

Approximately Counting Independent Sets of a Given Size in Bounded-Degree Graphs

Ewan Davies   

Department of Computer Science, University of Colorado, Boulder, CO, USA

Will Perkins  

Department of Mathematics, Statistics, and Computer Science,
University of Illinois at Chicago, IL, USA

Abstract

We determine the computational complexity of approximately counting and sampling independent sets of a given size in bounded-degree graphs. That is, we identify a critical density $\alpha_c(\Delta)$ and provide (i) for $\alpha < \alpha_c(\Delta)$ randomized polynomial-time algorithms for approximately sampling and counting independent sets of given size at most αn in n -vertex graphs of maximum degree Δ ; and (ii) a proof that unless $\text{NP}=\text{RP}$, no such algorithms exist for $\alpha > \alpha_c(\Delta)$. The critical density is the occupancy fraction of hard core model on the clique $K_{\Delta+1}$ at the uniqueness threshold on the infinite Δ -regular tree, giving $\alpha_c(\Delta) \sim \frac{e}{1+e} \frac{1}{\Delta}$ as $\Delta \rightarrow \infty$.

2012 ACM Subject Classification Theory of computation \rightarrow Approximation algorithms analysis; Mathematics of computing \rightarrow Approximation algorithms; Theory of computation \rightarrow Algorithm design techniques; Theory of computation \rightarrow Random walks and Markov chains

Keywords and phrases approximate counting, independent sets, Markov chains

Digital Object Identifier 10.4230/LIPIcs.ICALP.2021.62

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: <https://arxiv.org/abs/2102.04984>

Funding *Will Perkins*: supported in part by NSF grants DMS-1847451 and CCF-1934915.

1 Introduction

Counting and sampling independent sets in graphs are fundamental computational problems arising in several fields including algorithms, statistical physics, and combinatorics. Given a graph G , let $\mathcal{I}(G)$ denote the set of independent sets of G . The independence polynomial of G is

$$Z_G(\lambda) = \sum_{I \in \mathcal{I}(G)} \lambda^{|I|} = \sum_{k \geq 0} i_k(G) \lambda^k,$$

where $i_k(G)$ is the number of independent sets of size k in G . The independence polynomial also arises as the partition function of the hard-core model from statistical physics.

With G and λ as inputs, exact computation of $Z_G(\lambda)$ is $\#\text{P}$ -hard [32, 18], but the complexity of approximating $Z_G(\lambda)$ has been a major topic in recent theoretical computer science research. There is a detailed understanding of the complexity of approximating $Z_G(\lambda)$ for the class of graphs of maximum degree Δ , in particular showing that there is a *computational threshold* which coincides with a certain probabilistic phase transition as one varies the value of λ .

The *hard-core model* on G at fugacity λ is the probability distribution on $\mathcal{I}(G)$ defined by

$$\mu_{G,\lambda}(I) = \frac{\lambda^{|I|}}{Z_G(\lambda)}.$$



© Ewan Davies and Will Perkins;
licensed under Creative Commons License CC-BY 4.0

48th International Colloquium on Automata, Languages, and Programming (ICALP 2021).

Editors: Nikhil Bansal, Emanuela Merelli, and James Worrell; Article No. 62; pp. 62:1–62:18

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



Defined on a lattice like \mathbb{Z}^d (through an appropriate limiting procedure), this is a simple model of a gas (the hard-core lattice gas) and it exhibits an order/disorder phase transition as λ changes. The hard-core model can also be defined on the infinite Δ -regular tree (the *Bethe lattice*). Kelly [23] determined the critical threshold for uniqueness of the infinite volume measure on the tree, namely

$$\lambda_c(\Delta) = \frac{(\Delta - 1)^{\Delta-1}}{(\Delta - 2)^\Delta}. \quad (1)$$

This value of λ also marks a computational threshold for the complexity of approximating $Z_G(\lambda)$ on graphs of maximum degree Δ . One can approximate $Z_G(\lambda)$ up to a relative error of ε in time polynomial in n and $1/\varepsilon$ with several different methods, provided G is of maximum degree Δ and $\lambda < \lambda_c(\Delta)$. The first such algorithm is based on correlation decay on trees and is due to Weitz [33], but recently alternative algorithms based on polynomial interpolation [3, 27, 28] and Markov chains [2, 6, 7] for this problem have also been given. Conversely, for $\lambda > \lambda_c(\Delta)$ a result of Sly and Sun [31] and Galanis, Štefankovič, and Vigoda [16] (following Sly [30]) states that unless NP=RP there is no polynomial-time algorithm for approximating $Z_G(\lambda)$ on graphs of maximum degree Δ . Counting and sampling are closely related, and by standard reduction techniques the same computational threshold holds for the problem of approximately sampling independent sets from the hard-core distribution.

The hard-core model is an example of the *grand canonical ensemble* from statistical physics, where one studies physical systems that can freely exchange particles and energy with a reservoir. Closely related is the *canonical ensemble*, where one removes the reservoir and considers a system with a fixed number of particles. In the context of independent sets in graphs, this corresponds to the uniform distribution on independent sets of some fixed size k . Here the number $i_k(G)$ of independent sets of size k in G plays the role of the partition function. In this paper we answer affirmatively the natural question of whether there is a similar complexity phase transition for the problem of approximating $i_k(G)$, and the related problem of sampling independent sets of size k approximately uniformly. Analogous to the critical fugacity in the hard-core model, we identify a critical density $\alpha_c(\Delta)$, and for $\alpha < \alpha_c(\Delta)$ we give a fully polynomial-time randomized approximation scheme (FPRAS, defined below) for counting independent sets of size k in n -vertex graphs of maximum degree Δ , where $0 \leq k \leq \alpha n$. We also show that unless NP=RP there is no such algorithm for $\alpha > \alpha_c(\Delta)$.

In statistical physics the grand canonical ensemble and the canonical ensemble are known to be equivalent in some respects under certain conditions, and the present authors, Jenssen, and Roberts [12] used this idea to give a tight upper bound on $i_k(G)$ for large k in large Δ -regular graphs G (see also [10] for the case of small k). Here, the main idea in our proofs is also to exploit the equivalence of ensembles. For algorithms at subcritical densities we approximately sample independent sets from the hard-core model and show that with sufficiently high probability we get an independent set of the desired size, distributed approximately uniformly. For hardness at supercritical densities we construct an auxiliary graph G' such that $i_k(G')$ is approximately proportional to $Z_G(\lambda)$ for some $\lambda > \lambda_c(\Delta)$, and hence is hard to approximate. Our counting and sampling algorithms for independent sets of size k permit higher densities than previous algorithms for this problem based on Markov chains [5, 1], and an algorithm implicit in [10] based on the cluster expansion.

A pleasant feature of our methods is the incorporation of several advances from recent research on related topics. From the geometry of polynomials we use a state-of-the-art zero-free region for $Z_G(\lambda)$ due to Peters and Regts [28] and a central limit theorem of

Michelen and Sahasrabudhe [26, 25] (though an older result of Lebowitz, Pittel, Ruelle and Speer [24] would also suffice), and we also apply the very recent development that a natural Markov chain (the Glauber dynamics) for sampling from the hard-core model mixes rapidly at fugacities $\lambda < \lambda_c(\Delta)$ on all graphs of maximum degree at most Δ [1, 7]. Finally, our results also show a connection between these algorithmic and complexity-theoretic problems and extremal combinatorics problems for bounded-degree graphs [9, 12, 10], see also the survey [34].

1.1 Preliminaries

Given an error parameter ε and real numbers z, \hat{z} , we say that \hat{z} is a *relative ε -approximation* to z if $e^{-\varepsilon} \leq \hat{z}/z \leq e^\varepsilon$. A *fully polynomial-time randomized approximation scheme* or *FPRAS* for a counting problem is a randomized algorithm that with probability at least $3/4$ outputs a relative ε -approximation to the solution of the problem in time polynomial in the size of the input and $1/\varepsilon$. If the algorithm is deterministic (i.e. succeeds with probability 1) then it is a *fully polynomial-time approximation scheme* (*FPTAS*). An ε -approximate sampling algorithm for a probability distribution μ outputs a random sample from a distribution $\hat{\mu}$ such that the total variation distance $\|\mu - \hat{\mu}\|_{TV} \leq \varepsilon$, and an *efficient sampling scheme* is, for all $\varepsilon > 0$ an ε -approximate sampling algorithm which runs in time polynomial in the size of the input and $\log(1/\varepsilon)$. Note that approximate sampling schemes whose running times are polynomial in $1/\varepsilon$ or in $\log(1/\varepsilon)$ are common in the literature, but we adopt the stronger definition for this paper. The inputs to our algorithms are graphs, and input size corresponds to the number of vertices of the graph.

An *independent set* in a graph $G = (V, E)$ is a subset $I \subset V$ such that no edge of E is contained in I . The *density* of such an independent set I is $|I|/|V|$, and it will be convenient for us to parametrize independent sets by their density instead of their size. We write $\mathcal{I}(G)$ for the set of all independent sets in G , $\mathcal{I}_k(G)$ for the set of independent sets of size k in G , and $i_k(G) = |\mathcal{I}_k(G)|$ for the number of such sets. Recall the hard-core distribution $\mu_{G,\lambda}$ on $\mathcal{I}(G)$ is given by $\mu_{G,\lambda}(I) = \lambda^{|I|}/Z_G(\lambda)$. We also define the *occupancy fraction* $\alpha_G(\lambda)$ of the hard-core model on G at fugacity λ to be the expected density of a random independent set drawn according to $\mu_{G,\lambda}$. Let \mathcal{G}_Δ be the set of graphs of maximum degree Δ .

The critical density that we show constitutes a computational threshold for the problems of counting and sampling independent sets of a given size in graphs of maximum degree Δ is

$$\alpha_c(\Delta) = \frac{\lambda_c(\Delta)}{1 + (\Delta + 1)\lambda_c(\Delta)} = \frac{(\Delta - 1)^{\Delta-1}}{(\Delta - 2)^\Delta + (\Delta + 1)(\Delta - 1)^{\Delta-1}},$$

with λ_c the critical fugacity as in (1). This may seem unexpected at first sight, but has a natural interpretation. The threshold is in fact the quantity $\alpha_{K_{\Delta+1}}(\lambda_c(\Delta))$, the occupancy fraction of the clique on $\Delta + 1$ vertices at the critical fugacity $\lambda_c(\Delta)$. This is a natural threshold because the occupancy fraction is a monotone increasing function of λ , and the clique on $\Delta + 1$ vertices has the minimum occupancy fraction over all graphs of maximum degree Δ . Thus, for any $G \in \mathcal{G}_\Delta$, the value of λ which makes $\alpha_G(\lambda) > \alpha_c(\Delta)$ must be greater than $\lambda_c(\Delta)$. Conversely, if $\alpha < \alpha_c(\Delta)$ then for every graph $G \in \mathcal{G}_\Delta$ there is some $\lambda < \lambda_c(\Delta)$ such that $\alpha_G(\lambda) = \alpha$.

1.2 Our results

We are now ready to state our main result.

► **Theorem 1.**

- (a) For every $\alpha < \alpha_c(\Delta)$ there is an FPRAS for $i_{\lfloor \alpha n \rfloor}(G)$ and an efficient sampling scheme for the uniform distribution on $\mathcal{I}_{\lfloor \alpha n \rfloor}(G)$ for n -vertex graphs G of maximum degree Δ .
- (b) Unless $\text{NP}=\text{RP}$, for every $\alpha \in (\alpha_c(\Delta), 1/2)$ there is no FPRAS for $i_{\lfloor \alpha n \rfloor}(G)$ for n -vertex, Δ -regular graphs G .

The assumption $\text{NP} \neq \text{RP}$, which is that polynomial-time algorithms using randomness cannot solve all problems in NP, is standard in computational complexity theory. Indeed, this assumption is used in [30, 31, 16] to show hardness of approximation for $Z_G(\lambda)$ on regular graphs at supercritical fugacities, which we apply directly. The upper bound of $1/2$ on α in b is required since in a regular graph (of degree ≥ 1) there are no independent sets of density greater than $1/2$ and counting those of density $1/2$ amounts to counting connected components in a bipartite graph (which can be done in polynomial time). For graphs of maximum degree Δ there is no such barrier, and in this case our methods can also be used to prove b for $\alpha \in (\alpha_c(\Delta), 1)$.

On the algorithmic side, Bubbly and Dyer [5] showed via path coupling that a natural Markov chain for sampling independent sets of size k in n -vertex graphs of maximum degree Δ mixes rapidly when $k < n/(2\Delta + 2)$, and recently this was slightly improved to $k < n/(2\Delta)$ via the method of high-dimensional expanders by Alev and Lau [1] (who also gave an improved bound in terms of the smallest eigenvalue of the adjacency matrix of G). The fast mixing of this Markov chain provides a randomized algorithm for approximate sampling and an FPRAS for approximate counting for this range of k . Implicit in the work of the present authors and Jenssen [10] is an alternative method based on the cluster expansion that yields an FPTAS for $i_k(G)$ when $k < e^{-5}n/(\Delta + 1)$, and although we did not try to optimize the constant it seems unlikely that without significant extension the cluster expansion approach could yield a sharp result. Considering asymptotics as $\Delta \rightarrow \infty$, these previous algorithms work for densities up to $(c + o(1))/\Delta$ with the constant c being $1/2$ or $e^{-5} \approx 0.007$ respectively. Here, our algorithms work for densities α satisfying

$$\alpha < \alpha_c(\Delta) = (1 + o(1)) \frac{e}{1 + e \Delta},$$

as $\Delta \rightarrow \infty$. The constant $e/(1 + e)$ is approximately 0.731, and our hardness proof shows that this is tight.

Our sampling algorithm is based on searching over possible values of λ until we find one for which the mean size of an independent set from the hard-core model is close to the target k . We then repeatedly sample from the hard-core model until we obtain an independent set of size k and output this independent set. Our approximate counting algorithm is based on a standard reduction of approximate counting to approximate sampling.

The method of sampling from the canonical ensemble by sampling from the grand canonical ensemble and conditioning on obtaining an object the desired size is quite old and appears in the seminal papers of Jerrum and Sinclair [20, 21]. The key technical step in applying this method is to prove a lower bound on the probability of obtaining the desired size; in [20, 21] this is accomplished by using log-concavity of the specific distribution on sizes. Since the hard-core model does not have this property in general, we need a different argument; our new argument is based on the rapid mixing of Glauber dynamics.

Harris and Kolmogorov [19] have recently investigated the general problem of estimating the coefficients of partition functions given access to samples from the corresponding Gibbs distribution. Applying their ideas and results to this problem could likely lead to more

efficient run times for the problem of approximating the entire sequence $\{i_k(G)\}$, $1 \leq k \leq \alpha n$. Their ideas could also be applied to improve the efficiency of the reduction from counting to sampling, as they describe a more efficient “cooling schedule” for the simulated annealing technique that we use for this type of reduction.

1.3 Triangle-free graphs

As an additional application of our techniques we find an approximate computational threshold for the class of triangle-free graphs.

► **Theorem 2.** *For every $\delta > 0$ there is Δ_0 large enough so that the following is true.*

- (a) *For $\Delta \geq \Delta_0$ and $\alpha < \frac{1-\delta}{\Delta}$ there is an FPRAS and efficient sampling scheme for $i_{\lfloor \alpha n \rfloor}(G)$ for the class of triangle-free graphs of maximum degree Δ .*
- (b) *For $\Delta \geq \Delta_0$ and $\alpha \in (\frac{1+\delta}{\Delta}, 1/2)$ there is no FPRAS for $i_{\lfloor \alpha n \rfloor}(G)$ for the class of triangle-free graphs of maximum degree Δ .*

The proof of this theorem uses a result on the occupancy fraction of triangle-free graphs from [11], see Section 4.

1.4 Related work

Counting independent sets of a specified size has arisen in various places as a natural fixed-parameter version of counting independent sets, and is equivalent to counting cliques of a specified size in the complement graph. Exact computation of $i_k(G)$ in an n -vertex graph H is trivially possible in time $O(k^2 n^k)$, though improvements can be made via fast matrix multiplication algorithms (see e.g. [15]). Another branch of research concerns the complexity (in both time and number of queries to the graph data structure) of counting and approximately counting cliques. For example, in [14] the authors gave a randomized approximation algorithm for approximating the number of cliques of size k . Results of this kind perform poorly in our setting, which is equivalent to counting cliques in the complement of bounded-degree graphs, because such graphs are very dense. In particular, the main result of [14] has expected running time $\Omega((nk/e)^k)$ in our setting.

With a focus on bounded-degree graphs and connections to statistical physics, our work is closer in spirit to that of Curticapean, Dell, Fomin, Goldberg, and Lapinskas [8]. There, the authors consider the problem of counting independent sets of size k in bipartite graphs from the perspective of parametrized complexity. They give algorithms for exact computation and approximation of $i_k(G)$ in bipartite graphs (of bounded degree and otherwise), including a fixed parameter tractable randomized approximation scheme, though their running times are exponential in k . We note that the complexity of approximately counting the total number of independent sets in bipartite graphs (a problem known as #BIS) is unknown [13].

1.5 Questions and future directions

For the hard-core model, the algorithm of Weitz [33] gives a deterministic approximation algorithm (FPTAS) for $Z_G(\lambda)$ for $\lambda < \lambda_c(\Delta)$. The approach of Barvinok along with results of Patel and Regts and Peters and Regts give another FPTAS for the same range of parameters [3, 27, 28]. Our algorithm for approximating the number of independent sets of a given size uses randomness, but we conjecture that there is a deterministic algorithm that works for the same range of parameters. (The cluster expansion approach of [10] gives an FPTAS but only for smaller values of α).

► **Conjecture 3.** *There is an FPTAS for $i_{\lfloor \alpha n \rfloor}(G)$ for $G \in \mathcal{G}_\Delta$ and all $\alpha < \alpha_c(\Delta)$.*

The Markov chain analyzed in [5, 1] is the “down/up” Markov chain: starting from an independent set $I_t \in \mathcal{I}_k(G)$ at step t , pick a uniformly random vertex $v \in I_t$ and a uniformly random vertex $w \in V$. Let $I' = (I_t \setminus v) \cup w$. If $I' \in \mathcal{I}_k(G)$, let $I_{t+1} = I'$; if not, let $I_{t+1} = I_t$.

► **Conjecture 4.** *The down/up Markov chain for sampling from $\mathcal{I}_{\lfloor \alpha n \rfloor}(G)$ mixes rapidly for $\alpha < \alpha_c(\Delta)$ and all $G \in \mathcal{G}_\Delta$.*

One of the steps of our proof leads to a natural probabilistic conjecture concerning the hard-core model in bounded degree graphs.

► **Conjecture 5.** *Suppose G is a graph on n vertices of maximum degree Δ . Then if $\lambda < \lambda_c(\Delta)$ and $k = \lfloor \mathbb{E}_{G,\lambda} |\mathbf{I}| \rfloor$, we have*

$$\mathbb{P}_{G,\lambda}[|\mathbf{I}| = k] = \Omega(n^{-1/2}),$$

where the implied constant only depends on Δ and λ and the expectation and probability are with respect to the hard-core model on G at fugacity λ .

Lemma 8 below gives the weaker bound $\Omega(n^{-1} \log^{-1} n)$. A stronger conjecture would be that a local central limit theorem for $|\mathbf{I}|$ holds whenever $\lambda < \lambda_c(\Delta)$.

Finally, our proofs of Theorems 1 and 2 show a close connection between the computational threshold for sampling independent sets of a given size in bounded-degree graphs and the extremal combinatorics problem of minimizing the occupancy fraction in the hard-core model over a class of bounded-degree graphs. We expect that a rigorous connection between the two problems can be proved.

2 Algorithms

In this section, we fix $\Delta \geq 3$ and $\alpha < \alpha_c(\Delta)$. We first give an algorithm that, for $G \in \mathcal{G}_\Delta$ on n vertices and $k \leq \alpha n$, returns an ε -approximate uniform sample from $\mathcal{I}_k(G)$ and runs in time polynomial in n and $\log(1/\varepsilon)$; this proves the sampling part of Theorem 1a. We then use this algorithm to approximate $i_k(G)$ using a standard simulated annealing process to prove the approximate counting part of Theorem 1a.

Given $\lambda \geq 0$, let \mathbf{I} be a random independent set from the hard-core model on G at fugacity λ . We will write $\mathbb{P}_{G,\lambda}$ for probabilities over the hard-core measure $\mu_{G,\lambda}$, so e.g. $\mathbb{P}_{G,\lambda}(|\mathbf{I}| = k)$ is the probability that \mathbf{I} is of size exactly k . Often we will suppress the dependence on G .

A key tool that we use for probabilistic analysis and to approximately sample from $\mu_{G,\lambda}$ is the *Glauber dynamics*. This is a Markov chain with state space $\mathcal{I}(G)$ and stationary distribution $\mu_{G,\lambda}$. Though the algorithm of Weitz [33] was the first to give an efficient approximate sampling algorithm for $\mu_{G,\lambda}$ for $\lambda < \lambda_c(\Delta)$ and all $G \in \mathcal{G}_\Delta$, a randomized algorithm with better running time now follows from recent results showing that the Glauber dynamics mix rapidly for this range of parameters [2, 6, 7]. The *mixing time* $T_{\text{mix}}(\mathcal{M}, \varepsilon)$ of a Markov chain \mathcal{M} is the number of steps from the worst-case initial state I_0 for the resulting state to have a distribution within total variation distance ε of the stationary distribution. We will use the following result of Chen, Liu, and Vigoda [7], and the sampling algorithm that it implies.

► **Theorem 6 ([7]).** *Given $\Delta \geq 3$ and $\xi \in (0, \lambda_c(\Delta))$, there exists $C > 0$ such that the following holds. For all $0 \leq \lambda < \lambda_c(\Delta) - \xi$ and graphs $G \in \mathcal{G}_\Delta$ on n vertices, the mixing time $T_{\text{mix}}(\mathcal{M}, \varepsilon)$ of the Glauber dynamics \mathcal{M} for the hard-core model on G with fugacity λ is at most $Cn \log(n/\varepsilon)$. This implies an ε -approximate sampling algorithm for $\mu_{G,\lambda}$ for $G \in \mathcal{G}_\Delta$ that runs in time $O(n \log n \log(n/\varepsilon))$.*

The sampling algorithm follows from the mixing time bound; the extra factor $\log n$ is the cost of implementing one step of Glauber dynamics (which requires reading $O(\log n)$ random bits to sample a vertex uniformly). Note that the implicit constant in the running time depends on how close λ is to $\lambda_c(\Delta)$, but in applications of this theorem we will have $\lambda \leq \lambda_c(\Delta) - \xi$ for some fixed $\xi > 0$, so that the implicit constant depends only on ξ , which in turn depends on α .

2.1 Approximate sampling

The algorithm *Sample- k* listed in Algorithm 1 uses Theorem 6 and a binary search on values of λ to generate samples from $\mathcal{I}_k(G)$. The algorithm requires access to distributions $\hat{\mu}_\lambda$ which are meant to be approximate versions of the hard-core model on G at fugacity λ . For the algorithm to work, we require that there is some value of λ such that sampling from the hard-core model at fugacity λ is likely to yield a set of size exactly k , and we will establish that this holds for some λ in a set of size $O(n^2)$ and for a suitable definition of “likely”. Quantitatively, these results inform the length $C \log n$ of the for loop and the value of N . An intuitive choice for λ would be the unique value that makes k the expected size of an independent set from the hard-core model on G , and in fact it suffices to find a value sufficiently close to this.

Algorithm 1 *Sample- k* .

input : $\alpha < \alpha_c$; $\varepsilon > 0$; $G \in \mathcal{G}_\Delta$ of size n ; integer $k \leq \alpha n$

output : $I \in \mathcal{I}_k(G)$ with distribution within ε total variation distance of the uniform distribution of $\mathcal{I}_k(G)$

- 1 Let $\lambda_* = \frac{\alpha}{1 - \alpha(\Delta + 1)}$
 - 2 For $t = 0, \dots, \lfloor 2\lambda_* n^2 \rfloor$, let $\lambda_t = t / (2n^2)$
 - 3 Let $\Lambda_0 = \{\lambda_t : t = 0, \dots, \lfloor 2\lambda_* n^2 \rfloor\}$
 - 4 **for** $i = 1, \dots, C \log n$, **do**
 - 5 Let λ be a median of the set Λ_{i-1}
 - 6 With $N = C' n^2 \log \left(\frac{\log n}{\varepsilon} \right)$, take N independent samples I_1, \dots, I_N from a distribution $\hat{\mu}_\lambda$ on $\mathcal{I}(G)$
 - 7 Let $\kappa = \frac{1}{N} \sum_{j=1}^N |I_j|$
 - 8 If $|\kappa - k| \leq 1/4$ and there exists $j \in \{1, \dots, N\}$ so that $|I_j| = k$, then output I_j for the smallest such j and **halt**
 - 9 If $\kappa \leq k$, let $\Lambda_i = \{\lambda' \in \Lambda_{i-1} : \lambda' > \lambda\}$. If instead $\kappa > k$, let $\Lambda_i = \{\lambda' \in \Lambda_{i-1} : \lambda' < \lambda\}$
 - 10 **end**
 - 11 If no independent set of size k has been obtained by the end of the for loop (or if $\Lambda_j = \emptyset$ at any step), use a greedy algorithm and output an arbitrary $I \in \mathcal{I}_k(G)$
-

► **Theorem 7.** *Let C be the constant in Line 4 and N be as in Line 6 of *Sample- k* (Algorithm 1). If the distributions $\hat{\mu}_\lambda$ are each within total variation distance $\varepsilon / (2CN \log n)$ of $\mu_{G,\lambda}$, the output distribution of *Sample- k* is within total variation distance ε of the uniform distribution of $\mathcal{I}_k(G)$. The running time of *Sample- k* is $O(N \log n \cdot T(n, \varepsilon))$ where $T(n, \varepsilon)$ is the running time required to produce a sample from $\hat{\mu}_\lambda$ satisfying the above guarantee.*

The sampling part of Theorem 1 follows immediately from Theorem 7 since by Theorem 6 we can obtain $\varepsilon/(2CN \log n)$ -approximate samples from $\mu_{G,\lambda}$ in time $O(n \log n \log(n \log n \cdot N/\varepsilon))$. Thus, the total running time of Sample- k with this guarantee on $\hat{\mu}_\lambda$ is

$$O(N \cdot n \log^2 n \cdot \log(nN/\varepsilon)) \leq n^3 \log^3 n \cdot \text{polylog}\left(\frac{\log n}{\varepsilon}\right).$$

Note that the use of Glauber dynamics in Sample- k could be replaced by any other polynomial-time approximate sampler for $\mu_{G,\lambda}$, such as the algorithm due to Weitz [33] based on the method of correlation decay. We do however use rapid mixing of the Glauber dynamics in Lemma 8 below to prove correctness of our algorithm, by establishing a lower bound on the probability that a set sampled from $\mu_{G,\lambda}$ has size close to its mean. Note that e.g. a resolution of our Conjecture 5 could imply the necessary probabilistic statement without appealing to the Glauber dynamics, however.

Before we prove Theorem 7, we collect a number of preliminary results that we will use. The first is a bound on the probability of getting an independent set of size close to the mean from the hard-core model when $\lambda < \lambda_c(\Delta)$. We use the notation $n\alpha_G(\lambda)$ for the expected size of an independent set from the hard-core model on G at fugacity λ to avoid ambiguities.

► **Lemma 8.** *For $\Delta \geq 3$ and $\alpha < \alpha_c(\Delta)$, there is a unique $\lambda_* < \lambda_c(\Delta)$ so that $\alpha_{K_{\Delta+1}}(\lambda_*) = \alpha$, and the following holds. For any $G \in \mathcal{G}_\Delta$ on n vertices and any $1 \leq k \leq \alpha n$, there exists an integer $t \in \{0, 1, \dots, \lfloor 2\lambda_* n^2 \rfloor\}$ so that*

$$|n\alpha_G(t/(2n^2)) - k| \leq 1/2. \quad (2)$$

Moreover, if t satisfies (2) then

$$\mu_{G,t/(2n^2)}(\mathcal{I}_k(G)) = \Omega\left(\frac{1}{n \log n}\right).$$

To prove this lemma we need several more results. The first is an extremal bound on $\alpha_G(\lambda)$ for $G \in \mathcal{G}_\Delta$. The statement of the theorem follows from a stronger property proved by Cutler and Radcliffe in [9]; see [12] for discussion.

► **Theorem 9 ([9]).** *For all $G \in \mathcal{G}_\Delta$ and all $\lambda \geq 0$,*

$$\alpha_G(\lambda) \geq \alpha_{K_{\Delta+1}}(\lambda) = \frac{\lambda}{1 + \lambda(\Delta + 1)}.$$

We next rely on a zero-free region for $Z_G(\lambda)$ due to Peters and Regts [28], so that we can apply the subsequent central limit theorem.

► **Theorem 10 ([28]).** *Let $\Delta \geq 3$ and $\xi \in (0, \lambda_c(\Delta))$. Then there exists $\delta > 0$ such that for every $G \in \mathcal{G}_\Delta$ the polynomial Z_G has no roots in the complex plane that lie within distance δ of the real interval $[0, \lambda_c(\Delta) - \xi]$.*

The *probability generating function* of a discrete random variable \mathbf{X} distributed on the non-negative integers is the polynomial in z given by $f(z) = \sum_{j \geq 0} \mathbb{P}(\mathbf{X} = j)z^j$, and the above result shows that at subcritical fugacity the probability generating function of $|\mathbf{I}|$ has no zeros close to 1 in \mathbb{C} . This lets us use the following result of Michelen and Sahasrabudhe [25].

► **Theorem 11 ([25]).** *For $n \geq 1$ let \mathbf{X}_n be a random variable taking values in $\{0, \dots, n\}$ with mean μ_n , standard deviation σ_n , and probability generating function f_n . If f_n has no roots within distance δ_n of 1 in \mathbb{C} , and $\sigma_n \delta_n / \log n \rightarrow \infty$, then $(X_n - \mu_n)/\sigma_n$ tends to a standard normal in distribution.*

The final tools we need are simple bounds on the variance of the size of an independent set from the hard-core model.

► **Lemma 12.** *Let G be a graph on n vertices and let \mathbf{I} be a random independent set drawn from the hard-core model on G at fugacity λ . Then, if the maximum degree of G is at most Δ and $M \geq n/(\Delta + 1)$ is the size of a largest independent set in G , we have*

$$\frac{\lambda}{(1 + \lambda)^{2+\Delta}} M \leq \text{var}(|\mathbf{I}|) \leq n^2 \frac{\lambda}{1 + \lambda}.$$

Proof. For the upper bound note that $|\mathbf{I}|$ is the sum of the indicator random variables \mathbf{X}_v that the vertex $v \in V(G)$ is in \mathbf{I} . Then because $\mathbb{P}(\mathbf{X}_v = 1) \leq \lambda/(1 + \lambda)$ for all v , from the Cauchy–Schwarz inequality in the form $\text{cov}(\mathbf{X}_u, \mathbf{X}_v)^2 \leq \text{var}(\mathbf{X}_u) \text{var}(\mathbf{X}_v)$ we obtain

$$\text{var}(|\mathbf{I}|) = \sum_{u \in V(G)} \sum_{v \in V(G)} \text{cov}(\mathbf{X}_u, \mathbf{X}_v) \leq n^2 \frac{\lambda}{1 + \lambda}.$$

For the lower bound, let J be some fixed independent set in G of maximum size M . Now write $\mathbf{X} = |\mathbf{I}|$, and let $\mathbf{K} = \mathbf{I} \setminus J$. By the law of total variance,

$$\text{var}(\mathbf{X}) = \mathbb{E}[\text{var}(\mathbf{X}|\mathbf{K})] + \text{var}(\mathbb{E}[\mathbf{X}|\mathbf{K}]) \geq \mathbb{E}[\text{var}(\mathbf{X}|\mathbf{K})].$$

But we have $\mathbf{X} = |\mathbf{K}| + |\mathbf{I} \cap J|$, and conditioned on \mathbf{K} the set $|\mathbf{I} \cap J|$ is distributed according to the hard-core model on $J \setminus N_G(\mathbf{K})$, the subset of J uncovered by \mathbf{K} . Since J is independent, this is a sum of at most $|J|$ independent, identically distributed Bernoulli random variables with probability $\lambda/(1 + \lambda)$.

Now, writing $\mathbf{U} = |J \setminus N_G(\mathbf{K})|$ for the number variables in the sum we have

$$\text{var}(\mathbf{X}) \geq \mathbb{E}[\text{var}(\mathbf{X}|\mathbf{K})] = \frac{\lambda}{(1 + \lambda)^2} \mathbb{E}\mathbf{U}.$$

A vertex $u \in J$ is uncovered by \mathbf{K} precisely when $N(u) \cap \mathbf{K} = \emptyset$. Then by successive conditioning and the maximum degree condition, the probability that u is uncovered by \mathbf{K} is at least $(1 + \lambda)^{-\Delta}$. This means $\mathbb{E}\mathbf{U} \geq |J|(1 + \lambda)^{-\Delta}$ and hence

$$\text{var}(\mathbf{X}) \geq \frac{\lambda}{(1 + \lambda)^{2+\Delta}} M.$$

The assertion $M \geq n/(\Delta + 1)$ follows from the fact that any n -vertex graph of maximum degree Δ contains an independent set of size at least $n/(\Delta + 1)$, which is easy to prove by analyzing a greedy algorithm. ◀

Now we are ready to prove Lemma 8.

Proof of Lemma 8. A standard calculation gives

$$\frac{\partial}{\partial \lambda} \alpha_G(\lambda) = \frac{1}{n} \frac{\partial}{\partial \lambda} \frac{\lambda Z'_G(\lambda)}{Z_G(\lambda)} = \frac{1}{n\lambda} \text{var}(|\mathbf{I}|),$$

and so Lemma 12 gives that $0 < \alpha'_G(\lambda) \leq n$ for all $\lambda > 0$.

Next, let $\lambda_* < \lambda_c(\Delta)$ be the solution to the equation $\alpha_{K_{\Delta+1}}(\lambda_*) = \alpha$. This means

$$\lambda_* = \frac{\alpha}{1 - \alpha(\Delta + 1)},$$

62:10 Approximately Counting Independent Sets of a Given Size

as defined in Sample- k . The fact that $\lambda_* < \lambda_c(\Delta)$ follows from the fact that $\alpha < \alpha_c(\Delta) = \alpha_{K_{\Delta+1}}(\lambda_c(\Delta))$, and that occupancy fractions are strictly increasing. Then using Theorem 9 we have that

$$\alpha_G(\lambda_*) \geq \alpha_{K_{\Delta+1}}(\lambda_*) = \alpha, \quad (3)$$

and so there exists $\lambda \in (0, \lambda_*]$ such that $n\alpha_G(\lambda) = k$. Using the upper bound on $\alpha'_G(\lambda)$, we see that as λ increases over an interval of length $1/(2n^2)$, the function $n\alpha_G(\lambda)$ can increase by at most $1/2$. Hence, there is at least one integer $t \in \{1, \dots, \lfloor 2\lambda_* n^2 \rfloor\}$ such that $|n\alpha_G(t/(2n^2)) - k| \leq 1/2$.

The second statement of Lemma 8 follows from a central limit theorem for $|\mathbf{I}|$ and rapid mixing of the Glauber dynamics. There is a close connection between zeros of the probability generating function of $|\mathbf{I}|$ and the zeros of the partition function itself. The probability generating function of $|\mathbf{I}|$ is

$$f(z) = \sum_{j \geq 0} \mathbb{P}_\lambda(|\mathbf{I}| = j) z^j = \sum_{j \geq 0} \frac{i_j(G) \lambda^j z^j}{Z_G(\lambda)} = \frac{Z_G(\lambda z)}{Z_G(\lambda)}.$$

Then for λ such that $Z_G(\lambda) \neq 0$, z is a root of f if and only if $z\lambda$ is a root of $Z_G(\lambda)$. By our assumptions on t , when $\lambda = t/(2n^2)$ Theorem 10 gives the existence of $\delta > 0$ such that for all $G \in \mathcal{G}_\Delta$ there are no complex zeros of f within distance δ/λ of 1. This is because Theorem 10 means that $Z_G(z\lambda) = 0$ implies $|z\lambda - \lambda| \geq \delta$. The condition of Theorem 11 which states that $\sigma_n \delta_n / \log n \rightarrow \infty$ is met because $\lambda < \lambda_c(\Delta) \leq 4$ and so

$$\sigma_n \delta_n \geq \sqrt{\frac{\lambda}{(1+\lambda)^{2+\Delta}} \frac{n}{\Delta+1}} \cdot \frac{\delta}{\lambda} \geq \Omega\left(\sqrt{n/\lambda}\right) > \omega(\log n).$$

Now, let $\lambda = t/(2n^2)$ and suppose that (2) holds, meaning that k is within $1/2$ of $n\alpha_G(\lambda)$. The standard deviation of the size of a set drawn from $\mu_{G,\lambda}$ is at least a constant, which follows from the lower bound in Lemma 12 of $\Omega(\sqrt{\lambda n})$ and the fact that $\lambda \geq \Omega(1/n)$. To see this, note that $\alpha_G(\lambda) \leq \lambda/(1+\lambda)$ for any graph and non-negative fugacity λ . This holds because when \mathbf{I} is a random independent set from the hard-core model, conditioned on a vertex v having no neighbors in \mathbf{I} , $v \in \mathbf{I}$ with probability $\lambda/(1+\lambda)$. If v has a neighbor in \mathbf{I} then $v \notin \mathbf{I}$ with probability 1, and the bound follows. Then using (2), we have

$$n \frac{\lambda}{1+\lambda} \geq n\alpha_G(\lambda) \geq k - 1/2 \geq 1/2,$$

and so $\lambda \geq \Omega(1/n)$. We deduce that k is within some constant number $r > 0$ of standard deviations of the mean size $n\alpha_G(\lambda)$. The central limit theorem and standard properties of the normal distribution mean that there are constants $\rho > 0$ (small enough as a function of r) and n_0 such that for all $n \geq n_0$, with probability at least ρ , $|\mathbf{I}|$ is at least r standard deviations below the mean, and similarly with probability at least ρ it is at least r standard deviations above the mean. So we have $\mathbb{P}_{G,\lambda}(|\mathbf{I}| \geq k) \geq \rho$ and $\mathbb{P}_{G,\lambda}(|\mathbf{I}| \leq k) \geq \rho$.

The transition probabilities when we are at state I in the Glauber dynamics are given by the following random experiment. Choose a vertex $v \in V(G)$ uniformly at random and let

$$I' = \begin{cases} I \cup \{v\} & \text{with probability } \lambda/(1+\lambda), \\ I \setminus \{v\} & \text{with probability } 1/(1+\lambda). \end{cases}$$

Now if I' is independent in G move to state I' , otherwise stay in state I . This means that the sequence of sizes of set visited must take consecutive integer values. By Theorem 6, there is a constant C'' such that from an arbitrary starting state, in $C''n \log n$ steps the distribution π of the current state is within total variation distance $\rho/2$ of the hard-core model. Then the following statements hold.

- (i) Starting from an independent set of size at most k , with probability at least $\rho/2$ the state after $C''n \log n$ steps is an independent set of size at least k .
- (ii) Starting from an independent set of size at least k , with probability at least $\rho/2$ the state after $C''n \log n$ steps is an independent set of size at most k .

Consider starting from an initial state distributed according to $\mu_{G,\lambda}$. Then every subsequent state is also distributed according to $\mu_{G,\lambda}$, and the above facts mean that for any sequence of $C''n \log n$ consecutive steps, with probability at least $\rho/2$ we see a state of size exactly k . Recalling that $\lambda = t/(2n^2)$, this immediately implies that

$$\mu_{G,t/(2n^2)}(\mathcal{I}_k(G)) \geq \frac{\rho}{2C''n \log n},$$

as required. ◀

Proof of Theorem 7. We first prove the theorem under the assumption that each $\hat{\mu}_\lambda$ is exactly the hard-core measure $\mu_{G,\lambda}$, taking note of how many times we sample from any $\hat{\mu}_\lambda$.

We say a *failure* occurs at step i in the FOR loop if either of the following occur:

1. $|n\alpha_G(\lambda) - \kappa| > 1/4$.
2. $|n\alpha_G(\lambda) - k| \leq 1/2$ but the algorithm did not output an independent set of size k in step i .

We show that the probability that a failure occurs at any time during the algorithm is at most $\varepsilon/2$. By a union bound, it is enough to show that the probability of either type of failure at a given step i is at most $\frac{\varepsilon}{4C \log n}$.

Consider an arbitrary step i with its value of λ . To bound the quantity $\mathbb{P}(|n\alpha_G(\lambda) - \kappa| > 1/4)$, note that κ is the mean of N independent samples from $\hat{\mu}_\lambda$, which we currently assume to be $\mu_{G,\lambda}$. Then we have $\mathbb{E}\kappa = n\alpha_G(\lambda)$ and Hoeffding's inequality gives $\mathbb{P}(|n\alpha_G(\lambda) - \kappa| > 1/4) \leq 2e^{-N/(8n^2)}$, so for this to be at most $\varepsilon/(4C \log n)$ we need only $N \geq \Omega(n^2 \log(\frac{\log n}{\varepsilon}))$.

To bound the probability that the current step involves λ such that $|n\alpha_G(\lambda) - k| \leq 1/2$, but we fail to get a set of size k in the N samples, observe that we have N independent trials for getting a set of size k , and each trial succeeds with probability $p \geq c/(n \log n)$ by Lemma 8. Then the probability we see no successful trials is $(1 - \frac{c}{n \log n})^N$, which is at most $\varepsilon/(4C \log n)$ for $N \geq \Omega(n \log n \cdot \log(\frac{\log n}{\varepsilon}))$. Thus, we can take $N = \Theta(n^2 \log(\frac{\log n}{\varepsilon}))$, as in line 6 of Sample- k .

Next we show that in the event that no failure occurs during the running of the algorithm, the algorithm outputs an independent set I with distribution within $\varepsilon/2$ total variation distance of the uniform distribution on $\mathcal{I}_k(G)$.

We first observe that if no failure occurs, the algorithm at some point reaches a value of λ so that $|n\alpha_G(\lambda) - k| \leq 1/2$. This is a simple consequence of Lemma 8, which means there exists some t with this property, and the binary search structure of the algorithm. In particular, in each iteration of the FOR loop, at line (e) the size of the set Λ_i being searched goes down by (at least) half. Conditioned on no failures, the search also proceeds in the correct half of λ_i because we search the upper half only when $\kappa < k - 1/4$ and so conditioned on no failure we have $n\alpha_G(\lambda) \leq \kappa + 1/4 < k$ and hence using a larger value of λ must bring $n\alpha_G(\lambda)$ closer to k . The case $\kappa > k + 1/4$ is similar. This means that, conditioned on no failures, the algorithm must reach a value of λ such that $|n\alpha_G(\lambda) - k| \leq 1/4$.

62:12 Approximately Counting Independent Sets of a Given Size

Note that $\mu_{G,\lambda}$ conditioned on getting a set of size exactly k is precisely the uniform distribution on $\mathcal{I}_k(G)$, hence if the algorithm outputs an independent set of size k during the FOR loop, its distribution is exactly uniform distribution on $\mathcal{I}_k(G)$. Thus, under the assumption that each $\hat{\mu}_\lambda$ is precisely $\mu_{G,\lambda}$ we have shown that with probability at least $1 - \varepsilon/2$ no failures occur, and hence a perfectly uniform sample from $\mathcal{I}_k(G)$ is output during the FOR loop.

We do not have access to an efficient exact sampler for $\mu_{G,\lambda}$, so we make do with the approximate sampler from Theorem 6. One interpretation of total variation distance is that when each $\hat{\mu}_\lambda$ has total variation distance at most ξ from $\mu_{G,\lambda}$, there is a coupling between $\hat{\mu}_\lambda$ and $\mu_{G,\lambda}$ such that the probability they disagree is at most ξ . Then to prove Theorem 7 we consider a third failure condition: that during any of the calls to a sampling algorithm for any $\hat{\mu}_\lambda$ the output differs from what would have been given by $\mu_{G,\lambda}$ under this coupling. Since we make at most $CN \log n$ calls to such sampling algorithms, provided $\xi \leq \varepsilon/(2CN \log n)$ the probability of any failure of this kind is at most $\varepsilon/2$. Together with the above proof for samples distributed exactly according to $\mu_{G,\lambda}$ which successfully returns uniform samples from $\mathcal{I}_k(G)$ with probability $1 - \varepsilon/2$, we have now shown the existence of a sampler that with probability $1 - \varepsilon$ returns uniform samples from $\mathcal{I}_k(G)$, and makes at most $CN \log n$ calls to a $\varepsilon/(2CN \log n)$ -approximate sampler for $\mu_{G,\lambda}$ (at various values of λ). Interpreting this in terms of total variation distance, this means we have an ε -approximate sampler for the uniform distribution on $\mathcal{I}_k(G)$ with running time $O(N \log n \cdot T(n, \varepsilon))$. ◀

2.2 Approximate counting

Given a graph $G = (V, E)$ on n vertices and $j \geq 0$, let $f_j(G) = (j+1)i_{j+1}(G)/i_j(G)$. This $f_j(G)$ has an interpretation as the *expected free volume* over a uniform random independent set $\mathbf{J} \in \mathcal{I}_j(G)$, that is, $f_j = \mathbb{E}|V \setminus (\mathbf{J} \cup N(\mathbf{J}))|$. This holds because each vertex in $V \setminus (\mathbf{J} \cup N(\mathbf{J}))$ can be added to \mathbf{J} to make an independent set of size $j+1$, and each such set is counted $j+1$ times in this way. Then by a simple telescoping product we have

$$i_k(G) = \prod_{j=0}^{k-1} \frac{f_j(G)}{j+1}, \quad (4)$$

and hence if for $0 \leq j \leq k-1$ we can obtain a relative ε/k approximation to f_j in time polynomial in n and $1/\varepsilon$ then we can obtain a relative ε -approximation to $i_k(G)$ in time polynomial in n and $1/\varepsilon$. By the definition of f_j as an expectation over a uniform random independent set of size j , we can use an efficient sampling scheme for this distribution to approximate f_j , which is provided by Theorem 7. That is, by repeatedly sampling independent sets of size j approximately uniformly and recording the free volume we can approximate the expected free volume $f_j(G)$, and hence the corresponding term of the product in (4). Doing this for all $0 \leq j \leq k-1$ thus provides an approximation to $i_k(G)$. This scheme is an example of *simulated annealing*, which can be used as a general technique for obtaining approximation algorithms from approximate sampling algorithms. For more details, see e.g. [22, 29]. Here the integer j is playing the role of inverse temperature, and we approximate $i_k(G)$ by estimating $f_j(G)$ (by sampling from $\mathcal{I}_j(G)$) with the cooling schedule $j = 0, 1, \dots, k-1$. We expect that a more sophisticated cooling schedule can be used to decrease the running time of our reduction, see for example [19].

Since this annealing process is standard, we sketch a simple version of the method. Suppose that for all $0 \leq j \leq k-1$ we have a randomized algorithm that with probability at least $1 - \delta'$ returns a relative ε/k -approximation \hat{t}_j to $f_j(G)/(j+1)$ in time T' . Then (4)

implies that with probability at least $1 - k\delta$, the product $\hat{i}_k = \prod_{j=0}^{k-1} \hat{t}_j$ is a relative ε -approximation to $i_k(G)$, and this takes time kT' to compute. For the FPRAS in Theorem 1, it therefore suffices to design the hypothetical algorithm with $\delta' = 1/(4k)$ and T' polynomial in n and $1/\varepsilon$.

First, suppose that we have access to an exactly uniform sampler for $\mathcal{I}_j(G)$ for $0 \leq j \leq k-1$, but impose the smaller failure probability bound of $\delta'/2$. Then, for each j , let \hat{t}_j be the sample mean of m computations of $|V \setminus (\mathbf{J} \cup N(\mathbf{J}))|/(j+1)$ where \mathbf{J} is a uniform random independent set of size j . We note that as a random variable $|V \setminus (\mathbf{J} \cup N(\mathbf{J}))|/(j+1)$ has a range of at most $j\Delta/(j+1)$ in a graph of maximum degree Δ because $0 \leq |N(\mathbf{J})| \leq j\Delta$, and for $j \leq k-1$ and

$$k \leq \alpha n < \alpha_c(\Delta)n < \frac{e}{1+e} \frac{n}{\Delta},$$

we have

$$\frac{|V \setminus (\mathbf{J} \cup N(\mathbf{J}))|}{j+1} \geq \frac{n - j(\Delta + 1)}{j+1} \geq \frac{\Delta}{e} - 1.$$

Let S_j be the mean of m samples of $|V \setminus (\mathbf{J} \cup N(\mathbf{J}))|/(j+1)$. Then, using that for $\varepsilon' \leq 1$ it suffices to ensure $|S_j - \mu| \leq \varepsilon'\mu/2$ for S_j to be a relative ε' -approximation to μ , by Hoeffding's inequality,

$$m \geq \Omega(\varepsilon^{-2}k^2 \log(1/\delta')) = \Omega(\varepsilon^{-2}k^2 \log k)$$

samples are sufficient to obtain the required approximation accuracy ε' with the required success probability $1 - \delta'/2$. Since we do not have an exact sampler, we use the approximate sampler obtained in this section with total variation distance $\delta'/2$. Using the coupling between the exact and the approximate sampler that we used in the proof of Theorem 7, this suffices to obtain the required sampling accuracy with failure probability at most δ' . Recalling that $k \leq n$, it is now simple to check that the running time of the entire annealing scheme is polynomial in n and $1/\varepsilon$. This completes the proof of Theorem 1a.

3 Hardness

To prove hardness we will use the notion of an ‘‘approximation-preserving reduction’’ from [13]. We reduce the problem of approximating the hard-core partition function $Z_G(\lambda)$ on a Δ -regular graph G , which we recall is hard for $\lambda > \lambda_c$ (see [16, 31]), to the problem of approximating $i_k(G')$ for Δ -regular graph G' that can be constructed in time polynomial in the size of G . In particular, we show that it suffices to find an $\varepsilon/2$ -approximation to $i_k(G')$ in order to obtain an ε -approximation to $Z_G(\lambda)$.

Let $\text{IS}(\alpha, \Delta)$ be the problem of computing $i_{\lfloor \alpha n \rfloor}(G)$ for a Δ -regular graph G on n vertices. Let $\text{HC}(\lambda, \Delta)$ be the problem of computing $Z_G(\lambda)$ for a Δ -regular graph G .

► **Theorem 13.** *For every $\Delta \geq 3$ and $\alpha \in (\alpha_c(\Delta), 1/2)$, there exists $\lambda > \lambda_c(\Delta)$ so that there is an approximation-preserving reduction from $\text{HC}(\lambda, \Delta)$ to $\text{IS}(\alpha, \Delta)$.*

Theorem 13 immediately implies the hardness part of Theorem 1 as the results of [16, 31] show that there is no FPRAS for $\text{HC}(\lambda, \Delta)$ for any $\lambda > \lambda_c(\Delta)$ unless $\text{NP}=\text{RP}$.

Proof of Theorem 13. Fix $\Delta \geq 3$, and let $\alpha \in (\alpha_c(\Delta), 1/2)$ be given. We will construct a Δ -regular graph H on n_H vertices such that for some value $\lambda \in (\lambda_c(\Delta), \infty)$ we have $\alpha_H(\lambda) = \alpha$. Our reduction is then as follows: given a Δ -regular graph G on n vertices and

62:14 Approximately Counting Independent Sets of a Given Size

$\varepsilon > 0$, let G' be the disjoint union of G with rH , the graph of r disjoint copies of H , with $r = \lceil C\Delta n^2/\varepsilon \rceil$ for some absolute constant C . Let $N = |V(G')| = n + rn_H$. We will prove that

$$e^{-\varepsilon/2} \frac{i_k(G')}{i_k(rH)} \leq Z_G(\lambda) \leq e^{\varepsilon/2} \frac{i_k(G')}{i_k(rH)}, \quad (5)$$

where $k = \lfloor \alpha N \rfloor$.

Since G' can be constructed and $i_k(rH)$ computed in time polynomial in n , this provides the desired approximation-preserving reduction. What remains is to construct the graph H satisfying $\alpha_H(\lambda) = \alpha$ and then to prove (5).

Constructing H

The graph $H = H_{a,b}$ will consist of the union of a copies of the complete bipartite graph $K_{\Delta,\Delta}$ and b copies of the clique $K_{\Delta+1}$. Clearly H is Δ -regular. Since the occupancy fraction of any graph is a strictly increasing function of λ , and the relevant occupancy fractions satisfy $\alpha_{K_{\Delta+1}}(\lambda_c(\Delta)) = \alpha_c(\Delta)$ and $\lim_{\lambda \rightarrow \infty} \alpha_{K_{\Delta,\Delta}}(\lambda) = 1/2$, we see that there exist integers $a, b \geq 0$ (with at least one positive) and $\lambda > \lambda_c(\Delta)$ so that $\alpha_{H_{a,b}}(\lambda) = \alpha$. A given pair (a, b) provides a suitable $H_{a,b}$ when

$$\alpha_{H_{a,b}}(\lambda_c(\Delta)) < \alpha < \lim_{\lambda \rightarrow \infty} \alpha_{H_{a,b}}(\lambda) = \frac{a\Delta + b}{2a\Delta + b(\Delta + 1)},$$

and hence it can be shown that for all $\Delta \geq 3$ one of the pairs $(0, 1)$, $(1, 16)$, $(1, 6)$, $(1, 3)$, $(2, 3)$, $(2, 1)$, $(1, 0)$ suffices for (a, b) , and a suitable pair is easy to find efficiently. This provides us with the desired graph H . From here on, fix these values a, b, λ and let $n_H = 2a\Delta + b(\Delta + 1)$.

Proving (5)

We now form G' by taking the union of G (a Δ -regular graph on n vertices) and r copies of H . Let $N = n + rn_H$ be the number of vertices of G' , and write $k = \lfloor \alpha N \rfloor$. Let rH be the graph consisting of the disjoint union of r copies of H . We can write:

$$i_k(G') = \sum_{j=0}^n i_j(G) i_{k-j}(rH) = i_k(rH) \sum_{j=0}^n i_j(G) \frac{i_{k-j}(rH)}{i_k(rH)}.$$

Now to prove (5) it suffices to show that for $r \geq C\Delta n^2/\varepsilon$ and $0 \leq j \leq n$, we have

$$e^{-\varepsilon/2} \lambda^j \leq \frac{i_{k-j}(rH)}{i_k(rH)} \leq e^{\varepsilon/2} \lambda^j. \quad (6)$$

We have the exact formula (for any $0 \leq j \leq k$)

$$i_{k-j}(rH) = \frac{Z_{rH}(\lambda)}{\lambda^{k-j}} \mathbb{P}_{rH,\lambda}(|\mathbf{I}| = k-j)$$

and so

$$\frac{i_{k-j}(rH)}{i_k(rH)} = \lambda^j \frac{\mathbb{P}_{rH,\lambda}(|\mathbf{I}| = k-j)}{\mathbb{P}_{rH,\lambda}(|\mathbf{I}| = k)},$$

where $\mathbb{P}_{rH,\lambda}$ denotes probabilities with respect to an independent set \mathbf{I} drawn according to the hard-core model on rH at fugacity λ . It is then enough to show

$$e^{-\varepsilon/2} \leq \frac{\mathbb{P}_{rH,\lambda}(|\mathbf{I}| = k-j)}{\mathbb{P}_{rH,\lambda}(|\mathbf{I}| = k)} \leq e^{\varepsilon/2}.$$

This will follow from a local central limit theorem (e.g. [17]) since $|\mathbf{I}|$ is the sum of r i.i.d. random variables and the fact that $\mathbb{E}_{r,H,\lambda}|\mathbf{I}|$ is close to both k and $k - j$. The following theorem gives us what we need.

► **Theorem 14** (Gnedenko [17]). *Let X_1, \dots, X_r be i.i.d. integer valued random variables with mean μ and variance σ^2 , and suppose that the support of X_1 includes two consecutive integers. Let $S_r = X_1 + \dots + X_r$. Then*

$$\mathbb{P}(S_r = k) = \frac{1}{\sqrt{2\pi r\sigma}} \exp\left[-(k - r\mu)^2/(2r\sigma^2)\right] + o(r^{-1/2}),$$

with the error term $o(r^{-1/2})$ uniform in k .

This immediately implies that with μ and σ^2 the mean and standard deviation of the hard-core model on H at fugacity λ ,

$$\frac{\mathbb{P}_{r,H,\lambda}(|\mathbf{I}| = k - j)}{\mathbb{P}_{r,H,\lambda}(|\mathbf{I}| = k)} = \frac{e^{-[j^2 - 2(k - r\mu)j]/(2r\sigma^2)} + o(e^{(k - r\mu)^2/(2r\sigma^2)}/r)}{1 + o(e^{(k - r\mu)^2/(2r\sigma^2)}/r)}.$$

It therefore suffices to show that for large enough r , namely $r \geq C\Delta n^2/\varepsilon$, we can make $[j^2 - 2(k - r\mu)j]/(2r\sigma^2)$ small compared to ε and show that $(k - r\mu)^2/(2r\sigma^2)$ is bounded above by some absolute constant. Note that $\mu = \alpha n_H$, and by Lemma 12 we have for all $\Delta \geq 3$ (and any choices of α, λ, a, b made according to our conditions),

$$\sigma^2 \geq \frac{\lambda}{(1 + \lambda)^{2+\Delta}}(a\Delta + b) \geq \frac{\lambda_c(\Delta)}{(1 + \lambda_c(\Delta))^{2+\Delta}}(a\Delta + b) \geq \frac{0.00384}{\Delta}.$$

Since $k = \lfloor \alpha N \rfloor = \lfloor \alpha n + r\alpha n_H \rfloor$, we then have $(k - r\mu)^2 \leq \alpha^2 n^2 < n^2$, and hence $\frac{(k - r\mu)^2}{2r\sigma^2} \leq C'\Delta \frac{n^2}{r}$, where C' is an absolute constant. Now since $0 \leq j \leq n$ we also have

$$\left| \frac{j^2 - 2(k - r\mu)j}{2r\sigma^2} \right| \leq C'\Delta \frac{n^2}{r}.$$

This means that provided we take C to be a large enough absolute constant and $r \geq C\Delta n^2/\varepsilon$, we have (5) as required. ◀

4 Triangle-free graphs

In this section we briefly describe the modifications to the proofs in Sections 2 and 3 to yield Theorem 2, an analogue of Theorem 1 for triangle-free graphs.

4.1 Algorithms

We use the following lower bound on the occupancy fraction of triangle-free graphs.

► **Theorem 15** ([11]). *For every $\delta > 0$, there is Δ_0 large enough so that for every $\Delta \geq \Delta_0$, and every triangle-free $G \in \mathcal{G}_\Delta$,*

$$\alpha_G(\lambda_c(\Delta) - 1/\Delta^2) \geq \frac{1 - \delta}{\Delta}.$$

This statement follows from [11, Theorem 3] and some asymptotic analysis of the bound for $\lambda = \lambda_c(\Delta) - 1/\Delta^2$ as $\Delta \rightarrow \infty$. Now the algorithm for Theorem 2 is essentially the same as for Theorem 1, but since we assume the graph G is triangle free we can use a stronger

lower bound on the occupancy fraction than Theorem 9. Let $\delta > 0$ and $\alpha < (1 - \delta)/\Delta$ as in Theorem 2. Then Theorem 15 means that for $\Delta \geq \Delta_0$ and any triangle-free graph $G \in \mathcal{G}_\Delta$ we have

$$\alpha_G(\lambda_c(\Delta) - 1/\Delta^2) \geq \frac{1 - \delta}{\Delta} > \alpha.$$

But occupancy fractions are continuous and strictly increasing, so with $\lambda_* = \lambda_c(\Delta) - 1/\Delta^2$ there exists $\lambda \in (0, \lambda_*]$ such that $k = n\alpha_G(\lambda)$, as in the proof of Lemma 8 but permitting larger α . The analysis of the algorithm can then proceed exactly as in the proofs of Lemma 8 and Theorem 1, completing the proof of Theorem 2a.

4.2 Hardness

The proof of hardness for triangle-free graphs is the same as that for general graphs, but we replace $K_{\Delta+1}$ with a (constant-sized) random regular graph in the construction. Bhatnagar, Sly, and Tetali [4] showed that the local distribution of the hard-core model on the random regular graph converges to that of the unique translation-invariant hard-core measure on the infinite regular tree for a range of λ including $\lambda = \lambda_c(\Delta)$. This means that if K is a random Δ -regular graph on n vertices and α_{T_Δ} denotes the occupancy fraction of the unique translation-invariant hard-core measure on the infinite Δ -regular tree (see [4, 11]) we have with probability $1 - o_n(1)$,

$$\alpha_K(\lambda_c(\Delta)) = \alpha_{T_\Delta}(\lambda_c(\Delta)) + o_n(1) = \frac{1 + o_{n,\Delta}(1)}{\Delta},$$

where $o_n(1) \rightarrow 0$ as $n \rightarrow \infty$ and $o_{n,\Delta}(1) \rightarrow 0$ as both n and Δ tend to infinity. Thus, for fixed $\delta \in (0, 1)$, there is $n_0 = n_0(\delta)$ and $\Delta_0 = \Delta_0(\delta)$ such that with probability at least $1 - \delta$ a random Δ -regular graph K on n_0 vertices has $\alpha_G(\lambda_c(\Delta)) \leq (1 + \delta)/\Delta$. This means that in time bounded by a function of δ an exhaustive search over Δ -regular graphs on n_0 vertices must yield a graph K with the property $\alpha_K(\lambda_c(\Delta)) \leq (1 + \delta)/\Delta$. Now we replace $K_{\Delta+1}$ with the random Δ -regular graph K in the proof above, which for $\Delta \geq \Delta_0$ allows us to work with any $\alpha \in ((1 + \delta)/\Delta, 1/2)$ by the above argument. To finish the proof, we require that approximating $Z_G(\lambda)$ is hard for Δ -regular *triangle-free* graphs G when $\lambda > \lambda_c$. This follows directly from the proof of Sly and Sun [31], as their gadget which shows hardness for Δ -regular graphs contains no triangles. Thus, we have the following analogue of Theorem 13, where we let $\text{IS}'(\alpha, \Delta)$ be the problem of computing $i_{\lfloor \alpha n \rfloor}(G)$ for a Δ -regular triangle-free graph G on n vertices.

► **Theorem 16.** *Given $\delta > 0$ there exists Δ_0 such that the following holds for all $\Delta \geq \Delta_0$. For every $\alpha \in ((1 + \delta)/\Delta, 1/2)$, there exists $\lambda > \lambda_c(\Delta)$ so that there is an approximation-preserving reduction from $\text{HC}(\lambda, \Delta)$ to $\text{IS}'(\alpha, \Delta)$.*

This implies Theorem 2b.

References

- 1 Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1198–1211, Chicago IL USA, 2020. ACM. doi:10.1145/3357713.3384317.
- 2 N. Anari, K. Liu, and S. O. Gharan. Spectral Independence in High-Dimensional Expanders and Applications to the Hardcore Model. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1319–1330, 2020. doi:10.1109/FOCS46700.2020.00125.

- 3 Alexander Barvinok. *Combinatorics and Complexity of Partition Functions*, volume 30 of *Algorithms and Combinatorics*. Springer International Publishing, Cham, 2016. doi:10.1007/978-3-319-51829-9.
- 4 Nayantara Bhatnagar, Allan Sly, and Prasad Tetali. Decay of correlations for the hardcore model on the d -regular random graph. *Electronic Journal of Probability*, 21(0), 2016. doi:10.1214/16-EJP3552.
- 5 R. Buble and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pages 223–231, Miami Beach, FL, USA, 1997. IEEE Comput. Soc. doi:10.1109/SFCS.1997.646111.
- 6 Z. Chen, K. Liu, and E. Vigoda. Rapid Mixing of Glauber Dynamics up to Uniqueness via Contraction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1307–1318, November 2020. doi:10.1109/FOCS46700.2020.00124.
- 7 Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal Mixing of Glauber Dynamics: Entropy Factorization via High-Dimensional Expansion. *CoRR*, 2020. arXiv:2011.02075.
- 8 Radu Curticapean, Holger Dell, Fedor Fomin, Leslie Ann Goldberg, and John Lapinskas. A fixed-parameter perspective on #BIS. *Algorithmica*, 81(10):3844–3864, 2019.
- 9 Jonathan Cutler and A.J. Radcliffe. The maximum number of complete subgraphs in a graph with given maximum degree. *Journal of Combinatorial Theory, Series B*, 104:60–71, 2014. doi:10.1016/j.jctb.2013.10.003.
- 10 Ewan Davies, Matthew Jenssen, and Will Perkins. A proof of the Upper Matching Conjecture for large graphs. *CoRR*, April 2020. arXiv:2004.06695.
- 11 Ewan Davies, Matthew Jenssen, Will Perkins, and Barnaby Roberts. On the average size of independent sets in triangle-free graphs. *Proc. Amer. Math. Soc.*, 146(1):111–124, 2017. doi:10.1090/proc/13728.
- 12 Ewan Davies, Matthew Jenssen, Will Perkins, and Barnaby Roberts. Tight bounds on the coefficients of partition functions via stability. *Journal of Combinatorial Theory, Series A*, 160:1–30, November 2018. doi:10.1016/j.jcta.2018.06.005.
- 13 Martin Dyer, Leslie Ann Goldberg, Catherine Greenhill, and Mark Jerrum. The relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2004. doi:10.1007/978-3-642-04016-0.
- 14 Talya Eden, Dana Ron, and C. Seshadhri. On Approximating the Number of k -Cliques in Sublinear Time. *SIAM Journal on Computing*, 49(4):747–771, January 2020. doi:10.1137/18M1176701.
- 15 Friedrich Eisenbrand and Fabrizio Grandoni. On the complexity of fixed parameter clique and dominating set. *Theoretical Computer Science*, 326(1):57–67, 2004. doi:10.1016/j.tcs.2004.05.009.
- 16 Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the Partition Function for the Antiferromagnetic Ising and Hard-Core Models. *Combinatorics, Probability and Computing*, 25(4):500–559, July 2016. doi:10.1017/S0963548315000401.
- 17 B. V. Gnedenko. On a local limit theorem of the theory of probability. *Uspehi Matem. Nauk (N. S.)*, 3(3(25)):187–194, 1948.
- 18 C. Greenhill. The complexity of counting colourings and independent sets in sparse graphs and hypergraphs. *Comput. complex.*, 9(1):52–72, November 2000. doi:10.1007/PL00001601.
- 19 David G. Harris and Vladimir Kolmogorov. Parameter estimation for Gibbs distributions. *CoRR*, 2021. arXiv:2007.10824.
- 20 Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18(6):1149–1178, 1989.
- 21 Mark Jerrum and Alistair Sinclair. Polynomial-Time Approximation Algorithms for the Ising Model. *SIAM Journal on Computing*, 22(5):1087–1116, October 1993. doi:10.1137/0222066.
- 22 Mark Jerrum and Alistair Sinclair. The Markov chain Monte Carlo method: An approach to approximate counting and integration. In *Approximation Algorithms for NP-Hard Problems*. PWS Pub. Co, 1997.

- 23 F. P. Kelly. Stochastic Models of Computer Communication Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3):379–395, 1985.
- 24 J.L. Lebowitz, B. Pittel, D. Ruelle, and E.R. Speer. Central limit theorems, Lee–Yang zeros, and graph-counting polynomials. *Journal of Combinatorial Theory, Series A*, 141:147–183, July 2016. doi:10.1016/j.jcta.2016.02.009.
- 25 Marcus Michelen and Julian Sahasrabudhe. Central limit theorems and the geometry of polynomials. *CoRR*, 2019. arXiv:1908.09020.
- 26 Marcus Michelen and Julian Sahasrabudhe. Central limit theorems from the roots of probability generating functions. *Advances in Mathematics*, 358:106840, December 2019. doi:10.1016/j.aim.2019.106840.
- 27 Viresh Patel and Guus Regts. Deterministic Polynomial-Time Approximation Algorithms for Partition Functions and Graph Polynomials. *SIAM J. Comput.*, 46(6):1893–1919, January 2017. doi:10.1137/16M1101003.
- 28 Han Peters and Guus Regts. On a Conjecture of Sokal Concerning Roots of the Independence Polynomial. *Michigan Math. J.*, 68(1):33–55, April 2019. doi:10.1307/mmj/1541667626.
- 29 Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, July 1989. doi:10.1016/0890-5401(89)90067-9.
- 30 Allan Sly. Computational Transition at the Uniqueness Threshold. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 287–296, Las Vegas, NV, USA, October 2010. IEEE. doi:10.1109/F0CS.2010.34.
- 31 Allan Sly and Nike Sun. Counting in two-spin models on d-regular graphs. *Ann. Probab.*, 42(6):2383–2416, November 2014. doi:10.1214/13-AOP888.
- 32 Leslie G. Valiant. The Complexity of Enumeration and Reliability Problems. *SIAM J. Comput.*, 8(3):410–421, August 1979. doi:10.1137/0208032.
- 33 Dror Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing - STOC '06*, page 140, Seattle, WA, USA, 2006. ACM Press. doi:10.1145/1132516.1132538.
- 34 Yufei Zhao. Extremal regular graphs: independent sets and graph homomorphisms. *The American Mathematical Monthly*, 124(9):827–843, 2017. doi:10.4169/amer.math.monthly.124.9.827.