.618 \ldots, \log 1.545 \ldots, \log 1.494 \ldots$ *and* $\log 1.453 \ldots$.

Remark. One important corollary of this result is that the asymptotic value of the log-partition function (limit of free energy) is the same for every r-regular graph with large girth. In particular, this result validates the non-rigorous statistical physics approach for computing free energy, where only locally-tree like structure and regularity is used in computation of free energy. We note that such insensitivity result cannot be obtained by the Markov Chain sampling technique.

We now state our main result for coloring. As we already mentioned, we only consider the special case $\lambda_j = 1, 1 \le j \le q$, that is the problem of counting the number of colorings. The reason for this limitation will be apparent when we discuss a result by Jonasson [20] on coloring regular trees.

Theorem 2.   *For every $q \geq r + 1$, the number of q-colorings of graphs $G \in \mathcal{G}(n, g, r)$ satisfies*

$$\lim_{g \to \infty} \sup_{G \in \mathcal{G}(n,g,r)} \left| \frac{\log Z(q, G)}{n} - \log \left[q \left(1 - \frac{1}{q} \right)^{\frac{r}{2}} \right] \right| = 0.$$

Remark.   As an immediate corollary of Theorem 2 we obtain that for every constant $\alpha \geq 1$, the number of $q = \lfloor \alpha r \rfloor + 1$ colorings of graphs $G \in \mathcal{G}(n, g, r)$ is approximately $(q e^{-\frac{1}{2\alpha}})^n$ as $g, r \to \infty$. Recently Bezakova, Stefankovic, Vazirani, Vigoda [7] obtained the following lower bound on $|\mathcal{C}(q, G)|$ in arbitrary n-node graph with maximum degree r (no restriction on girth): $|\mathcal{C}(q, G)| \geq (q - r(1 - e^{-1}))^n$. Thus, when r is large and $q = \alpha r$ for some constant α, their bound becomes approximately $(q(1 - \alpha^{-1} + (\alpha e)^{-1}))^n$. It is not hard to see that our lower bound is strictly superior. For example, when $\alpha = 1$, their bound gives approximately $(q e^{-1})^n$ colorings, whereas, per our result, the correct limiting value (in log scale) is $(q/\sqrt{e})^n$. Of course our tight estimate comes at a cost of the large girth requirement.

3.2. Applications to Random Regular Graphs

Random graphs are obtained by drawing a graph from some family of graphs at random according to some (typically uniform) distribution. Specifically, an r-regular n-node random graph $G_r(n)$ is obtained by selecting an r-regular graph uniformly at random from the set of all r-regular graphs on n-nodes. An important feature of such a regular graph is that the number of small cycles is small. In particular, for every constant C the expected number of size-C cycles is $O(1)$ in terms of the number of nodes n, [19]. Thus, *essentially* such graphs have a large girth and we may expect that our results for regular graphs with large girth extend to this class of graphs. It is indeed the case as we state below. The derivation of these results is very similar to the one used for the class $\mathcal{G}(n, g, r)$.

Theorem 3.   *For every r and every $\lambda < (r-1)^{r-1}/(r-2)^r$, the (random) partition function $Z(\lambda, G_r(n))$ of a random r-regular graph $G_r(n)$ corresponding to the Gibbs distribution on independent sets satisfies*

$$\frac{\log Z(\lambda, G_r(n))}{n} \to \log \left[x^{-\frac{r}{2}} (2 - x)^{-\frac{r-2}{2}} \right],$$

with high probability (w.h.p.), as $n \to \infty$, where x is the unique positive solution of $x = 1/(1 + \lambda x^{r-1})$. In particular, when $r = 2, \ldots, 5$ and $\lambda = 1$, $\log Z(\lambda, G_r(n))/n$ converges w.h.p. to $\log 1.618 \ldots, \log 1.545 \ldots, \log 1.494 \ldots$ and $\log 1.453 \ldots$, respectively, as $n \to \infty$.

The results corresponding to the counting problem $\lambda = 1$ are obtained only for $r \leq 5$, precisely because the condition $(r - 1)^{r-1}/(r - 2)^r > 1$ holds only for this range.

Our corresponding result for colorings is as follows.

Theorem 4.   *For every r and every $q \geq r + 1$, the (random) partition function $Z(q, G_r(n))$ of a random r-regular graph $G_r(n)$ corresponding to the uniform distribution on proper q-colorings satisfies*

$$\frac{\log Z(q, G_r(n))}{n} \to \log \left[q \left(1 - \frac{1}{q} \right)^{\frac{r}{2}} \right].$$

w.h.p. as $n \to \infty$.

Theorem 4 is in fact not new. Using the second moment method it was established in [3], that logarithm of the number of q colorings of a graph $G_r(n)$ divided by n converges w.h.p. to $\log[q(1 - \frac{1}{q})^{\frac{r}{2}}]$, matching our expression. In fact the range for q for which this is the case includes $q < r$. However, the (second moment) argument relies strongly on randomness of the graph. We stress that our general result Theorem 2 holds for every regular graph with large girth.

3.3. Counting Independent Sets and Colorings

As a byproduct of our analysis we can compute approximately the sizes of the sets $\mathcal{I}$ and $\mathcal{C}(q)$, even when the underlying graph is not regular. Specifically, we are interested in approximating the exponents corresponding to the cardinalities of these sets.

Definition 2.   Value $\alpha > 0$ is defined to be ϵ-approximation of the log-partition function $\log Z(\lambda, G)$ if

$$(1 - \epsilon)\frac{\log Z(\lambda, G)}{n} \leq \alpha \leq (1 + \epsilon)\frac{\log Z(\lambda, G)}{n}.$$

where $\epsilon > 0$ is the error tolerance.

Given a family of graphs $\mathcal{G}$, an algorithm $\mathcal{A}$ is said to be Polynomial approximation scheme (PAS) for computing the log-partition function if for every $G \in \mathcal{G}$ it produces an ϵ-approximation of $\log Z(G)$ in time which is polynomial in n.

Here it is worth noting that in the literature there are alternative Markov chain based approaches for solving the counting problems, these typically provide approximation for the partition functions itself and not just a logarithm of the partition functions, as in our approach does. Also it typically runs in time which is also polynomial in ϵ^{-1}. Thus it is called fully polynomial randomized approximation scheme (FPRAS). On the other hand it provides approximation only with some probabilistic guarantee. We stress that the algorithms proposed in this paper provide deterministic guarantee, and thus are PAS, albeit the dependence on ϵ can be exponential. A natural intersection of two classes is fully polynomial approximation scheme (FPAS). The difference between different types of approximations is non-trivial and is not fully understood. For example, it is yet not clear that FPAS is always possible whenever FPRAS is possible. In fact Dyer, Goldberg, and Jerrum [13] provide an evidence to the contrary.

Our first algorithm result establishes existence of PAS for computing the logarithm of the number of independent sets in graphs.

Theorem 5.   *For every family $\mathcal{G}$ of graphs G with maximum degree $r \leq 4$ and large girth, the problem of computing $\log Z(\lambda, G)$ when $\lambda = 1$ is PAS.*

The family of graphs on which we establish algorithmic results is limited because of the high girth requirement. Our analysis is based on approximations, the quality of which depends on the size of the girth. For the case of the graph with small girth our algorithm is just a brute force computation. Since the requirement of the family of graphs $\mathcal{G}$ to have increasing girth, this brute force computation is performed on a constant size graphs and thus takes a constant amount of time.

As we will see in the next section under the assumptions of the Theorem 5 we have a *correlation decay* property which will allow us to efficiently estimate the marginal probabilities by concentrating only on "local" neighborhoods. It is worthwhile to mention here that if such a property can be proved then these efficient approximation of the marginal probabilities can be used to create a polynomial time algorithm to approximate the logarithm of the partition function. Even though we could not find a reference for this kind of an algorithm but it has been pointed out to us by an anonymous referee that this is well known to the researchers working in this filed. But we would like to emphasize that our work not only proves the correlation decay (Proposition 3) leading to Theorem 5, but also provides a better way of approximating the marginal probabilities by doing computation only on the trees. Here we do point out that we can do that because of our special choice of large girth graphs.

Our corresponding result for counting proper colorings does not require any upper bound on the maximum degree. Also it is more explicit and its algorithmic implication is immediate. Though in Section 5, we do describe an algorithm for completeness.

Theorem 6. *Given constants $q \geq r+1$, the number of q-coloring of graphs $G \in \mathcal{G}_0(n, g, r)$ satisfies*

$$\lim_{g \to \infty} \sup_{G \in \mathcal{G}_0(n,g,r)} \left| \frac{\log Z(q, G)}{n} - \frac{1}{n} \sum_{1 \leq k \leq n} \log \left[q \left(1 - \frac{1}{q} \right)^{r(v_k, G_{k-1})} \right] \right| = 0.$$

In particular, for every family $\mathcal{G}$ of graphs G with maximum degree r and large girth, the problem of computing $\log Z(q, G)$ is PAS.

Note that the bound in theorem above does not put any lower bound restriction on the number of nodes n. This is because the quality of approximation is completely controlled by the girth size. Implicitly, however, there is a trivial restriction, since when $n < g$, the graph has in fact infinite girth, namely, it is a tree. In this case, it can be verified directly, that the expression for $Z(q, G)$ is the exact number of colorings.

4. STUDY OF INDEPENDENT SETS

The key method for obtaining the results in this article is to establish a form of correlation decay, appropriately defined. Correlation decay is one of the key concepts in statistical physics which has been used to established the uniqueness of Gibbs distribution on infinite graphs (on finite graphs Gibbs distribution is unique by definition). These questions of uniqueness and correlation decay have been considered primarily on regular trees. Here we reconstruct some of these results and extend them to nonregular trees. A correlation decay property which we will establish will then be used to project our results to arbitrary graphs with large girth (and additional restrictions dictated by a particular context).

4.1. Independent Sets on Trees and Correlation Decay

Let T be an arbitrary rooted tree with root v_0 and depth at most t. That is the distance from v_0 to any other node $v \in T$ is at most t. Denote by $B(T)$ the boundary of the tree – the set of nodes with distance exactly t from the root. Any function $b : B(T) \to \{0, 1\}$ is called

a boundary condition b. When $B(T)$ is empty the boundary condition is not defined. We think of boundary condition as conditioning on which nodes on the boundary belong to an independent set (corresponding value is 1) and which do not (value is zero). In particular, for any boundary condition b, we denote by $\mathbb{P}(v_0 \in \boldsymbol{I}|b)$ the probability of the event "v_0 belongs to the random independent set $\boldsymbol{I}$," conditioned on the event "$\{v \in B(T) : v \in \boldsymbol{I}\} = \{v \in B(T) : b(v) = 1\}$," with respect to the Gibbs measure. Denote by $\mathcal{B}(T)$ the set of all boundary conditions b on T, and denote by $\mathcal{T}(t, r)$ the set of all trees with maximum degree at most r and depth at most t.

Our first result establishes the key correlation decay property of Gibbs distributions of independent sets on trees with maximum degree at most 4.

Proposition 3.   *The following bounds holds for every $t \geq 2$, $T \in \mathcal{T}(t, 4)$, $b, b_1, b_2 \in \mathcal{B}(T)$*

$$\frac{1}{2} \leq \mathbb{P}(v_0 \notin \boldsymbol{I}|b) \leq \frac{8}{9}. \tag{5}$$

and

$$|\mathbb{P}(v_0 \notin \boldsymbol{I}|b_1) - \mathbb{P}(v_0 \notin \boldsymbol{I}|b_2)| \leq (.9)^{t-2}. \tag{6}$$

where $\mathbb{P}(\cdot)$ is with respect to the Gibbs distribution with $\lambda = 1$.

Moreover, given λ satisfying $\lambda < (r - 1)^{r-1}/(r - 2)^r$, let x be the unique non-negative solution of the equation $x = 1/(1 + \lambda x^{r-1})$. Suppose all the nodes of T except for leaves and the root have degree r, and suppose the root has degree $r - 1$. Then for all $b \in \mathcal{B}(T), t \geq 2$,

$$|\mathbb{P}(v_0 \notin \boldsymbol{I}|b) - x| \leq \alpha^t, \tag{7}$$

for some constant $\alpha = \alpha(\lambda, r) < 1$. If, on the other hand, all the nodes except for leaves, have degree r (including the root), then

$$\left| \mathbb{P}(v_0 \notin \boldsymbol{I}|b) - \frac{1}{2 - x} \right| \leq \alpha^t, \tag{8}$$

for the same constant α.

Remark.   The second part of the proposition is a known result established by Kelly [24]. and we simply refer to Kelly's work for the proof. See also [10] (where w corresponds to $1/x - 1$), and [5] where the latter work is concerned with the extension of Kelly's result to general Galton-Watson type random trees. The constant $\alpha(\lambda, r)$ approaches unity as λ approaches $(r - 1)^{r-1}/(r - 2)^r$ and can expressed explicitly, but this is not required for our paper.

Proof.   At the beginning we assume no restrictions on λ and r. We fix a tree $T \in \mathcal{T}(t, r)$ with a root v_0 and an activity $\lambda > 0$. Denote by $v_1, \ldots, v_k, k \leq r$ the neighbors $N(v_0, T)$ of the root v_0 of the tree T. This includes the possibility $k = 0$ (the tree consists of only node v_0). For every node $v \in T$, let $T(v)$ be the subtree rooted at v containing only the descendants of v. Note that in particular $T(v_0) = T$. By $b(T(v))$ we denote the natural restriction of a boundary condition $b \in \mathcal{B}(T)$ to $T(v)$. For every node v, let $T(v|b)$ be the tree obtained by deleting the leaves $v' \in T(v)$ which have $b(v') = 1$, as well as their parent nodes, and any siblings, and be deleting the leaves $v' \in T(v)$ which have $b(v') = 0$ (but no parent and sibling nodes are deleted).

Given two independent sets I, I' on $T(v)$ observe that if they have the same intersection with the boundary $I \cap B(T(v)) = I' \cap B(T(v)) = b$, then also their intersection with the complement of $T(v|b)$ in $T(v)$ is the same. This is because the only non-boundary nodes in $T \setminus T(v|b)$ are parents of the boundary nodes which belong to I and I'. These parents nodes are not in I, I', by independence set constraints. Thus the intersection of I, I' with $T \setminus T(v|b)$ is the same as asserted. Then

$$
\begin{aligned}
\mathbb{P}_{T(v)}(\boldsymbol{I} = I | b) &= \frac{\mathbb{P}_{T(v)}(\boldsymbol{I} = I)}{\mathbb{P}_{T(v)}(\boldsymbol{I} \cap B(T(v)) = b)} \\
&= \frac{\lambda^{|I|}}{\sum_{I':I' \cap B(T(v))=b} \lambda^{|I'|}} \\
&= \frac{\lambda^{|I \cap T(v|b)|} \lambda^{|I \cap (T \setminus T(v|b))|}}{\sum_{I':I' \cap B(T(v))=b} \lambda^{|I' \cap T(v|b)|} \lambda^{|I' \cap (T \setminus T(v|b))|}} \\
&= \frac{\lambda^{|I \cap T(v|b)|}}{\sum_{I':I' \cap B(T(v))=b} \lambda^{|I' \cap T(v|b)|}} \\
&= \mathbb{P}_{T(v|b)}(\boldsymbol{I} = I \cap T(v|b)).
\end{aligned}
$$

Using convention $\prod_{1 \le j \le k}(\cdot) = 1$ when $k = 0$, we obtain

$$
Z(\lambda, T(v_0|b)) = \sum_{I \in \mathcal{I}(T(v_0|b))} \lambda^{|I|} = \prod_{1 \le j \le k} \left(\sum_{I \in \mathcal{I}(T(v_j|b))} \lambda^{|I|} \right) + \lambda \prod_{1 \le j \le k} \left(\sum_{I \in \mathcal{I}(T(v_j|b)), v_j \notin I} \lambda^{|I|} \right) \quad (9)
$$

We recognize that

$$
\frac{\prod_{1 \le j \le k} \left(\sum_{I \in \mathcal{I}(T(v_j|b))} \lambda^{|I|} \right)}{Z(\lambda, T(v_0|b))} = \frac{\prod_{1 \le j \le k} Z(\lambda, T(v_j|b))}{Z(\lambda, T(v_0|b))} = \mathbb{P}_{T(v_0)}(v_0 \notin \boldsymbol{I} | b)
$$

A similar identity holds for each $j \le k$ when $Z(\lambda, T(v_j|b))$ replaces $Z(\lambda, T(v_0|b))$ in the denominator and in the numerator the product is over neighbors of v_j in $T(v_j|b)$. Dividing both parts of (9) by $\prod_{1 \le j \le k} \left(\sum_{I \in \mathcal{I}(T(v_j|b))} \lambda^{|I|} \right) = \prod_{1 \le j \le k} Z(\lambda, T(v_j|b))$ we obtain

$$
\mathbb{P}_{T(v_0)}(v_0 \notin \boldsymbol{I} | b) = \frac{1}{1 + \lambda \prod_{1 \le j \le k} \mathbb{P}_{T(v_j)}(v_j \notin \boldsymbol{I} | b)}. \quad (10)
$$

Note, that similar recursion applies to any node v substituting the root v_0, by replacing T with $T(v)$. Specifically, we now take any node v which is a parent of a leaf in level t in a main tree T, if any exist. That is v is located on level $t - 1$. It has $r(v) - 1$ children which we denote by $v_1, \ldots, v_{r(v)-1}$. For every child $v_j, j \le r(v) - 1$ (if there are any) the value $\mathbb{P}_{T(v_j)}(v_j \notin \boldsymbol{I} | b)$ is either zero or one depending on whether $b(v_j) = 0$ or $= 1$, respectively. The recursive equation (10) implies that $\mathbb{P}_{T(v)}(v \notin \boldsymbol{I} | b) \in [(1 + \lambda)^{-1}, 1]$.

Now, suppose that v is any node on level $t - 2$. Then applying the same recursion and the previously obtained bounds on its $r(v) - 1$ children, we get

$$
\frac{1}{1 + \lambda} \le \mathbb{P}_{T(v)}(v \notin \boldsymbol{I} | b) \le \frac{1}{1 + \lambda(1 + \lambda)^{-r(v)+1}} \le \frac{1}{1 + \lambda(1 + \lambda)^{-r+1}}.
$$

For every node v in level $t - 2$ define $a(v) = 1/(1 + \lambda)$ and $c(v) = 1/(1 + \lambda(1 + \lambda)^{-r+1})$. Now we will define inductively $a(v), c(v)$ for every node $v \in T$. These values will be such that $\mathbb{P}_{T(v)}(v \notin \boldsymbol{I}|b) \in [a(v), c(v)]$, and

$$\frac{1}{1 + \lambda} \leq a(v) \leq c(v) \leq \frac{1}{1 + \lambda(1 + \lambda)^{-r+1}}. \tag{11}$$

for every node v. Suppose for every node u in levels $\tau + 1, \ldots, t - 2$ values $a(u)$ and $c(u)$ satisfying these two conditions are already defined. In particular, this is the case for $\tau + 1 = t - 2$. Consider any node v in level τ. If v has no children, then we set $a(v) = c(v) = 1/(1 + \lambda)$. Otherwise, denote the children by $v_1, v_2, \ldots, v_{r(v)-1}$. Applying (10), to $v_1, v_2, \ldots, v_{r(v)-1}$, we obtain

$$a(v) \triangleq \frac{1}{1 + \lambda \prod_{1 \leq j \leq r(v)-1} c(v_j)} \leq \mathbb{P}_{T(v)}(v \notin \boldsymbol{I}|b) \leq \frac{1}{1 + \lambda \prod_{1 \leq j \leq r(v)-1} a(v_j)} \triangleq c(v). \tag{12}$$

Also, since by inductive assumption, $a(v_j) \geq 1/(1 + \lambda), c(v_j) \leq 1/(1 + \lambda(1 + \lambda)^{-r+1}) \leq 1$, we obtain by the same argument as above that the same bounds hold for $a(v), c(v)$ for all the node v in levels up to $t - 2$:

$$\frac{1}{1 + \lambda} \leq a(v) \leq c(v) \leq \frac{1}{1 + \lambda(1 + \lambda)^{-r+1}}. \tag{13}$$

We have defined values $a(v), c(v)$ for all nodes $v \in T$.

We note that these bounds only depend on the tree T but not the boundary condition b. We now return to our special case $\lambda = 1, r \leq 4$ and show that the length of the bounding interval $c(v) - a(v)$ is geometrically decreasing as a function of the level of v.

Lemma 1.   *Suppose $r = 4, \lambda = 1$. Then for every node $v \in T$ in level τ, $c(v) - a(v) \leq (.9)^{t-2-\tau}$.*

Proof.   The proof proceeds by reverse induction in τ starting with $\tau = t - 2$. For $\tau = t - 2$ the bound holds trivially from $0 \leq a(v), c(v) \leq 1$. Assume it holds for levels $\tau + 1, \ldots, t - 2$ and consider any node v in level τ with children $v_1, \ldots, v_k, 0 \leq k \leq r - 1$. If $k = 0$ then $a(v) = c(v) = 1/(1 + \lambda)$ and the bound holds trivially. Now suppose $k > 0$. Introduce function $f : [(1 + \lambda)^{-1}, (1 + \lambda(1 + \lambda)^{-r+1})^{-1}]^k \to \mathbb{R}$ given by $f(z) = f(z_1, \ldots, z_k) = (1 + \lambda \prod_{1 \leq j \leq k} z_j)^{-1}$. We rewrite (12) as $f(c(v_1), \ldots, c(v_k)) = a(v) \leq c(v) = f(a(v_1), \ldots, a(v_k))$, where $a(v_j), c(v_j)$ satisfy the bounds in (13). Function f is differentiable on its domain. By mean value theorem, there exists $z \in [(1 + \lambda)^{-1}, (1 + \lambda(1 + \lambda)^{-r+1})^{-1}]^k$ such that

$$\begin{aligned}
c(v) - a(v) &= \nabla f(z)(a(v_1) - c(v_1), \ldots, a(v_k) - c(v_k)) \\
&\leq \|\nabla f(z)\|_1 \max_{1 \leq j \leq k} |a(v_j) - c(v_j)| \\
&\leq \|\nabla f(z)\|_1 (.9)^{t-2-\tau-1},
\end{aligned}$$

where the last bound follows from the inductive assumption. It then suffices to prove that $\|\nabla f(z)\|_1 < .9$. We expand $\|\nabla f(z)\|_1$ as

$$\|\nabla f(z)\|_1 = \frac{\lambda \prod_{1 \leq j \leq k} z_j \sum_{1 \leq j \leq k} z_j^{-1}}{\left(1 + \lambda \prod_{1 \leq j \leq k} z_j\right)^2}.$$

We now resort to our specific assumption $r \le 4, \lambda = 1$. The remainder of the proof is computer assisted. For given $k = 1, 2, 3$, consider a resolution .001 grid on the rectangle $[(1+\lambda)^{-1}, (1+\lambda(1+\lambda)^{-r+1})^{-1}]^k, 1 \le k \le 4$. We note that the right end $(1+\lambda(1+\lambda)^{-r+1})^{-1}$ of the rectangle is largest when $r = 4$, so we consider the set of vectors $z = (z_1, \ldots, z_k)$ of the form $z_j = .001m_j$, for some $m_j \in \mathbb{N}$ such that $1/2 = (1+\lambda)^{-1} \le z_j \le (1+2^{-3})^{-1}$ for all j. We have checked numerically using MATLAB that for every $k = 1, 2, 3$ and every point z on this k-dimensional grid, the value of $\|\nabla f(z)\|_1$ is at most .8736. Specifically, the maximum values for $k = 1, 2, 3$ (using rational computations) turn out to be $1089/2500 \approx 0.4356$, $109/165 \approx 0.6606$, $825/943 \approx 0.8749$, respectively.

We now use first-order Taylor approximation to obtain a bound on $\max \|\nabla f(z)\|_1$ over the entire domain of f. We will show that this bound is smaller than .9. For every z in the rectangle find an arbitrary grid point $\hat{z} = (\hat{z}_1, \ldots, \hat{z}_k)$ satisfying $|z_j - \hat{z}_j| < 0.001$ (typically many such points exist and we choose any of them). We denote the function $\|\nabla f\|_1$ by g for a notational convenience. Now let us show that for every two vectors z^1, z^2 which coincide in all the coordinates except for one, and such that $\|z^1 - z^2\|_1 < 0.001$, we have

$$|g(z^1) - g(z^2)| < 7 \times 10^{-3}. \tag{14}$$

This gives $|g(z) - g(\hat{z})| < 7 \times 10^{-3}k \le 7 \times 10^{-3} \times 3 < 0.021$. Combining with our bound for the points on the grid, we obtain that for every point z on the domain $\nabla f(z) < 0.8749 + 0.021 = 0.8959 < 0.9$ and the proof of the lemma would be complete.

To estimate the difference $|g(z^1) - g(z^2)|$ we assume, w.l.g. that the two vectors differ in the first variable z_1. Applying second-order Taylor expansion for the first variable z_1 we obtain that for some value θ between z_1^1 and z_1^2,

$$g(z^2) = g(z^1) + \frac{\partial g(z^1)}{\partial z_1}(z_1^2 - z_1^1) + \frac{1}{2}\frac{\partial_2 g(\theta)}{\partial^2 z_1}(z_1^2 - z_1^1)^2 \tag{15}$$

For convenience, denote $\prod_{2 \le j \le k} z_j$ by A, and $\prod_{2 \le j \le k} z_j \sum_{2 \le j \le k} z_j^{-1}$ by B. Trivially, we have $A < 1, B < k - 1 \le 2$. We have $g(z) = \frac{Bz_1 + A}{(1+Az_1)^2}$, and

$$\frac{\partial g(z)}{\partial z_1} = \frac{B(1+Az_1)^2 - 2(Bz_1 + A)(1+Az_1)A}{(1+Az_1)^4}$$
$$= \frac{B(1+Az_1) - 2(Bz_1 + A)A}{(1+Az_1)^3}$$

The absolute value of the last expression does not exceed $\max(B, 2(B + A)A) \le 6$, using the bounds on A, B and $1 + Az_1 > 1, 0 < z_1 < 1$. Then the absolute value of the second term in the sum in (15) is bounded by 6×10^{-3}. We now bound the term corresponding to the second derivative, which find to be

$$\frac{\partial_2 g(z)}{\partial^2 z_1} = \frac{BA(1+Az_1)^3 - 2BA(1+Az_1)^3 - (B(1+Az_1) - 2(Bz_1 + A)A)3(1+Az_1)^2 A}{(1+Az_1)^6}.$$

We very crudely upper bound the absolute value of $\frac{\partial_2 g(z)}{\partial^2 z_1}$ as

$$BA + 2BA + (B + 2(B + A)A)(3A) < 2 + 4 + (2 + 6)3 = 30,$$

again using the bounds $A < 1, B < 2, 0 < z_1 < 1, Az_1 + 1 > 1$. Thus the third term in the sum (15) is upper bounded by $(1/2)30 \times .001^2 = 15 \times 10^{-6}$. Combining, we obtain

from (15) and the obtained bounds on the first and second derivative, that $|g(z^1) - g(z^2)| < 6 \times 10^{-3} + 15 \times 10^{-6} < 7 \times 10^{-3}$. We established (14). This completes the proof of the lemma.   ∎

Application of the lemma to the root node v_0 yields, $c(v_0) - a(v_0) \leq (0.9)^{t-2}$. Combining this with (12) applied to v_0 gives for every two boundary conditions b_1, b_2

$$|\mathbb{P}(v_0 \notin \boldsymbol{I}|b_1) - \mathbb{P}(v_0 \notin \boldsymbol{I}|b_2)| \leq c(v_0) - a(v_0) \leq (0.9)^{t-2}.$$

This establishes (6) and completes the proof the first part of the proposition.

As we mentioned earlier, the second part of the proposition is the result already established in Kelly [24] and we simply refer to his paper for the proof.   ∎

4.2. An Algorithm and the Proof of Theorem 5

Proposition 3 establishes the key correlation decay property for independent sets for trees with maximum degree at most 4. It shows that the marginal Gibbs probability at the root is asymptotically independent from the boundary. Equipped with this result and Proposition 1, we propose the following algorithm for estimating the number of independent sets of a given graph G.

Algorithm CountIND

```
INPUT: A graph G with a node set v₁,…,vₙ and parameter
ϵ > 0.
BEGIN
1. Compute the girth g(G). If (.9)^{g(G)/2 - 2} ≥ ϵ compute 𝓘(G) by
exhaustive enumeration. Otherwise,
2. Set Z = 1, t = g(G)/2.
FOR k = 1,2,…,n:
3. identify depth−t neighborhood T(vₖ) of vₖ in Gₖ₋₁ = G \
{v₁,…,vₖ₋₁};
4. Perform subroutine CountingTREE on T(vₖ) which results
in some value p(v). Set Z = Zp⁻¹(v).
END
OUTPUT: Z.
```

Subroutine CountingTREE

```
INPUT: A tree T with an identified root v and depth t.
BEGIN
1. Identify the nodes u in level t (if any exist) and set
p(u) = 1/2.
FOR l = t − 1, t − 2, …, 0
Identify a node u in level l (if any exist). If u has no
children, set p(u) = 1/2. Otherwise set p(u) = 1/(1 + ∏ p(uᵢ)),
where the product runs over children uᵢ of u in level l + 1
and the values p(uᵢ) were obtained in an earlier step.
END
OUTPUT: p(v).
```

Proof. *Proof of Theorem 5.* We claim that the algorithm `CountIND` provides PAS. Fix a family of graphs $\mathcal{G}$ with maximum degree $r \leq 4$ and large girth, a graph $G \in \mathcal{G}$ and $\epsilon > 0$. The algorithm first checks whether $g(G) > 4 + 2\log(1/\epsilon)/\log(10/9)$. By definition there exists a finite number of graphs in $\mathcal{G}$ with girth $\leq 4 + 2\log(1/\epsilon)/\log(10/9)$ and their corresponding values of $\mathcal{I}$ can be found in constant time, where the constant depends on ϵ and the growth rate f of girth.

Otherwise the girth satisfies $(0.9)^{\frac{g(G)}{2}-2} < \epsilon$ and in the remaining n steps of the algorithm the Gibbs marginal probability $\mathbb{P}(v_k \in \boldsymbol{I})$ is computed with respect to the depth $t = g(G)/2$ neighborhood $T(v_k)$ of the node v_k with respect to the graph G_{k-1}. By selection of t, $T(v_k)$ is a tree (the girth of each subgraph G_{k-1} is trivially at least $g(G)$). Let $B(T(v_k))$ be the boundary of $T(v_k)$ and consider the graph $\hat{G}_{k-1} = (G_{k-1} \setminus T(v_k)) \cup B(T(v_k))$, that is everything but the first $t-1$ levels of $T(v_k)$. For every independent set I on G_{k-1}, let $J = I \cap \hat{G}_{k-1}$ and let $J \cap B(T(v_k)) = I \cap B(T(v_k)) = b(I)$. $b(I)$ is a boundary condition implied by the entire independent set I. Observe that we have by Markovian property of Gibbs distributions

$$\mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I}) = \mathbb{E}_{\mathbb{P}_{G_{k-1}}}[\mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I}|\boldsymbol{J})] = \mathbb{E}_{\mathbb{P}_{G_{k-1}}}[\mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I}|b(\boldsymbol{I}))]$$
$$= \sum_b \mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I}|b(\boldsymbol{I}) = b)\mathbb{P}_{G_{k-1}}(b(\boldsymbol{I}) = b).$$

But for each boundary condition $b \in B(T(v_k))$

$$\mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I}|b(\boldsymbol{I}) = b) = \mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}|b(\boldsymbol{I}) = b).$$

Also

$$\mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}) = \mathbb{E}_{\mathbb{P}_{T(v_k)}}[\mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}|b(\boldsymbol{I}))] = \sum_b \mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}|b(\boldsymbol{I}) = b)\mathbb{P}_{T(v_k)}(b(\boldsymbol{I}) = b).$$

We see that $\mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I})$ and $\mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I})$ can be represented as convex combinations of values $\mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}|b(\boldsymbol{I}) = b)$, $b \in B(T(v_k))$. On the other hand, Proposition 3 implies that for every $b, b' \in B(T(v_k))$,

$$|\mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}|b) - \mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}|b')| < (.9)^{t-2} = (.9)^{\frac{g(G)}{2}-2} < \epsilon.$$

It follows that
$$|\mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}) - \mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I})| < \epsilon.$$

The lower bound part of (5) gives $\mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}) \geq 1/(1+\lambda) = .5$. Then

$$\left|\mathbb{P}_{T(v_k)}^{-1}(v_k \notin \boldsymbol{I}) - \mathbb{P}_{G_{k-1}}^{-1}(v_k \notin \boldsymbol{I})\right| = \mathbb{P}_{G_{k-1}}^{-1}(v_k \notin \boldsymbol{I})\left|\frac{\mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I}) - \mathbb{P}_{G_{k-1}}(v_k \notin \boldsymbol{I})}{\mathbb{P}_{T(v_k)}(v_k \notin \boldsymbol{I})}\right|$$
$$< \mathbb{P}_{G_{k-1}}^{-1}(v_k \notin \boldsymbol{I})\frac{\epsilon}{.5}.$$

We conclude

$$\mathbb{P}_{G_{k-1}}^{-1}(v_k \notin \boldsymbol{I})(1 - 2\epsilon) \leq \mathbb{P}_{T(v_k)}^{-1}(v_k \notin \boldsymbol{I}) \leq \mathbb{P}_{G_{k-1}}^{-1}(v_k \notin \boldsymbol{I})(1 + 2\epsilon)$$

The value $\mathbb{P}_{T(v_k)}^{-1}(v_k \notin I)$ is what algorithm CountTREE outputs as $p^{-1}(v)$. Therefore, applying Proposition 1, we have that Z, the product of these outputs satisfies

$$Z(1, G)(1 - 2\epsilon)^n = \prod_{k=1}^{n} \mathbb{P}_{G_{k-1}}^{-1}(v_k \notin I)(1 - 2\epsilon)^n$$

$$\leq Z \leq \prod_{k=1}^{n} \mathbb{P}_{G_{k-1}}^{-1}(v_k \notin I)(1 + 2\epsilon)^n = Z(1, G)(1 + 2\epsilon)^n.$$

Using $|\log(1 - 2\epsilon)| < 3\epsilon$ for sufficiently small ϵ, we obtain

$$\left| \frac{\log Z}{n} - \frac{\log Z(1, G)}{n} \right| < 3\epsilon.$$

Finally, we observe that since, by bounds (13) each element of the product Z belongs to the interval $[1 + \lambda(1 + \lambda)^{-r+1}, (\lambda + 1)/\lambda] = [9/8, 2/1]$, then $\log Z/n \geq \log(9/8)$. Therefore

$$(1 - 3\epsilon \log^{-1}(9/8)) \leq \frac{\log Z}{\log Z(1, G)} \leq (1 + 3\epsilon \log^{-1}(9/8)).$$

Thus the algorithm CountIND is PAS for counting independent sets. ∎

4.3. Regular Graphs and Proof of Theorem 1

The second part of Proposition 3 provides an explicit limiting expression for the probability that a given node belongs to an independent set selected according to the Gibbs distribution. In this subsection, we use it to obtain explicit asymptotics for the logarithm of the number of independent sets in regular graphs. Theorem 5 provides a way in principle for computing number of independent sets in regular graph. The problem is, however, in the fact that the cavity step expressed in (1) destroys regularity: when node v_1 is removed, the remaining graph is no longer regular and it is not clear how to estimate product (2) explicitly. The help comes from a trick introduced by Mezard and Parisi [32], also used in [34] in a context of random regular graph. Given an n-node r-regular graph G fix any two nodes v_1, v_2 which are not neighbors, and do not have common neighbors (if there are any) and denote their nonoverlapping neighbor sets by $v_{11}, \ldots, v_{1r}$ and $v_{21}, \ldots, v_{2r}$, respectively. Consider a modified graph G^o obtained from G by deleting v_1, v_2 and connecting v_{1j} to v_{2j} by an edge for all $j = 1, \ldots, r$, see Fig. 1 for an example when $r = 3$. The resulting graph is r-regular again. We call this operation "rewiring" or "rewire" operation. Rewiring was used in [32] and [34] in the context of random regular graphs, and it was performed on two nodes selected randomly from the graph. The main question is whether we can relate the partition functions of the original and modified graphs and whether the resulting graph still has a sufficiently large girth, provided the original one does. The first issue has been addressed in [34] and is essentially a simple combination of arguments similar to that used to derive (1). The second issue was not addressed in [34] in a rigorous way. It was just postulated that the resulting graph again has a large girth if the two nodes are selected uniformly at random.

We begin by addressing the second issue first.

Lemma 2. *Given an n-node r-regular graph G, consider any integer $4 \leq g \leq g(G)$. The rewiring operation can be performed for at least $(n/2) - (2g + 1)r^{2g}$ steps on pairs of nodes*
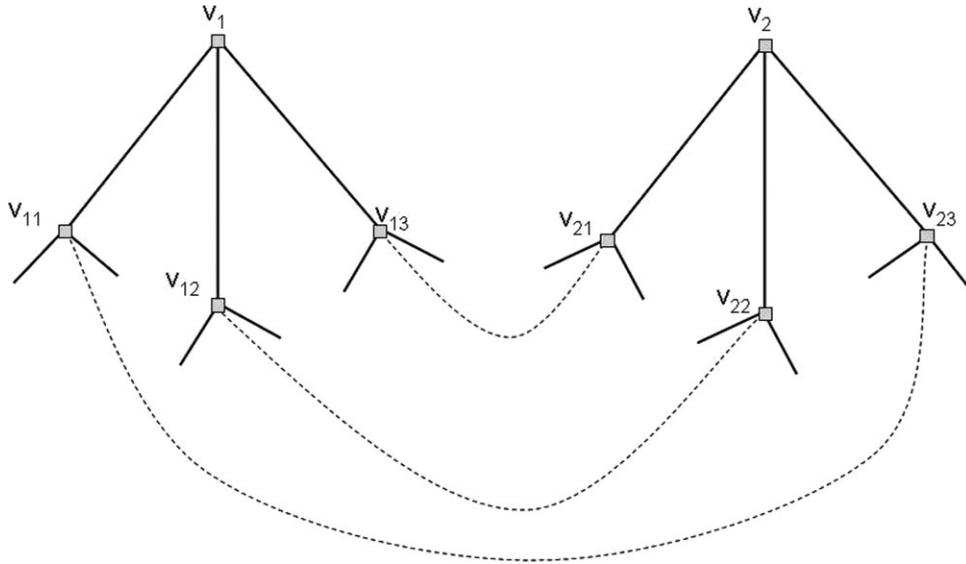
Fig. 1. Rewiring on nodes v_1 and v_2.

which are at least $2g + 1$ distance apart. In every step the resulting graph is r-regular with girth at least g.

Proof. In every step of the rewiring we delete two nodes in the graph. Thus when we performed $t \leq (n/2) - (2g + 1)r^{2g}$ successful rewiring steps, in the end we obtain a graph with at least $n - 2((n/2) - (2g + 1)r^{2g}) = 2(2g + 1)r^{2g}$ nodes. Suppose in step $t \leq (n/2) - (2g + 1)r^{2g}$ we have a graph G_t which is r-regular and has girth at least g. We claim that the diameter of this graph is at least $2g + 1$. Indeed, if the diameter is smaller, then for a given node v any other nodes is reachable from v by a path with distance at most $2g$ and the total number of nodes is at most $\sum_{0 \leq k \leq 2g} r^k < (2g + 1)r^{2g}$, which is a contradiction. Now select any two nodes $v_1, v_2 \in G_t$ which are at the distance equal to the diameter of this graph, and thus are at least $2g + 1$ edges apart. We already showed that the graph G_{t+1} obtained by rewiring G_t on v_1, v_2 is r-regular. It remains to show it has a girth at least g. Suppose, for the purposes of contradiction, G_t has girth $\leq g - 1$. Suppose $k \geq 1$ is the number of newly created edges which participate in creating a cycle with length $\leq g - 1$. If $k = 1$ and v_{1j}, v_{2j} is the pair creating the unique participating edge, then the original distance between v_{1j} and v_{2j} was at most $g - 2$ by following a path on the cycle which does not use the new edge. But then the distance between v_1 and v_2 is at most $g < 2g + 1$, gives a contradiction. Now suppose $k > 1$, then there exists a path of length at most $(g - 1)/k \leq (g - 1)/2$ which uses only the original edges (the edges of the graph G_t) and connects a pair v, v' of nodes from the set $v_{11}, \ldots, v_{1r}, v_{21}, \ldots, v_{2r}$. If the pair is from the same set, say $v_{11}, \ldots, v_{1r}$, then, since these two nodes are connected to v_1, we obtain a cycle in G_t with length $(g - 1)/2 + 2 < g$, leading to a contradiction (since by assumption $g > 3$). If these two nodes are from different sets, for example $v = v_{1j}, v' = v_{2l}$, then we obtain that the distance between v_1 and v_2 in G_t is at most $(g - 1)/2 + 2 < 2g + 1$, which also leads to a contradiction. So we conclude that G_t got to have girth at least g as well. ∎

We now turn to the second problem of estimating the relative change of the partition function after rewiring. This relative change is called *energy shift* in [34]. First we provide an elementary analogue of (1).

Lemma 3. *Given an r-regular graph G, given $\lambda > 0$ and graph G^o obtained from G by rewiring on nodes $v_1, v_2 \in G$, the following relation holds*

$$\frac{Z(\lambda, G^o)}{Z(\lambda, G)} = \mathbb{P}_G(v_1, v_2 \notin \boldsymbol{I})\mathbb{P}_{G\setminus\{v_1,v_2\}}(\wedge_{1\leq j\leq r}(v_{1j} \notin \boldsymbol{I} \vee v_{2j} \notin \boldsymbol{I}))$$

where $v_{ij}, j = 1, \ldots, r$ is the set of neighbors of $v_i, i = 1, 2$ in G.

Proof. The proof is almost identical to the one of Proposition 1. The partition function $Z(\lambda, G^o)$ is obtained as a sum $\lambda^{|I|}$ over the set of independent subsets $I \subset V(G)$, which do not contain v_1, v_2 and which contain at most one of the two nodes v_{1j}, v_{2j} for each $j = 1, 2, \ldots, r$.   ∎

We now obtain a very simple limiting expression for the probability in Lemma 3.

Lemma 4. *Given $r \geq 3, \lambda < (r-1)^{r-1}/(r-2)^r$ and $\epsilon > 0$, there exists a sufficiently large constant $g = g(r, \epsilon, \lambda)$ such that for every graph G with girth $g(G) \geq g$, and for every pair of nodes $v_1, v_2 \in G$ at distance at least $2g + 1$*

$$\left| \mathbb{P}_G(v_1, v_2 \notin \boldsymbol{I}) - \frac{1}{(2-x)^2} \right| < \epsilon, \tag{16}$$

and

$$|\mathbb{P}_{G\setminus\{v_1,v_2\}}(\wedge_{1\leq j\leq r}(v_{1j} \notin \boldsymbol{I} \vee v_{2j} \notin \boldsymbol{I})) - (2x - x^2)^r| < \epsilon, \tag{17}$$

where $v_{ij}, j = 1, \ldots, r$ is the set of neighbors of v_i in G, $i = 1, 2$, and x is the unique positive solution of $x = 1/(1 + \lambda x^{r-1})$.

Proof. The proof consists of several steps, each ideologically very similar to the one for Theorem 5. Fix $\epsilon > 0$ and let $g = g(\epsilon, r, \lambda)$ be a large value to be specified later. Select $\alpha = \alpha(\lambda, r)$ as in Proposition 3. We consider any r-regular graph with girth at least g and consider any two nodes v_1, v_2 in G at distance at least $2g + 1$, if such two nodes exist. Consider depth $t = \lfloor g/2 \rfloor$ neighborhoods $T(v_1), T(v_2)$. By the distance assumption, they do not intersect, and by the girth assumption, each neighborhood is a depth-t r-regular tree. First we estimate the impact of deleting these nodes v_1, v_2 from G. That is we consider $Z(\lambda, G \setminus \{v_1, v_2\})/Z(\lambda, G)$. Then we will consider G^o obtained by rewiring G on v_1, v_2 and estimate $Z(\lambda, G^o)/Z(\lambda, G \setminus \{v_1, v_2\})$.

Fix any independent set I on $\hat{G} = B(T(v_1)) \cup B(T(v_2)) \cup (G \setminus (T(v_1) \cup T(v_2)))$, where $B(T)$ is again the boundary of a tree T. Let $b_i = I \cap B(T(v_i)), i = 1, 2$. Let $\boldsymbol{I}$ be the random independent set in G selected according to the Gibbs distribution with parameter λ. Using Markov property of the Gibbs distribution

$$\mathbb{P}_G(v_1, v_2 \notin \boldsymbol{I}|\boldsymbol{I} \cap \hat{G} = I) = \mathbb{P}_G(v_1 \notin \boldsymbol{I}|\boldsymbol{I} \cap \hat{G} = I)\mathbb{P}_G(v_2 \notin \boldsymbol{I}|\boldsymbol{I} \cap \hat{G} = I)$$
$$= \mathbb{P}_{T(v_1)}(v_1 \notin \boldsymbol{I}|b_1)\mathbb{P}_{T(v_2)}(v_2 \notin \boldsymbol{I}|b_2)$$

From the second part of Proposition 3

$$\left| \mathbb{P}_{T(v_i)}(v_i \notin \boldsymbol{I} | b_i) - \frac{1}{2-x} \right| < \alpha^t, \quad i = 1, 2,$$

which results in

$$\left| \mathbb{P}_G(v_1, v_2 \notin \boldsymbol{I} | \boldsymbol{I} \cap \hat{G} = I) - \left(\frac{1}{2-x} \right)^2 \right| \leq \alpha^t + \alpha^t \frac{1}{2-x}.$$

Because the relation holds for every independent set I then

$$\left| \mathbb{P}_G(v_1, v_2 \notin \boldsymbol{I}) - \left(\frac{1}{2-x} \right)^2 \right| \leq \alpha^t + \alpha^t \frac{1}{2-x}.$$

We take $g = g(\epsilon, r, \lambda)$ sufficiently large, so that the absolute difference above is at most ϵ (note that the choice depends on α which itself is controlled by λ and r). This concludes the proof of the first part of the lemma.

Now consider $\mathbb{P}_{G_1^o}(\wedge_{1 \leq j \leq r}(v_{1j} \notin \boldsymbol{I} \vee v_{2j} \notin \boldsymbol{I}))$. We take depth-$(t-1)$ neighborhoods of $v_{ij}, k = 1, 2, j = 1, \ldots, r$ and again observe that they are all nonintersecting trees because of the girth and distance between v_1 and v_2 assumption. By conditioning on the realizations I of a random independent set $\boldsymbol{I}$ in $\hat{G}_1 = (G_1^o \setminus \cup_{i,j} T(v_{ij})) \cup (\cup_{i,j} B(T(v_{ij})))$, letting $b_{ij} = I \cap B(T(v_{ij}))$ and using the same argument as above, we obtain

$$\mathbb{P}_{G_1^o}(\wedge_{1 \leq j \leq r}(v_{1j} \notin \boldsymbol{I} \vee v_{2j} \notin \boldsymbol{I}) | \boldsymbol{I} \cap \hat{G}_1 = I)$$
$$= \prod_{1 \leq j \leq r} \left(\mathbb{P}_{T(v_{1j})}(v_{1j} \notin \boldsymbol{I} | b_{1j}) + \mathbb{P}_{T(v_{2j})}(v_{2j} \notin \boldsymbol{I} | b_{2j}) - \mathbb{P}_{T(v_{1j})}(v_{1j} \notin \boldsymbol{I} | b_{1j}) \mathbb{P}_{T(v_{2j})}(v_{2j} \notin \boldsymbol{I} | b_{2j}) \right)$$
$$= \prod_{1 \leq j \leq r} \left(1 - \mathbb{P}_{T(v_{1j})}(v_{1j} \in \boldsymbol{I} | b_{1j}) \mathbb{P}_{T(v_{2j})}(v_{2j} \in \boldsymbol{I} | b_{1j}) \right)$$

Again we use bound provided by Proposition 3

$$\left| \mathbb{P}_{T(v_{ij})}(v_{ij} \in \boldsymbol{I} | b_{ij}) - (1-x) \right| < \alpha^{t-1}, i = 1, 2, j = 1, 2, \ldots, r,$$

(we recall that each tree $T(v_{ij})$ has depth $t-1$ and the root v_{ij} of this tree has degree $r-1$). We now take $t = g/2 = g(\epsilon, r, \lambda)/2$ sufficiently large so that

$$\left| \mathbb{P}_{G_1^o}(\wedge_{1 \leq j \leq r}(v_{1j} \notin \boldsymbol{I} \vee v_{2j} \notin \boldsymbol{I}) | \boldsymbol{I} \cap \hat{G}_1 = I) - (1 - (1-x)^2)^r \right| < \epsilon.$$

Since the relation holds for every independent set I, then

$$\left| \mathbb{P}_{G_1^o}(\wedge_{1 \leq j \leq r}(v_{1j} \notin \boldsymbol{I} \vee v_{2j} \notin \boldsymbol{I})) - (2x - x^2)^r \right| < \epsilon.$$

This completes the proof of the lemma.   ■

Proof. *Proof of Theorem 1.* The proof is obtained by combining the results of Lemmas 2,3,4. From the last two lemmas, for every ϵ we can find $g = g(\epsilon, r, \lambda)$ sufficiently large so that for every graph G with girth at least $g + 1$ and for every two nodes v_1, v_2 at distance at

least $2g + 1$, the graph G^o obtained from G by rewiring on v_1, v_2 satisfies, after simplifying $(2 - x)^{-2}(2x - x^2)^r$ to $x^r(2 - x)^{r-2}$, the following bounds.

$$(1 - \epsilon)x^r(2 - x)^{r-2} \leq \frac{Z(\lambda, G^o)}{Z(\lambda, G)} \leq (1 + \epsilon)x^r(2 - x)^{r-2}.$$

Here we note that in order to combine the individual absolute differences (16) and (17), we need to take $g = g(\epsilon, r, \lambda)$ which is sufficiently large with taking x into account. But x itself depends only on λ. Therefore such g indeed exists. By Lemma 2, if the original graph G has n nodes, then the rewiring can be performed for at least $N = n/2 - C = n/2 - C(g, r) = n/2 - C(\epsilon, r, \lambda)$ steps, and at most $n/2$ steps, where constant $C = C(g, r) = (2g + 1)r^{2g}$. Let G^* denote the graph obtained from G after N rewiring steps. Then from the bound above

$$(1 - \epsilon)^{\frac{n}{2}-C}(x^r(2 - x)^{r-2})^{\frac{n}{2}-C} \leq \frac{Z(\lambda, G^*)}{Z(\lambda, G)} \leq (1 + \epsilon)^{\frac{n}{2}}(x^r(2 - x)^{r-2})^{\frac{n}{2}}$$

Since the number of nodes in G^* is at most $2C$, then trivially $Z(\lambda, G^*) \leq (1 + \lambda)^{2C}$. We obtain for sufficiently large $n(\epsilon, r, x, C) = n(\epsilon, r, \lambda)$, that for all $n \geq n(\epsilon, r, \lambda)$

$$\left| \frac{\log Z(\lambda, G)}{n} - \log x^{-\frac{r}{2}}(2 - x)^{-\frac{r-2}{2}} \right| < 2\epsilon.$$

This concludes the proof of the first part of the theorem.

The case $\lambda = 1$ corresponds to the counting problem. We check that $(r-1)^{r-1}/(r-2)^r > 1$ only for $r = 2, 3, 4, 5$ and thus for these values we can obtain the asymptotics of the log-partition function, and we do so now.

In the special case $r = 2$ and $\lambda = 1$ we find that $x = \frac{\sqrt{5}-1}{2} \approx 0.6180$, derived from the golden ratio equation $x = 1/(1+x)$. Thus the total number of independent sets $\mathcal{I}(G)$ in every 2-regular graphs with large girth is $\approx (\frac{2}{\sqrt{5}-1})^n \approx (1.618\ldots)^n$. As a sanity check there is a simple way to see the validity of this answer, for example in a special case when the graph is an n-cycle. We note that for every node v on a cycle, if it belongs to the independent set, its right-hand side neighbor v' does not, but if v does not, then v' either belongs or does not belong to the independent set. It is a simple exercise to see that the number of independent sets which can be created on a path of length k starting from v and going to the right is

$$(1 \quad 1) \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^{k-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The growth rate of this expression is determined by the largest eigenvalue of the matrix, which is the golden ration value $2/(\sqrt{5} - 1)$. Thus on the path of length n the number of independent sets is $\approx (2/(\sqrt{5} - 1))^n$. The number of independent sets on a cycle differs from this only by a constant factor (to adjust for the fact that the last node and the first node v of the interval are merged to create a cycle).

When $r = 3, \lambda = 1$, the solution x to the equation $x = 1/(1 + x^2)$ is found numerically to be $x = 0.682\ldots$. Thus $\mathcal{I}(G)$ for every 3-regular graph with large girth has cardinality $\approx (1.545\ldots)^n$. When $r = 4, \lambda = 1$, we find similarly that the answer is $\approx (1.494\ldots)^n$ and when $r = 5$ it is $\approx (1.453\ldots)^n$. This concludes the proof of Theorem 1. ∎

4.4. Random Regular Graphs and Proof of Theorem 3

We now prove Theorem 3.

Proof. We use the following fact about random regular graphs (see [19]): given any constant $g > 0$ the expected total number of cycles with length $< g$ is at most some constant $k(g)$. Our proof approach consists of "taking care" of small cycles by modifying a graph into a new graph $\hat{G}$ which is also r-regular, but has girth at least a given constant g. First, suppose a graph $\hat{G}$ is obtained from G by some constant k number of the following operations: adding/deleting a node or adding/deleting and edge. We claim that

$$|\log Z(\lambda, G) - \log Z(\lambda, \hat{G})| \leq c(\lambda, k), \tag{18}$$

that is the difference is at most a constant. Indeed every time we add a node, without adding any edges, the partition function increases by at most a factor $1 + \lambda$. Therefore, by the reverse argument, deleting a node decreases the partition function also by at most this factor. Now suppose we delete an edge. Any newly created independent set can be obtain from an independent set of the original graph by adding one node, which is one of the end nodes of the deleted edge. Also each independent set in the original graph can participate in creating at most two new independent sets. Thus the sum $\sum_I \lambda^{|I|}$ corresponding to newly created independent sets is at most $2\lambda Z(\lambda, G)$, and, as a result, the partition function increases by at most factor $(1 + 2\lambda)$. The case of edge addition is considered similarly. This proves the assertion (18).

We now return to the proof of the theorem. We fix $\epsilon > 0$ and find sufficiently large $g > 0$ such that for every r-regular graph with girth at least g,

$$\left| \frac{\log Z(\lambda, G)}{n} - \log\left(x^{-\frac{r}{2}}(2 - x)^{-\frac{r-2}{2}}\right) \right| < \epsilon/2. \tag{19}$$

Such g exists by Theorem 1. We also find constants $k = k(g)$ and n_0 sufficiently large, so that for $n \geq n_0$, the probability that the random n-node r-regular graph $G_r(n)$ has more than k cycles with length $< g$ is at most ϵ. Consider a random r-regular graph $G = G_r(n)$. Let $C_1, \ldots, C_k$ denote all cycles in G which have length $< g$. In each cycle C_i fix an arbitrary edge (u_1^i, u_2^i), $1 \leq i \leq k$. Note, that the edges might repeat. Observe that if these edges are removed, the resulting graph has girth at least g. We construct now a modified graph $\hat{G}$ from G as follows. Consider an arbitrary constant size r-regular graph H which has girth at least g. The size m of this graph is irrelevant to us as long as it is a constant, namely m does not grow with n. Delete any edge (u_1, u_2) in H. Observe that the distance between u_1 and u_2 in the modified graph $H_0 = H \setminus \{(u_1, u_2)\}$ is at least $g - 1$, since the girth of G_0 is at least g. We now take the original graph G and replace each edge (u_1^i, u_2^i) with a copy of the graph H_0, where u_1^i, u_2^i replace a copy of u_1, u_2 in each replica of the graph, respectively. We denote the created graph by $\hat{G}$. It is clearly r-regular. It is obtained from G by some constant number c of node/edge additions/deletions. Therefore (18) holds for G and $\hat{G}$ for some constant $c(\lambda, H)$.

We claim that $\hat{G}$ has girth at least g. Note that any cycle in $\hat{G}$ can be decomposed into an alternating sequence of paths of two kinds, the first kind being paths from G without the deleted edges (u_1^i, u_2^i), and the second kind being paths in replicas of the graph H_0 connecting some pairs u_1^i to u_2^i. If there are no paths of the second kind (that is there is only one path of the first kind which is also a cycle) then this is a cycle in the original graph G without edges (u_1^i, u_2^i). The length of this cycle is at least g. But the path of the second type

has length at least $g - 1$, since the distance between u_1, u_2 in H_0 is at least $g - 1$. Therefore the total cycle length is again at least g. We conclude that the girth of $\hat{G}$ is at least g. As a result, (19) holds for $\hat{G}$ in place of G. Combining with (18), we obtain that if $n \geq n_0$, then with probability at least $1 - \epsilon$

$$
\left| \frac{\log Z(\lambda, G)}{n} - \log\left(x^{-\frac{r}{2}}(2 - x)^{-\frac{r-2}{2}} \right) \right| \leq \left| \frac{\log Z(\lambda, G)}{n} - \frac{\log Z(\lambda, \hat{G})}{n} \right|
$$

$$
+ \left| \frac{\log Z(\lambda, \hat{G})}{n} - \log\left(x^{-\frac{r}{2}}(2 - x)^{-\frac{r-2}{2}} \right) \right|
$$

$$
\leq \frac{c(\lambda, H)}{n} + \frac{\epsilon}{2}
$$

Now assuming $n \geq \max(n_0, 2c(\lambda, H)/\epsilon)$, we obtain that the difference is at most ϵ. ∎

5. STUDY OF COLORINGS

The general approach for solving the problem of counting the number of proper colorings is the same as for independent sets. We establish correlation decay property for arbitrary graphs with bounded degree and large girth. We construct an algorithm exploiting this correlation decay property. Then we focus on regular graphs, where explicit results can be obtained. Unlike the results for independent sets, our results for coloring do not need explicit bounds on the degree of the graph.

5.1. Coloring of Trees and Correlation Decay

We use the definitions and notations of Subsection 4.1: $T, B(T), \mathcal{B}(T)$ denote respectively an arbitrary depth-t tree with maximum degree at most r, the boundary of the tree and the set of boundary conditions. The latter, however, is defined as the set of functions $b : B(T) \to \{1, 2, \ldots, q\}$ mapping nodes to colors. The root of this tree is v_0. Similarly to the case of independent set, we use notation $\mathbb{P}_G(\boldsymbol{C}(v) = j|b)$ to indicate probability that the random coloring $\boldsymbol{C}$ assigns color j to the node $v \in T$, subject to the boundary condition b, where probability is with respect to the Gibbs measure, (in this case uniform distribution) on the set of all proper colorings.

We need an analog of Proposition 3, and in this case we use the following result by Jonasson [20]. This result was used to establish uniqueness of Gibbs measures for coloring infinite trees, but the main underlying result is a correlation decay property. (We note that Jonasson uses $r + 1$ in place of r for the degree of a tree).

Theorem 7 (Jonasson [20]). *Suppose $q \geq r + 1$. There exists a computable value $\beta = \beta(r) < 1$ such that for every r-regular tree T with depth t*

$$
\sup_{b \in \mathcal{B}(T)} \left| \mathbb{P}(\boldsymbol{C}(v_0) = j|b) - \frac{1}{q} \right| \leq \beta^t,
$$

for every $j = 1, 2, \ldots, q$.

This result says that the color received by the root v_0 is asymptotically independent from the colors of the boundary as a function of the tree depth. Note that the decay constant β does not even depend on q provided that $q \geq r+1$. The analysis of the proof in [20] reveals that the same result holds for non-regular trees as well.

Corollary 1. *The result of Theorem 7 holds when T is an arbitrary depth-t tree with maximum degree r.*

5.2. An Algorithm and the Proof of Theorem 6

We propose the following algorithm for estimating the number of q-colorings of a given graph G.

Algorithm CountCOLOR

```
INPUT: A graph G with maximum degree r, a positive
integer q, such that q ≥ r + 1, and parameter ε > 0.
BEGIN
1. Compute the girth g(G). If β^(g(G)/2 − 2)   ≥   ε compute C(G) by
exhaustive enumeration. Otherwise
2. Set G′ = G, Z = 1, t = g(G)/2.
3. Find any node v ∈ G′ and its degree r′ = r(v, G) ≤ r. Set Z
equal to
```

$$
Z\left[q \left(1 - \frac{1}{q} \right)^{r'} \right]
$$

```
4. Set G′ = G′ \ {v} and go to step 2.
END
OUTPUT: Z.
```

Proof. Proof of Theorem 6. The proof is very similar to the one of Theorem 5. Applying Proposition 2 we need to estimate in each step of the algorithm the expected value of used colors $\mathbb{E}_{G_k}[|C(N(v_k, G_{k-1}))|]$. By fixing any boundary condition on depth-t neighborhood of v_k in the graph G_{k-1} the probability of any particular coloring of the nodes in $N(v_k, G_{k-1})$ is product of individual coloring probabilities. Each individual coloring probability is asymptotically $1/q$ provided t is large by Corollary 1. Therefore given a fixed color $i \leq q$, the probability that this color was never used in coloring nodes $N(v_k, G_{k-1})$ is asymptotically $(1 - 1/q)^{|N(v_k, G_{k-1})|}$. Therefore $q - \mathbb{E}_{G_k}[|C(N(v_k, G_{k-1}))|]$ is asymptotically $q(1 - 1/q)^{|N(v_k, G_{k-1})|}$, provided that $t = g(G)/2$ is sufficiently large.

The rest of the argument follows the lines the proof of Theorem 5. ∎

5.3. Regular Graphs and Proof of Theorem 2

Our main tool is again rewiring performed on regular graphs with large girth. Given an arbitrary graph G and nodes $v_1, v_2 \in G$ such that v_1 and v_2 are not neighbors, and they do not have a common neighbor, let G^o be obtained from G by rewiring on v_1, v_2. Proposition 2 already relates the partition function of G to the one of $G \setminus \{v_1, v_2\}$. We now relate it to

the one of G^o. Consider a random uniform q-coloring C selected in $G \setminus \{v_1, v_2\}$. The lemma below does not rely on assumptions of regularity or the girth size of the underlying graph G.

Lemma 5. *The following relation holds*

$$\frac{Z(q, G)}{Z(q, G^o)} = \frac{\mathbb{E}_{G \setminus \{v_1, v_2\}}[(q - |C(N(v_1, G))|)(q - |C(N(v_2, G))|)]}{\mathbb{P}_{G \setminus \{v_1, v_2\}}(C(v_{1j}) \neq C(v_{2j}), 1 \leq j \leq r)},$$

where $v_{ij}, j = 1, \ldots, r$ is the set of neighbors of $v_i, i = 1, 2$ in G.

Proof. Using the same argument as in Proposition 2 we obtain that

$$\frac{Z(q, G)}{Z(q, G \setminus \{v_1, v_2\})} = \mathbb{E}_{G \setminus \{v_1, v_2\}}[(q - |C(N(v_1, G))|)(q - |C(N(v_2, G))|)].$$

On the other hand $\frac{Z(q, G_0)}{Z(q, G \setminus \{v_1, v_2\})}$ is the probability that a randomly selected coloring in $G \setminus \{v_1, v_2\}$ assigns different colors to each pair $v_{1j}, v_{2j}, j = 1, 2, \ldots, r$. Combining, we obtain the result. ∎

The following lemma is an analogue of Lemma 4.

Lemma 6. *Given $r \in \mathbb{N}$, $q \geq r + 1$ $\epsilon > 0$, there exists a sufficiently large constant $g = g(r, \epsilon)$ such that for every r-regular graph G with girth $g(G) > g$, for every pair of nodes $v_1, v_2 \in G$ at distance at least $2g + 1$*

$$\left| \mathbb{E}_{G \setminus \{v_1, v_2\}}[(q - |C(N(v_1, G))|)(q - |C(N(v_2, G))|)] - q^2 \left(1 - \frac{1}{q}\right)^{2r} \right| < \epsilon. \quad (20)$$

$$\left| \mathbb{P}_{G \setminus \{v_1, v_2\}}(\wedge_{1 \leq j \leq r} C(v_{1j}) \neq C(v_{2j})) - \left(\frac{q-1}{q}\right)^r \right| < \epsilon. \quad (21)$$

Proof. The proof is very similar to the one of Lemma 4. In the graph $G \setminus \{v_1, v_2\}$ consider depth-$t = g/2$ neighborhoods of nodes v_{ij}. By the girth assumption these neighborhoods are non-intersecting r-regular trees T_{ij}, with the exception that each root v_{ij} has degree $r - 1$. Fix any collection of colors $c_{ij} \in \{1, 2, \ldots, q\}, i = 1, 2, j = 1, 2, \ldots, r$. Applying Corollary 1 and using the fact that the tree T_{ij} are non-intersecting, we obtain that

$$\left| \mathbb{P}_{G \setminus \{v_1, v_2\}}(C(v_{ij}) = c_{ij}, \forall i, j) - \frac{1}{q^{2r}} \right|, \quad (22)$$

can be made arbitrarily small by making $g = g(\epsilon, r, q)$ sufficiently large. Thus, under $\mathbb{P}_{G \setminus \{v_1, v_2\}}$ the random colors $\{C(v_{ij})\}$ are asymptotically independent and each uniformly distributed on the set of colors $\{1, 2, \ldots, q\}$. Thus (20) and (21) follows by making g sufficiently large. ∎

Proof of Theorem 2. The proof follows the same steps as the proof of Theorem 1. The results of Corollary 1 and Lemmas 2, 5, 6 are combined to obtain the limiting expression after the cancelation of $(\frac{q-1}{q})^r$. ∎

5.4. Random Regular Graphs and Proof of Theorem 4

The proof of Theorem 4 can be established in exactly the same manner as done in the proof of Theorem 3.

6. CONCLUSIONS

We have presented in this paper a new method for solving approximately some counting problems, which is not based on the Markov Chain sampling technique. We applied our method to independent sets and colorings in low degree graphs with large girth. The primary technical tool is a derivation of a certain correlation decay property which features prominently in statistical physics literature in connections with a completely different topic: uniqueness of Gibbs distributions on infinite trees. We certainly hope that our approach is more general and can be applied to other combinatorial problems. This constitutes an interesting direction for further research.

Another research direction is removing the requirement of large girth, and here the difficulty is establishing correlation decay in non-tree like graphs. Such correlation decay was already established by Dobrushin [14] back in 70's for lattice like graphs, but there is a recent extension by Weitz [40] to a more general graphs. Perhaps this correlation decay (long-range independence) can be exploited to obtain non-Markov chain type algorithms for counting problems. It would be also of interest to weaken the form of the correlation decay used in this paper, perhaps to the correlation decay with respect to the typical boundary condition obtained by a broadcasting mechanism (reconstruction). The relevance of this form of the correlation decay is yet to be seen.

Finally, it would be interesting to see if our approach can be converted to an algorithm for sampling from the uniform distribution, for example of independent set or coloring in the same class of low degree graphs with large girth. This would be a nice supplement to the classical approach of rapidly mixing Markov chains.

Remark. After the preparation of the manuscript we have been informed about a recent paper by Weitz [41] on algorithms for counting independent sets in general graphs with maximum degree 5, which do not necessarily have large girth. Weitz's work is perhaps an even stronger testimony to the viability of statistical physics based approach for counting problems.

REFERENCES

[1] D. Aldous and A. Bandyopadhyay, A survey of max-type recursive distributional equations, Ann Appl Probab 15 (2005), 1047–1110.

[2] D. Aldous, The $\zeta(2)$ limit in the random assignment problem, Random Struct Algorithms (2001), 381–418.

[3] D. Achlioptas and C. Moore, The chromatic number of random regular graphs, 8th edition, Workshop on Randomization and Computation (RANDOM), 2004.

[4] D. Aldous and J. M. Steele, The objective method: Probabilistic combinatorial optimization and local weak convergence, In: H. Kesten, editor, Discrete combinatorial probability, Springer-Verlag, New York, 2003, pp. 1–72.

[5] A. Bandyopadhyay, Hard-core model on random graphs, in preparation.

[6] A. Bandyopadhyay, Bivariate uniqueness in the logistic fixed point equation, Technical Report 629, Department of Statistics, UC, Berkeley, 2002.

[7] I. Bezakova, D. Stefankovic, V. Vazirani, and E. Vigoda, Accelerated simulated annealing algorithm for the permanent and combinatorial counting problems, SIAM J Comput 37 (2008), 1429–1454.

[8] G. Brightwell and P. Winkler, Random colorings of a Cayley tree, In B. Bollobas, editor, Contemporary Combinatorics, Bolyai Society Mathematical Studies, 2002, pp. 247–276.

[9] G. Brightwell and P. Winkler, Graph homomorphisms and long range action, In Nesetril and Winkler, editors, Graphs, morphisms and statistical physics, DIMACS series in discrete mathematics and computer science, 2004, pp. 29–47.

[10] G. Brightwell and P. Winkler, A second threshold for the hard-core model on a Bethe lattice, Random Struct Algorithms 24 (2004), 303–314.

[11] M. E. Dyer, A. Frieze, and R. Kannan, A random polynomial time algorithm for approximating the volume of convex bodies, J Assoc Comput Machinery 38 (1991), 1–17.

[12] M. Dyer, A. Frieze, T. Hayes, and E. Vigoda, Randomly coloring constant degree graphs, In Proceedings of 45th IEEE Symposium on Foundations of Computer Science, Rome, Italy, October 2004.

[13] M. Dyer, L. A. Goldberg, and M. Jerrum, Counting and sampling H-colourings, Inform Comput 189 (2004), 1–16.

[14] R. L. Dobrushin, Prescribing a system of random variables by the help of conditional distributions, Theory Probab Appl 15 (1970), 469–497.

[15] D. Gamarnik, Linear phase transition in random linear constraint satisfaction problems, Probab Theory Related Fields 129 (2004), 410–440.

[16] H. O. Georgii, Gibbs measures and phase transitions, de Gruyter studies in mathematics 9, Walter de Gruyter & Co., Berlin, 1988.

[17] D. Gamarnik, T. Nowicki, and G. Swirszcz, Invariant probability measures and dynamics of exponential linear type maps, Ergodic Theory and Dynamical Systems, to appear.

[18] D. Gamarnik, T. Nowicki, and G. Swirscsz, Maximum weight independent sets and matchings in sparse random graphs. Exact results using the local weak convergence method, Random Struct Algorithms 28 (2006), 76–106.

[19] S. Janson, T. Łuczak, and A. Rucinski, Random graphs, Wiley, New York, 2000.

[20] J. Jonasson, Uniqueness of uniform random colorings of regular trees, Statistics Probab Lett 57 (2002), 243–248.

[21] M. Jerrum and A. Sinclair, Approximating the permanent, SIAM J Comput 18 (1989), 1149–1178.

[22] M. Jerrum and A. Sinclair, The Markov chain Monte Carlo method: An approach to approximate counting and integration, In: D. Hochbaum, editor, Approximation algorithms for NP-hard problems, PWS Publishing Company, Boston, MA, 1997.

[23] M. Jerrum, A. Sinclair, and E. Vigoda, A polynomial-time approximation algorithms for permanent of a matrix with non-negative entries, J Assoc Comput Mach 51 (2004), 671–697.

[24] F. Kelly, Stochastic models of computer communication systems, J R Statist Soc B 47 (1985), 379–395.

[25] R. Kannan, L. Lovasz, and M. Simonovits, Random walks and $O^*(n^5)$ volume algorithm for convex bodies, Random Struct Algorithms 11 (1997), 1–50.

[26] M. Luby and E. Vigoda, Approximately counting up to four, Proc 29d Ann ACM Symposium on the Theory of Computing (STOC), El Paso, Texas, May 1997.

[27] L. Lovasz and S. Vempala, Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm, Proceedings of the 44th annual IEEE Symposium on Foundations of Computer Science, Cambridge, MA, October 2003, pp. 650–659.

[28] J. Martin, Reconstruction thresholds on regular trees, In: C. Banderier and C. Krattenthaler, editors, Discrete Random Walks, DRW'03, Discrete Mathematics and Theoretical Computer Science Proceedings AC, 2003, pp. 191–2003.

[29] M. Mezard and A. Montanari, Reconstruction on trees and spin glass transition, J Stat Phys 124 (2006).

[30] E. Mossel, Survey: information flow on trees, In J. Nestril and P. Winkler, editors, Graphs, morphisms and statistical physiscs, DIMACS series in discrete mathematics and theoretical computer science, American Mathematical Society, 2004, pp. 155–170.

[31] E. Mossel and Y. Peres, Information flow on trees, Ann Appl Probab 13 (2003), 817–844.

[32] M. Mezard and G. Parisi, The cavity method at zero temperature, http://fr.arxiv.org/ps/cond-mat/0207121, 2005.

[33] M. Mezard, G. Parisi, and M. A. Virasoro, Spin-glass theory and beyond, Vol. 9: Lecture Notes in Physics, World Scientific, Singapore, 1987.

[34] O. Rivoire, G. Biroli, O. C. Martin, and M. Mezard, Glass models on Bethe lattices, Eur Phys J B 37 (2004), 55–78.

[35] F. Spitzer, Markov random fields on an infinite tree, Ann Probab 3 (1975), 387–398.

[36] M. Talagrand, The high temperature case of the K-sat problem, Probab Theory Related Fields 119 (2001), 187–212.

[37] M. Talagrand, Parisi formula, Ann Math 163 (2006), 221–263.

[38] L. G. Valiant, The complexity of computing the permanent, Theor Comput Sci 8 (1979), 189–201.

[39] J. Warren, Dynamics and endogeny for recursive processes on trees, http://arxiv.org/abs/math.PR/0506038, 2005.

[40] D. Weitz, Combinatorial criteria for uniqueness of Gibbs measures, Random Struct Algorithms 27 (2005), 445–475.

[41] D. Weitz, Counting independent sets up to the tree threshold, In Proc 38th Ann. Symposium on the Theory of Computing, Seattle, WA, May 2006.

[42] S. Zachary, Countable state space Markov random fields and Markov chains on tree, Ann Prob 11 (1983), 894–903.