



HAL
open science

Sublinear-time algorithms for monomer–dimer systems on bounded degree graphs

Marc Lelarge, Hang Zhou

► **To cite this version:**

Marc Lelarge, Hang Zhou. Sublinear-time algorithms for monomer–dimer systems on bounded degree graphs. *Theoretical Computer Science*, 2014, 548, pp.10. 10.1016/j.tcs.2014.06.040 . hal-01090580

HAL Id: hal-01090580

<https://inria.hal.science/hal-01090580v1>

Submitted on 28 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Sublinear-Time Algorithms for Monomer-Dimer Systems on Bounded Degree Graphs

Marc Lelarge and Hang Zhou

INRIA, École Normale Supérieure, France
{marc.lelarge, hang.zhou}@ens.fr

Abstract. For a graph G , let $Z(G, \lambda)$ be the partition function of the monomer-dimer system defined by $\sum_k m_k(G) \lambda^k$, where $m_k(G)$ is the number of matchings of size k in G . We consider graphs of bounded degree and develop a sublinear-time algorithm for estimating $\log Z(G, \lambda)$ at an arbitrary value $\lambda > 0$ within additive error ϵn with high probability. The query complexity of our algorithm does not depend on the size of G and is polynomial in $1/\epsilon$, and we also provide a lower bound quadratic in $1/\epsilon$ for this problem. This is the first analysis of a sublinear-time approximation algorithm for a $\#P$ -complete problem. Our approach is based on the correlation decay of the Gibbs distribution associated with $Z(G, \lambda)$. We show that our algorithm approximates the probability for a vertex to be covered by a matching, sampled according to this Gibbs distribution, in a near-optimal sublinear time. We extend our results to approximate the average size and the entropy of such a matching within an additive error with high probability, where again the query complexity is polynomial in $1/\epsilon$ and the lower bound is quadratic in $1/\epsilon$. Our algorithms are simple to implement and of practical use when dealing with massive datasets. Our results extend to other systems where the correlation decay is known to hold as for the independent set problem up to the critical activity.

1 Introduction

The area of sublinear-time algorithms is an emerging area of computer science which has its root in the study of massive data sets [6,25]. Internet, social networks or communication networks are typical examples of graphs with potentially millions of vertices representing agents, and edges representing possible interactions among those agents. In this paper, we present sublinear-time algorithms for graph problems. We are concerned more with problems of counting and statistical inference and less with optimization. For example, in a mobile call graphs, phone calls can be represented as a matching of the graph where each edge has an activity associated to the intensity of the interactions between the pair of users. Given such a graphs, with local activities on edges, we would like to answer questions like: what is the size of a typical matching? for a given user what is the probability of being matched? As another example, models of statistical physics have been proposed to model social interactions. In particular,

spin systems are a general framework for modeling nearest-neighbor interactions on graphs. In this setting, the activity associated to each edge allows to model a perturbed best-response dynamics [2]. Again in this setting, it is interesting to compute estimations for the number of agents playing a given strategy or the probability for an agent in the graph to play a given strategy at equilibrium.

There are now quite a few results on sublinear-time approximation algorithms for graph optimization problems: minimum spanning tree weight [5], minimum set cover [22], maximum matching [22,33] and minimum vertex cover [22,23,24]. There are also a couple of works on sublinear-time algorithms for statistical and counting problems, e.g., approximating the average degree of a graph [7,11] and approximating the number of occurrences of a certain structure (such as a star) in a graph [12]. Our focus in this paper is on the algorithmic problems arising in statistical physics and classical combinatorics [31]. We now present the monomer-dimer problem which will be the main focus of our paper.

Let $G = (V, E)$ be an undirected graph with $|V| = n$ vertices and $|E| = m$ edges, where we allow G to contain parallel edges and self-loops. We denote by $N(G, v)$ the set of neighbors of v in G . We consider bounded degree graphs with $\max_v |N(G, v)| \leq \Delta$. In a monomer-dimer system, the vertices are covered by non-overlapping arrangement of monomers (molecules occupying one vertex of G) and dimers (molecules occupying two adjacent vertices of G) [13]. It is convenient to identify monomer-dimer arrangements with matchings; a matching in G is a subset $M \subset E$ such that no two edges in M share an endpoint. Thus, a matching of cardinality $|M| = k$ corresponds exactly to a monomer-dimer arrangement with k dimers and $n - 2k$ monomers. Let \mathbb{M} be the set of matchings of G . To each matching M , a weight $\lambda^{|M|}$ is assigned, where $\lambda > 0$ is called the activity. The partition function of the system is defined by $Z(G, \lambda) = \sum_{M \in \mathbb{M}} \lambda^{|M|}$, and the Gibbs distribution on the space \mathbb{M} is defined by $\pi_{G, \lambda}(M) = \frac{\lambda^{|M|}}{Z(G, \lambda)}$. The function $Z(G, \lambda)$ is also of combinatorial interest and called the *matching polynomial* in this context [20]. For example, $Z(G, 1)$ enumerates all matchings in G . From an algorithmic viewpoint, no feasible method is known for computing $Z(G, \lambda)$ exactly for general monomer-dimer system; indeed, for any fixed value of $\lambda > 0$, the problem of computing $Z(G, \lambda)$ exactly in a graph of bounded degree Δ is complete for the class #P of enumeration problems, when $\Delta \geq 5$ (see [29]). The focus on computing $Z(G, \lambda)$ shifted to finding approximate solutions in polynomial time. For example, the Markov Chain Monte Carlo (MCMC) method yields a provably efficient algorithm for finding an approximate solution. Based on the equivalence between the counting problem (computing $Z(G, \lambda)$) and the sampling problem (according to $\pi_{G, \lambda}$) [16], this approach focuses on rapidly mixing Markov chains to obtain appropriate random samples. A Fully Polynomial-time Randomized Approximation Scheme (FPRAS) for computing the total number of matchings based on MCMC was provided by Jerrum and Sinclair [14,26].

Another related problem in the monomer-dimer system is the average size of a matching sampled according to $\pi_{G, \lambda}$, defined by $E(G, \lambda) = \sum_{M \in \mathbb{M}} |M| \pi_{G, \lambda}(M)$. Sinclair and Srivastava recently proved in [28] that for any fixed value of $\lambda > 0$, the problem of computing $E(G, \lambda)$ exactly in a bounded degree graph (allowing

parallel edges) is #P-hard, for any maximum degree $\Delta \geq 5$. Thus again we are interested in finding approximate solutions to this problem.

In order to study sublinear-time approximation algorithms for these problems, we use the approach based on the concept of correlation decay originating in statistical physics [21] and which has been used to get a deterministic approximation scheme for counting matchings in polynomial time [1]. It follows already from [13] that the marginals of the probability distribution $\pi_{G,\lambda}$ are local in nature: the local structure of the graph around a vertex v allows to compute an approximation of the corresponding marginal. In the computer science literature, this property follows from the so-called *correlation decay property*. Our algorithm is then simple to understand: we need only to sample a fixed number of vertices, approximate the marginals associated to these vertices locally and then from these values output an estimate for the desired quantity. The correlation decay property also holds for other systems such as the independent set problem [30], the coloring problem [9], and the two-state spin system [18,19,27]. In Appendix E, we extend our technique to the independent set problem. We believe that similar extensions can be done for other systems as soon as the correlation decay property holds.

A graph G is represented by two kinds of oracles \mathcal{D} and \mathcal{N} such that $\mathcal{D}(v)$ returns the degree of $v \in V$ and $\mathcal{N}(v, i)$ returns the i^{th} (with $1 \leq i \leq \mathcal{D}(v)$) neighbor of $v \in V$. The efficiency of an algorithm is measured by its query complexity, i.e. the total number of accesses to \mathcal{D} and \mathcal{N} . Let VAL denote a real value associated with the graph. We say that $\widehat{\text{VAL}}$ is an ϵ -approximation of VAL if $\widehat{\text{VAL}} - \epsilon \leq \text{VAL} \leq \widehat{\text{VAL}} + \epsilon$, where $\epsilon > 0$ is specified as an input parameter. An algorithm is called an ϵ -approximation algorithm for VAL if for any graph G , it computes an ϵ -approximation of VAL with high probability (e.g., at least $\frac{2}{3}$). In our model, we consider the case of constant maximum degree Δ as ϵ tends to zero, i.e., we always first take the limit as $\epsilon \rightarrow 0$ and then the limit $\Delta \rightarrow \infty$.

Our main contribution (Theorem 5) is an ϵn -approximation algorithm for $\log Z(G, \lambda)$ in a graph G of bounded degree Δ . The query complexity of the algorithm is $\tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta})}\right)$, which does not depend on the size of the graph. From the relation between the partition function and the matching statistics, we then obtain ϵn -approximation algorithms for the average size of a matching and the entropy of $\pi_{G,\lambda}$ with the same query complexity as before. We also provide the $\Omega(1/\epsilon^2)$ query lower bound for ϵn -approximation algorithms for $\log Z(G, \lambda)$ and the other two problems.

The main tool of the above algorithms is the approximation of the marginal $p_{G,\lambda}(v)$, which is the probability that the vertex v is not covered by a matching under the Gibbs distribution. We estimate $p_{G,\lambda}(v)$ for an arbitrary vertex $v \in V$ within an error of $\epsilon > 0$ with near-optimal query complexity $\tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta})}\right)$.

The rest of the paper is organized as follows. In Section 2, we prove our first main result concerning local computations for matchings. Based on this result, we construct an ϵn -approximation algorithm for the partition function $Z(G, \lambda)$ in Section 3 and ϵn -approximation algorithms for the average size of a matching

and the entropy of $\pi_{G,\lambda}$ in Section 4. We also provide query lower bounds in these two sections. In Section 5, we give some applications of our technique for approximating the permanent of constant degree expander graphs and the size of a maximum matching (in this last case, our algorithm is outperformed by [33]). In Appendix F, we also show the efficiency of our algorithms by testing on large real-world networks.

2 Local Computations for Matchings

Recall that we defined for all $\lambda > 0$, the Gibbs distribution on matchings of a graph G by:

$$\forall M \in \mathbb{M}, \quad \pi_{G,\lambda}(M) = \frac{\lambda^{|M|}}{Z(G,\lambda)} \text{ where } Z(G,\lambda) = \sum_{M \in \mathbb{M}} \lambda^{|M|}.$$

The focus in this section is on the approximation of the probability that a vertex $v \in V$ is not covered by a matching:

$$p_{G,\lambda}(v) := \sum_{M \not\ni v} \pi_{G,\lambda}(M),$$

where $M \not\ni v$ is a matching not covering v .

First notice that

$$p_{G,\lambda}(v) = \frac{Z(G \setminus \{v\}, \lambda)}{Z(G, \lambda)}, \quad (1)$$

where $G \setminus \{v\}$ is the graph obtained from G by removing the vertex v and all incident edges. Then we have

$$Z(G, \lambda) = Z(G \setminus \{v\}, \lambda) + \lambda \sum_{u \in N(G,v)} Z(G \setminus \{u, v\}, \lambda),$$

so that dividing by $Z(G \setminus \{v\}, \lambda)$, we get

$$p_{G,\lambda}(v) = \frac{1}{1 + \lambda \sum_{u \in N(G,v)} p_{G \setminus \{v\}, \lambda}(u)}. \quad (2)$$

This recursive expression for $p_{G,\lambda}(v)$ is well-known and allows to compute the marginal $p_{G,\lambda}(v)$ exactly for each $v \in V$. We follow the approach of Godsil [10]. First, we recall the notion of *path-tree* associated with a rooted graph: if G is any rooted graph with root v_0 , we define its path-tree $T_G(v_0)$ as the rooted tree whose vertex-set consists of all *finite simple paths* starting at the root v_0 ; whose edges are the pairs $\{P, P'\}$ of the form $P = v_0 \dots v_k$, $P' = v_0 \dots v_k v_{k+1}$ ($k \geq 0$); and whose root is the single-vertex path v_0 . By a *finite simple path*, we mean here a finite sequence of distinct vertices $v_0 \dots v_k$ ($k \geq 0$) such that $v_i v_{i+1} \in E$ for all $0 \leq i < k$. Note that the notion of path-tree is similar to the more standard

notion of computation tree, the main difference being that for any finite graph G , the path-tree is always finite (although its size might be much larger than the size of the original graph G).

For every node u in the path-tree $T_G(v)$, define $\text{Ch}(u)$ to be the set of children of u in $T_G(v)$. The recursion (2) easily implies $p_{G,\lambda}(v) = p_{T_G(v),\lambda}(v)$ and $p_{T_G(v),\lambda}(v) = x_v(v)$, where the vector $\mathbf{x}(v) = (x_u(v), u \in T_G(v))$ solves the recursion:

$$\forall u \in T_G(v), \quad x_u(v) = \frac{1}{1 + \lambda \sum_{w \in \text{Ch}(u)} x_w(v)} \quad (3)$$

(by convention a sum over the empty set is zero).

In order to approximate $p_{G,\lambda}(v)$, we will show that it suffices to solve the recursion (3) restricted to a truncated path-tree of $T_G(v)$. For any $h \geq 1$, let $T_G^h(v)$ be the path-tree truncated at depth h and let $\mathbf{x}^h(v) = (x_u^h(v), u \in T_G^h(v))$ be the solution of the recursion (3) when the path-tree is replaced by the truncated version $T_G^h(v)$. Clearly $x_v^h(v) = p_{G,\lambda}(v)$ for any $h \geq n$ and the following lemma gives a quantitative estimate on how large h needs to be in order to get an ϵ -approximation of $p_{G,\lambda}(v)$.

Lemma 1. *There exists $\bar{h}(\epsilon, \Delta)$ such that $|\log x_v^h(v) - \log p_{G,\lambda}(v)| \leq \epsilon$ for any $h \geq \bar{h}(\epsilon, \Delta)$. Moreover $\bar{h}(\epsilon, \Delta) = \tilde{O}\left(\sqrt{\Delta} \log(1/\epsilon)\right)$ and satisfies*

$$\lim_{\Delta \rightarrow \infty} \frac{1}{\sqrt{\Delta}} \lim_{\epsilon \rightarrow 0} \frac{\bar{h}(\epsilon, \Delta)}{\log(1/\epsilon)} = \sqrt{\lambda}.$$

Proof. Theorem 3.2 in [1] proves that:

$$|\log x_v^h(v) - \log p_{G,\lambda}(v)| \leq \left(1 - \frac{2}{\sqrt{1 + \lambda\Delta} + 1}\right)^{h/2} \log(1 + \lambda\Delta). \quad (4)$$

The lemma then follows directly by taking $\bar{h}(\epsilon, \Delta)$ to be the h such that the right-hand side equals ϵ . \square

We now present the algorithmic implication of Lemma 1. We start with a simple remark. The exact value for $\bar{h}(\epsilon, \Delta)$ follows from the proof of the lemma, however this value will not be required in what follows as shown by the following argument: the fact that $(z_1, \dots, z_\Delta) \mapsto \left(1 + \lambda \sum_{i=1}^{\Delta} z_i\right)^{-1}$ is strictly decreasing in each positive variable z_i implies (by a simple induction) that for any $k \geq 0$, we have

$$x_v^{2k+1}(v) \leq x_v^{2k+3}(v) \leq p_{G,\lambda}(v) \leq x_v^{2k+2}(v) \leq x_v^{2k}(v). \quad (5)$$

Consider an algorithm that computes $x_v^h(v)$ for increasing values of h and stops at the first time two consecutive outputs are such that $|\log x_v^{h+1}(v) - \log x_v^h(v)| \leq \epsilon$. By Lemma 1, it takes at most $\bar{h}(\epsilon, \Delta)$ iterations and the last output will be an ϵ -approximation of $\log p_{G,\lambda}(v)$.

The algorithm APPROX-MARGINAL(λ, ϵ, v) provides an estimate of $p_{G,\lambda}(v)$, based on the Depth-First-Search (DFS) on the truncated path-tree rooted at v . In the algorithm DFS(λ, h, s, ℓ), integer h is the truncated level of the path tree $T_G(v)$; $s \in V$ is the current node in the graph G visiting by the DFS; and $path$ maintains an array of nodes in G which form the path from v to s during the DFS. This path also corresponds to a node in the path-tree $T_G^h(v)$ and let ℓ be the length of $path$. The algorithm DFS(λ, h, s, ℓ) computes recursively the marginal probability of $path$ in $T_G^h(v)$. Recall that $\mathcal{D}(v)$ returns the degree of $v \in V$ and $\mathcal{N}(v, i)$ returns the i^{th} (with $1 \leq i \leq \mathcal{D}(v)$) neighbor of $v \in V$.

```

APPROX-MARGINAL( $\lambda, \epsilon, v$ )
1   $x[1] \leftarrow \text{DFS}(\lambda, 1, 1, v)$ 
2   $x[2] \leftarrow \text{DFS}(\lambda, 1, 2, v)$ 
3   $h \leftarrow 2$ 
4  while  $|\log x[h] - \log x[h-1]| > \epsilon/e$ 
5      do  $h \leftarrow h + 1$ 
6       $x[h] \leftarrow \text{DFS}(\lambda, h, v, 1)$ 
7  return  $x[h]$ 

```

```

DFS( $\lambda, h, s, \ell$ )
1  if  $\ell = h$ 
2      then return 1
3   $A \leftarrow 1, path[\ell] \leftarrow s$ 
4  for  $i \leftarrow 1$  to  $\mathcal{D}(s)$ 
5      do  $t \leftarrow \mathcal{N}(s, i)$ 
6          if  $\forall j \in [1, \ell], t \neq path[j]$ 
7              then  $A \leftarrow A + \text{DFS}(\lambda, h, t, \ell + 1)$ 
8  return  $1/(\lambda A)$ 

```

Proposition 2. *The algorithm APPROX-MARGINAL(λ, ϵ, v) gives an estimate \hat{p} of $p_{G,\lambda}(v)$, such that $|\hat{p} - p_{G,\lambda}(v)|$ and $|\log \hat{p} - \log p_{G,\lambda}(v)|$ are both smaller than ϵ . Its query complexity is $\overline{Q}(\epsilon, \Delta) = \tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta})}\right)$. In addition, $Q = \overline{Q}(\epsilon, \Delta)$ satisfies*

$$\lim_{\Delta \rightarrow \infty} \frac{1}{\sqrt{\Delta} \log \Delta} \lim_{\epsilon \rightarrow 0} \frac{\log Q}{\log(1/\epsilon)} = \sqrt{\lambda}. \quad (6)$$

Proof. Let h be the final truncated level of the path-tree $T_G(v)$ in the algorithm. We have $\hat{p} = x_v^h(v)$ and $|\log x_v^h(v) - \log x_v^{h-1}(v)| < \epsilon/e$. Thus $|\log \hat{p} - \log p_{G,\lambda}(v)| < \epsilon/e$ by Inequality (5). Since $p_{G,\lambda}(v)$ and \hat{p} are at most 1, we then have $|\hat{p} - p_{G,\lambda}(v)| < \epsilon$. The number of nodes visited by the algorithm is $O(\Delta^h)$, so the number of queries is also $O(\Delta^h)$. The proposition follows by applying the upper bound $\bar{h}(\epsilon, \Delta)$ on h from Lemma 1. \square

Remark. *In Section 3, we need to estimate the marginal probability at the node v in the graph $G_v = \{u \in V \mid u \succeq v\}$ instead of the graph G , where \succ is some*

total order over V . To achieve this, we only need to add an additional constraint $t \succeq v$ to line 6 of the DFS algorithm. Denote $\text{APPROX-MARGINAL}^*(\lambda, \epsilon, v)$ to be the modified version of $\text{APPROX-MARGINAL}(\lambda, \epsilon, v)$ with the underlying graph G_v . Again, Proposition 2 holds for the algorithm $\text{APPROX-MARGINAL}^*(\lambda, \epsilon, v)$ by replacing G by G_v .

The next proposition shows that there exists some $\underline{Q}(\epsilon, \Delta)$, such that $Q = \underline{Q}(\epsilon, \Delta)$ satisfies Equation (6) and that $\underline{Q}(\epsilon, \Delta)$ is a query lower bound for computing an ϵ -approximation of $p_{G,\lambda}(v)$. This implies that Algorithm APPROX-MARGINAL is optimal when the influence of ϵ is much larger than that of Δ . The idea of the lower bound proof is to construct two instances of *almost full Δ -ary trees* whose marginal probabilities at the root differ by more than ϵ , while any approximation algorithm using a small number of queries cannot distinguish them. See Appendix A for a detailed proof of this proposition.

Proposition 3. *In order to approximate the marginal $p_{G,\lambda}(v)$ within an additive error ϵ , any deterministic or randomized algorithm¹ requires $\Omega(\underline{Q}(\epsilon, \Delta))$ queries where $Q = \underline{Q}(\epsilon, \Delta)$ satisfies Equation (6).*

Remark. *As noted in the introduction, the model with λ_e ($e \in E$) varying across the edges is of practical interest (allowing to model various intensities on edges). As soon as there exists λ_{\max} such that for all $e \in E$, we have $\lambda_e \in [0, \lambda_{\max}]$, it is easy to extend the results of this section to the more general model defined by (note that λ is now a vector in $[0, \lambda_{\max}]^E$): $\pi_{G,\lambda}(M) = \frac{\prod_{e \in M} \lambda_e}{Z(G,\lambda)}$ where, $Z(G, \lambda) = \sum_{M \in \mathcal{M}} \prod_{e \in M} \lambda_e$. Results in this section and Sections 3 and 4 hold provided λ is replaced by λ_{\max} .*

3 Approximating the Partition Function

First, we need an arbitrary total order \succ over V . We can achieve this by assigning a random number $a_v \in [0, 1]$ to each vertex v and then defining $u \succ v$ as $a_u > a_v$. However, if there are only a small number of vertices involved in our computation, we do not need to generate random numbers for all vertices. Using the technique in [23], we generate a random number each time we visit a new vertex and then save its value for the later visits. Generating a random number can be done in sublinear time and the number of vertices in our computation is at most twice the number of queries, which will later be proved to be a constant independent of n . As a result, the total time complexity for this random generation is sublinear.

Define $G_v = \{u \in V \mid u \succeq v\}$. The following formula which allows us to compute the partition function from the marginals is obtained easily from (1):

$$\log Z(G, \lambda) = \sum_{v \in V} -\log p_{G_v, \lambda}(v). \tag{7}$$

¹ In the randomized case, the algorithm is expected to provide *always* an estimate with an additive error ϵ , and the proposition implies a lower bound on the *average* number of queries of such an algorithm.

The algorithm below estimates $\log Z(G, \lambda)$. We sample $\lceil C/\epsilon^2 \rceil$ vertices uniformly at random from V , where C is some fixed constant. For every sampled vertex u , we compute an estimate of the marginal $p_{G_u, \lambda}(u)$ using the algorithm APPROX-MARGINAL $^*(\lambda, \epsilon/2, u)$. We then obtain an estimate of $Z(G, \lambda)$ from the estimates of marginals at sampled vertices.

APPROX-PARTITION-FUNCTION(λ, ϵ)

- 1 $s \leftarrow \lceil C/\epsilon^2 \rceil$
- 2 $U \leftarrow$ random multi-subset of V with s elements
- 3 **return** $(n/s) \cdot (\sum_{u \in U} -\log(\text{APPROX-MARGINAL}^*(\lambda, \epsilon/2, u)))$

We recall a basic lemma which follows from Hoeffding's inequality and which will be used several times in the paper.

Lemma 4. (see [4]) *Let V be a set of n real numbers in $[A, B]$, where A and B are constant. Let V' be a multi-subset of V consisting of $\Theta(1/\epsilon^2)$ elements chosen uniformly and independently at random. Let AVG be the average of all elements and AVG' be the average of sampled elements. Then with high constant probability, we have: $\text{AVG}' - \epsilon \leq \text{AVG} \leq \text{AVG}' + \epsilon$.*

Theorem 5. APPROX-PARTITION-FUNCTION(λ, ϵ) is an ϵn -approximation algorithm for $\log Z(G, \lambda)$ with query complexity $\tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta})}\right)$.

Proof. Let $A = \sum_{v \in V} -\log(\text{APPROX-MARGINAL}^*(\lambda, \epsilon/2, v))$. By Proposition 2 and Equation (7), A is an $\epsilon n/2$ -approximation of $\log Z(G, \lambda)$. By Lemma 4, there exists some constant C such that approximating the marginal probability at $\lceil C/\epsilon^2 \rceil$ sampled nodes gives an $\epsilon n/2$ -approximation of A with high probability. This implies an ϵn -approximation of $\log Z(G, \lambda)$ with high probability. The query complexity of this algorithm is $\lceil C/\epsilon^2 \rceil \cdot \overline{\mathcal{Q}}(\epsilon/2, \Delta) = \tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta})}\right)$. \square

Note that the size of any maximal matching is always lower bounded by $\frac{m}{2\Delta-1}$, where m is the number of edges. In particular, since $Z(G, 1)$ is the total number of matchings, we have $\frac{m}{2\Delta-1} \log 2 \leq \log Z(G, 1) \leq m \log 2 \leq \frac{n\Delta}{2} \log 2$ so that if $m = \Omega(n)$, we also have $\log Z(G, 1) = \Theta(n)$. Hence, if ϵ and Δ are constants and $m = \Omega(n)$, the error in the output of our algorithm is of the same order as the evaluated quantity. This is in contrast with the FPTAS (Fully Polynomial-Time Approximation Scheme) in [1] or the FPRAS (Fully Polynomial-time Randomized Approximation Scheme) in [14,26] which outputs an ϵ -approximation instead of an ϵn -approximation. Of course, we can let ϵ tend to 0 with n like c/n in Theorem 5, so that our result (when Δ is constant) is consistent with the FPTAS in [1]. Indeed, in this case, clearly no sampling is required and if we replace the sampling step by a visit of each vertex, our algorithm is the same as in [1].

When we assume Δ to be fixed, the query complexity of the above algorithm is polynomial in $1/\epsilon$. Next we give a lower bound on the query complexity which is quadratic in $1/\epsilon$. In the proof, we use a lower bound result from [5], which

is based on Yao's Minimax Principle [32]. See Appendix B for the proof of the following theorem.

Theorem 6. *Any deterministic or probabilistic ϵn -approximation algorithm for $\log Z(G, \lambda)$ needs $\Omega(1/\epsilon^2)$ queries on average. It is assumed that $\epsilon > C/\sqrt{n}$ for some constant C .*

4 Approximating Matching Statistics

We define the average size $E(G, \lambda)$ and the entropy $S(G, \lambda)$ of a matching as:

$$E(G, \lambda) = \sum_{M \in \mathbb{M}} |M| \pi_{G, \lambda}(M) \quad \text{and} \quad S(G, \lambda) = - \sum_{M \in \mathbb{M}} \pi_{G, \lambda}(M) \log \pi_{G, \lambda}(M).$$

The following algorithm estimates $E(G, \lambda)$, where C is a fixed constant.

APPROX-MATCHING-STATISTICS(λ, ϵ)

- 1 $s \leftarrow \lceil C/\epsilon^2 \rceil$
- 2 $U \leftarrow$ random multi-subset of V with s elements
- 3 **return** $n - (n/2s) \cdot \sum_{u \in U} (\text{APPROX-MARGINAL}(\lambda, \epsilon/2, u))$

Theorem 7. APPROX-MATCHING-STATISTICS(λ, ϵ) is an ϵn -approximation algorithm for $E(G, \lambda)$ with query complexity $\tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta})}\right)$. In addition, any ϵn -approximation algorithm for $E(G, \lambda)$ needs $\Omega(1/\epsilon^2)$ queries.

Proof. Let $A = \sum_{v \in V} \text{APPROX-MARGINAL}(\lambda, \epsilon/2, v)$. By Proposition 2, A is an $\epsilon n/2$ -approximation of $\sum_{v \in V} p_{G, \lambda}(v)$. By Lemma 4, there exists some constant C such that approximating the marginal probability at $\lceil C/\epsilon^2 \rceil$ sampled nodes gives an $\epsilon n/2$ -approximation of A with high probability. This implies an ϵn -approximation of $\sum_{v \in V} p_{G, \lambda}(v)$ with high probability. Since $E(G, \lambda) = n - \sum_{v \in V} p_{G, \lambda}(v)/2$, we thus get an ϵn -approximation of $E(G, \lambda)$ with high probability. The query complexity of this algorithm is $\lceil C/\epsilon^2 \rceil \cdot \overline{Q}(\epsilon/2, \Delta) = \tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta})}\right)$. The query lower bound is obtained similarly as Theorem 6. \square

Corollary 8. *We have an ϵn -approximation algorithm for $S(G, \lambda)$ with query complexity $\tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta})}\right)$. In addition, any ϵn -approximation algorithm for $S(G, \lambda)$ needs $\Omega(1/\epsilon^2)$ queries.*

Proof. A simple calculation gives: $S(G, \lambda) = \log Z(G, \lambda) - \log \lambda \cdot E(G, \lambda)$. Let \widehat{Z} be the output of APPROX-PARTITION-FUNCTION($\lambda, \epsilon/2$) and \widehat{E} be the output of APPROX-MATCHING-STATISTICS($\lambda, \epsilon/(2 \log \lambda)$). By Theorem 5 and Theorem 7, $\widehat{Z} - \log \lambda \cdot \widehat{E}$ is an ϵn -estimate of $S(G, \lambda)$ with high probability. Both \widehat{Z} and \widehat{E} are computed using $\tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta})}\right)$ queries. The query lower bound is obtained similarly as Theorem 6. \square

5 Some Applications

So far, we did consider that the parameter λ is fixed. Letting λ grow with $\frac{1}{\epsilon}$ allows us to get new results for the permanent of a matrix. There is a FPRAS for the permanent of a matrix with non-negative entries [15]. When the matrix is the adjacency matrix of a *constant degree expander* graph, there is a PTAS to estimate the permanent within a multiplicative factor $(1 + \epsilon)^n$ [8]. Using a key technical result of [8], we get a sublinear-time algorithm within the same multiplicative factor (see Appendix C).

Note that for a fixed graph G , if $\lambda \rightarrow \infty$ then the distribution $\pi_{G,\lambda}$ converges toward the uniform distribution on maximum matchings. Indeed, using a bound derived in [3], we can show that if λ grows exponentially with $\frac{1}{\epsilon}$ our technique allows to approximate the size of a maximum matching (see Appendix D). However our algorithm performs badly with respect to [33].

Acknowledgments. The authors acknowledge the support of the French *Agence Nationale de la Recherche (ANR)* under reference ANR-11-JS02-005-01 (GAP project).

References

1. Bayati, M., Gamarnik, D., Katz, D., Nair, C., Tetali, P.: Simple deterministic approximation algorithms for counting matchings. In: STOC. pp. 122–127. ACM (2007)
2. Blume, L.E.: The statistical mechanics of strategic interaction. *Games Econom. Behav.* 5(3), 387–424 (1993)
3. Bordenave, C., Lelarge, M., Salez, J.: Matchings on infinite graphs. *Probability Theory and Related Fields* pp. 1–26 (2012)
4. Canetti, R., Even, G., Goldreich, O.: Lower bounds for sampling algorithms for estimating the average. *Information Processing Letters* 53(1), 17–25 (1995)
5. Chazelle, B., Rubinfeld, R., Trevisan, L.: Approximating the minimum spanning tree weight in sublinear time. *SIAM Journal on computing* 34(6), 1370–1379 (2005)
6. Czumaj, A., Sohler, C.: Sublinear-time algorithms. *Bulletin of the EATCS* 89, 23–47 (2006)
7. Feige, U.: On sums of independent random variables with unbounded variance and estimating the average degree in a graph. *SIAM Journal on Computing* 35(4), 964–984 (2006)
8. Gamarnik, D., Katz, D.: A deterministic approximation algorithm for computing the permanent of a 0, 1 matrix. *Journal of Computer and System Sciences* 76(8), 879–883 (2010)
9. Gamarnik, D., Katz, D.: Correlation decay and deterministic FPTAS for counting list-colorings of a graph. In: SODA. pp. 1245–1254. SIAM (2007)
10. Godsil, C.D.: Matchings and walks in graphs. *J. Graph Theory* 5(3), 285–297 (1981)
11. Goldreich, O., Ron, D.: Approximating average parameters of graphs. *Random Structures and Algorithms* 32(4), 473–493 (2008)
12. Gonen, M., Ron, D., Shavitt, Y.: Counting stars and other small subgraphs in sublinear-time. *SIAM Journal on Discrete Mathematics* 25(3), 1365–1411 (2011)

13. Heilmann, O.J., Lieb, E.H.: Theory of monomer-dimer systems. *Comm. Math. Phys.* 25, 190–232 (1972)
14. Jerrum, M.: Counting, sampling and integrating: algorithms and complexity. *Lectures in Mathematics ETH Zürich*, Birkhäuser Verlag, Basel (2003)
15. Jerrum, M., Sinclair, A., Vigoda, E.: A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM* 51(4), 671–697 (2004)
16. Jerrum, M., Valiant, L., Vazirani, V.: Random generation of combinatorial structures from a uniform distribution. *Theoret. Comput. Sci.* 43(2-3), 169–188 (1986)
17. Kahn, J., Kim, J.H.: Random matchings in regular graphs. *Combinatorica* 18, 201–226 (1998)
18. Li, L., Lu, P., Yin, Y.: Approximate counting via correlation decay in spin systems. In: *SODA*. pp. 922–940. SIAM (2012)
19. Li, L., Lu, P., Yin, Y.: Correlation decay up to uniqueness in spin systems. In: *SODA*. pp. 67–84. SIAM (2013)
20. Lovász, L., Plummer, M.D.: *Matching theory*. AMS Chelsea Publishing, Providence, RI (2009), corrected reprint of the 1986 original [MR0859549]
21. Mézard, M., Montanari, A.: *Information, physics, and computation*. Oxford Graduate Texts, Oxford University Press, Oxford (2009)
22. Nguyen, H., Onak, K.: Constant-time approximation algorithms via local improvements. In: *FOCS*. pp. 327–336. IEEE (2008)
23. Onak, K., Ron, D., Rosen, M., Rubinfeld, R.: A near-optimal sublinear-time algorithm for approximating the minimum vertex cover size. In: *SODA*. pp. 1123–1131. SIAM (2012)
24. Parnas, M., Ron, D.: Approximating the minimum vertex cover in sublinear time and a connection to distributed algorithms. *Theoret. Comput. Sci.* 381(1-3), 183–196 (2007)
25. Rubinfeld, R.: Sublinear time algorithms. In: *International Congress of Mathematicians. Vol. III*, pp. 1095–1110. Eur. Math. Soc., Zürich (2006)
26. Sinclair, A.: *Algorithms for random generation and counting*. Progress in Theoretical Computer Science, Birkhäuser Boston Inc., Boston, MA (1993)
27. Sinclair, A., Srivastava, P., Thurley, M.: Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. In: *SODA*. pp. 941–953. SIAM (2012)
28. Sinclair, A., Srivastava, P.: Lee-yang theorems and the complexity of computing averages. In: *STOC*. pp. 625–634. ACM (2013)
29. Vadhan, S.P.: The complexity of counting in sparse, regular, and planar graphs. *SIAM Journal on Computing* 31(2), 398–427 (2002)
30. Weitz, D.: Counting independent sets up to the tree threshold. In: *STOC*. pp. 140–149. ACM (2006)
31. Welsh, D.J.A.: *Complexity: knots, colourings and counting*, London Mathematical Society Lecture Note Series, vol. 186. Cambridge University Press (1993)
32. Yao, A.: Probabilistic computations: Toward a unified measure of complexity. In: *FOCS*. pp. 222–227. IEEE (1977)
33. Yoshida, Y., Yamamoto, M., Ito, H.: An improved constant-time approximation algorithm for maximum independent sets and maximum matchings. In: *STOC*. pp. 225–234. ACM (2009)

A Proof of Proposition 3

The idea is to define a certain distribution of graphs and show that on that random input, any deterministic algorithm for approximating $p_{G,\lambda}(v)$ within an additive error ϵ requires $\underline{Q}(\epsilon, \Delta)$ queries. By Yao's Minimax Principle [32], the result follows.

We start by studying the full Δ -ary tree of height h , let it be T^h . Define $y_1 = 1$ and $y_k = (1 + \lambda\Delta y_{k-1})^{-1}$ for $k \geq 2$. Clearly y_h is equal to the value computed at the root of T^h by the recursion (3).

Lemma 9. *We have $\lim_{k \rightarrow \infty} y_k = \frac{2}{1 + \sqrt{1 + 4\lambda\Delta}}$. Define $\underline{h}(\epsilon, \Delta) = \sup\{h, |y_h - y_{h-1}| \geq \epsilon\}$. Then we have $\lim_{\Delta \rightarrow \infty} \frac{1}{\sqrt{\Delta}} \lim_{\epsilon \rightarrow 0} \frac{\underline{h}(\epsilon, \Delta)}{\log(1/\epsilon)} = \sqrt{\lambda}$.*

Proof. To study y_k , we introduce the auxiliary sequence: $f_k = f_{k-1} + \lambda\Delta f_{k-2}$ for $k \geq 3$ and $f_1 = f_2 = 1$ so that $y_k = f_k / f_{k+1}$ for $k \geq 1$. Let $\alpha = (1 + \sqrt{1 + 4\lambda\Delta})/2$ and $\beta = (1 - \sqrt{1 + 4\lambda\Delta})/2$, we have $f_k = \frac{1}{2\alpha - 1}(\alpha^k - \beta^k)$ and the first statement of the lemma follows.

A simple computation gives: $|y_k - y_{k-1}| \sim C_\Delta (D_\Delta)^{k-1}$ as $k \rightarrow \infty$, where $C_\Delta = \frac{\alpha^2 + \beta^2 - 2\alpha\beta}{\alpha^3}$ and $D_\Delta = \frac{|\beta|}{\alpha}$. So $\lim_{\epsilon \rightarrow 0} \frac{\underline{h}(\epsilon, \Delta)}{\log(1/\epsilon)} = -\frac{1}{\log D_\Delta}$. Since $\log D_\Delta \sim -\frac{1}{\sqrt{\lambda\Delta}}$ as $\Delta \rightarrow \infty$, the second statement of the lemma follows. \square

Denote L to be the set of leaves of the tree T^h . Let $\mathbf{e} = (e_u, u \in L)$ be a vector of $\{0, 1\}^L$. Define $T^h(\mathbf{e})$ to be the tree obtained from T^h in the following way: every non-leaf node in T^h remains in $T^h(\mathbf{e})$, and a leaf u in T^h remains in $T^h(\mathbf{e})$ iff $e_u = 1$. (As a result, we see that the recursion (3) is valid with $x_u(v) = e_u$ for all $u \in L$). Define the vector $\mathbf{x}^h(\mathbf{e}) = (x_u^h(\mathbf{e}), u \in T^h(\mathbf{e}))$ as defined in (3). We denote by $x^h(\mathbf{e})$ the value of the component of $\mathbf{x}^h(\mathbf{e})$ corresponding to the root of the tree. For any node u of depth $k < h$ in the tree T^h , simple monotonicity arguments show that for $h - k$ even, $y_{h-k} = x_u^h(\mathbf{0}) \leq x_u^h(\mathbf{e}) \leq x_u^h(\mathbf{1}) = y_{h-k+1}$, and for $h - k$ odd, $y_{h-k+1} = x_u^h(\mathbf{1}) \leq x_u^h(\mathbf{e}) \leq x_u^h(\mathbf{0}) = y_{h-k}$. Define $d_u^h(\mathbf{e}) = |x_u^h(\mathbf{1}) - x_u^h(\mathbf{e})|$ for all $u \in T^h$. We have:

$$\begin{aligned} d_u^h(\mathbf{e}) &= \left| \frac{1}{1 + \lambda \sum_{w \in \text{Ch}(u)} x_w^h(\mathbf{1})} - \frac{1}{1 + \lambda \sum_{w \in \text{Ch}(u)} x_w^h(\mathbf{e})} \right| \\ &\leq \frac{1}{1 + \lambda \sum_{w \in \text{Ch}(u)} x_w^h(\mathbf{1})} \cdot \frac{1}{1 + \lambda \sum_{w \in \text{Ch}(u)} x_w^h(\mathbf{e})} \cdot \lambda \sum_{w \in \text{Ch}(u)} d_w^h(\mathbf{e}) \end{aligned}$$

Notice that the first term is y_{h-k+1} by definition. For every $w \in \text{Ch}(u)$, $x_w^h(\mathbf{e}) \geq \min(y_{h-k-1}, y_{h-k})$, so the second term $\leq \max(y_{h-k}, y_{h-k+1})$. Thus we have:

$$d_u^h(\mathbf{e}) \leq y_{h-k+1} \cdot \max(y_{h-k}, y_{h-k+1}) \cdot \lambda \sum_{w \in \text{Ch}(u)} d_w^h(\mathbf{e}). \quad (8)$$

Consider u to be the root of the tree. Then $d_u^h(\mathbf{e}) = |x^h(\mathbf{e}) - x^h(\mathbf{1})|$. Iterating $h - 1$ times the inequality (8), we have:

$$|x^h(\mathbf{e}) - x^h(\mathbf{1})| \leq \sum_{u \in L} e_u \cdot \prod_{i=2}^h \lambda y_i \cdot \max(y_{i-1}, y_i).$$

Using the fact that $|y_{i-1} - y_i|/y_i$ decreases exponentially, it is easy to show that there exists a constant C , such that $\prod_{i=2}^h \max(y_{i-1}, y_i) < C \cdot \prod_{i=2}^h y_i$, for all h . So we have:

$$|x^h(\mathbf{e}) - x^h(\mathbf{1})| \leq \sum_{u \in L} e_u \cdot C \prod_{i=2}^h \lambda y_i^2 = \sum_{u \in L} e_u \cdot \frac{C \lambda^{h-1}}{f_{h+1}^2} = \sum_{u \in L} e_u \cdot O((\lambda/\alpha^2)^h). \quad (9)$$

Let $h = \underline{h}(2\epsilon, \Delta)$ and consider the uniform distribution on the set of trees $T^h(\mathbf{e})$, for every $\mathbf{e} \in \{0, 1\}^L$. Suppose \mathcal{A} is a deterministic algorithm for approximating $p_{G, \lambda}(v)$ using $o(\Delta^h)$ queries under this distribution. Then \mathcal{A} has visited at most $M = o(\Delta^h)$ positions in the h^{th} level before stopping. Let \mathbf{e}^- (resp. \mathbf{e}^+) be a vector in $\{0, 1\}^L$, where the M positions visited by \mathcal{A} have fixed values and all other positions are 0 (resp. 1). We will show that \mathcal{A} cannot provide an ϵ -approximation of the marginal for both trees $T^h(\mathbf{e}^-)$ and $T^h(\mathbf{e}^+)$. By Equation (9), $|x^h(\mathbf{e}^+) - x^h(\mathbf{1})| = M \cdot O((\lambda/\alpha^2)^h) = o((\lambda\Delta/\alpha^2)^h) = o(|\beta/\alpha|^h) = o(\epsilon)$. Similarly we can show that $|x^h(\mathbf{e}^-) - x^h(\mathbf{0})| = o(\epsilon)$. By the definition of h , $|x^h(\mathbf{1}) - x^h(\mathbf{0})| \geq 2\epsilon$, so we have $|x^h(\mathbf{e}^-) - x^h(\mathbf{e}^+)| > \epsilon$. Since \mathcal{A} is deterministic, it cannot provide an estimate which is an ϵ -approximation of $x^h(\mathbf{e}^-)$ and an ϵ -approximation of $x^h(\mathbf{e}^+)$. Thus we obtain a lower bound $\Omega(\underline{Q}(\epsilon, \Delta))$ for approximating the marginal within an error ϵ , where $\underline{Q}(\epsilon, \Delta) = \Delta^{\underline{h}(2\epsilon, \Delta)}$ satisfies Equation (6).

Notice that the trees studied above have maximal degree $\Delta + 1$ but changing Δ to $\Delta + 1$ will not affect the statement of the proposition. Note also that Kahn and Kim did study in [17] similar recursions for matchings in regular graphs but with a different perturbation at each level of the tree.

B Proof of Theorem 6

For $s \in \{0, 1\}$, let \mathcal{D}_s denote the distribution induced by setting a binary random variable to 1 with probability $p_s = (1 + (-1)^s \epsilon)/2$ (and 0 else). We define a distribution \mathcal{D} on m -bit strings as follows: (1) pick $s = 1$ with probability $1/2$; (2) draw a random string from $\{0, 1\}^m$ by choosing each bit b_i from \mathcal{D}_s independently. The following lemma is proved in [5].

Lemma 10. *Any probabilistic algorithm that can guess the value of s with a probability of error below $1/4$ requires $\Omega(1/\epsilon^2)$ bit lookups on average.*

In order to get the lower bound query complexity of $\log Z(G, \lambda)$, the idea is to create an n -node random graphs G_s depending on $s \in \{0, 1\}$ such that

$\log Z(G_0, \lambda) - \log Z(G_1, \lambda) > \rho \epsilon n$ for some constant ρ with high probability. So if there exists a $(\rho \epsilon n/3)$ -approximation algorithm for $\log Z(G, \lambda)$ using $o(1/\epsilon^2)$ queries, then we can differentiate G_0 and G_1 thus obtain the value of s with high probability using also $o(1/\epsilon^2)$ queries, which contradicts with the lower bound complexity in Lemma 10.

Consider the graph G consisting of n isolated vertices v_1, \dots, v_n . Pick $s \in \{0, 1\}$ uniformly at random and take a random $\lfloor n/2 \rfloor$ -bit string $b_1, \dots, b_{\lfloor n/2 \rfloor}$ with bits drawn from \mathcal{D}_s independently. Next, add an edge between v_{2i-1} and v_{2i} if and only if $b_i = 1$. Notice that the function $\log Z(G, \lambda)$ is additive over disjoint components, so $\log Z(G_s, \lambda) = \sum_{i=1}^{\lfloor n/2 \rfloor} x_i$, where $\{x_i\}_{1 \leq i \leq \lfloor n/2 \rfloor}$ are independent random variables, and each x_i equals $\log(1 + \lambda)$ with probability $(1 + (-1)^s \epsilon)/2$ and equals 0 otherwise. For any two graphs G_0 and G_1 derived from \mathcal{D}_0 and \mathcal{D}_1 respectively, we have $\mathbb{E}[\log Z(G_0, \lambda)] - \mathbb{E}[\log Z(G_1, \lambda)] = \log(1 + \lambda) \cdot \epsilon \lfloor n/2 \rfloor$. When $\epsilon > C/\sqrt{n}$ for some constant C large enough, we have $|\mathbb{E}[\log Z(G_0, \lambda)] - \log Z(G_0, \lambda)| < \log(1 + \lambda) \cdot \epsilon n/10$ and $|\mathbb{E}[\log Z(G_1, \lambda)] - \log Z(G_1, \lambda)| < \log(1 + \lambda) \cdot \epsilon n/10$ with high probability. Thus $\log Z(G_0, \lambda) - \log Z(G_1, \lambda) > \log(1 + \lambda) \cdot \epsilon n/5$ with high probability. Together with Lemma 10, we know that any probabilistic $(\log(1 + \lambda) \cdot \epsilon n/15)$ -approximation algorithm for $\log Z(G, \lambda)$ needs $\Omega(1/\epsilon^2)$ queries on average, thus the statement of the theorem follows.

C Permanent of expander graphs

Consider a bi-partite graph G with the node set $V = X \cup Y$, where $|X| = |Y| = n$. For every $S \subset V$, denote $N(S)$ to be the set of nodes adjacent to at least one node in S . For $\alpha > 0$, a graph is an α -*expander* if for every subset $S \subset X$ and every subset $T \subset Y$, as soon as $|S| \leq n/2$, we have $|N(S) \cap T| \geq (1 + \alpha)|S|$. Let $A = (a_{i,j})$ be the corresponding adjacency matrix of G , i.e., the rows and columns of A are indexed by nodes of X and Y respectively, and $a_{i,j} = 1$ iff (x_i, y_j) is an edge in G . Let PERM denote the permanent of A . We already know that computing the permanent of a matrix is #P-complete, even when the entries are limited to 0 and 1, so we look for an estimate of PERM.

In [8], Gamarnik and Katz provided a polynomial-time approximation algorithm for PERM with multiplication factor $(1 + \epsilon)^n$. A key tool in their analysis is the following lemma.

Lemma 11. ([8]) *Let G be a bi-partite α -expander graph of bounded degree Δ . Then for every $\lambda > 0$, we have:*

$$1 \leq \frac{Z(G, \lambda)}{\lambda^n \text{PERM}} \leq e^{O(n\lambda^{-1} \log^{-1}(1+\alpha) \log \Delta)}.$$

Based on this lemma and combined with our result in Section 3, we then obtain an algorithm to estimate $\log \text{PERM}$ within an error ϵn in sublinear time. This improves the polynomial-time algorithm in [8] which has an equivalent approximation factor.

Proposition 12. *Let G be a bi-partite α -expander graph of bounded degree Δ . Then there is an ϵn -approximation algorithm for $\log \text{PERM}$ with query complexity $\tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta/(\epsilon\alpha)})}\right)$.*

Proof. Take $\lambda = \Theta(\log \Delta/(\epsilon\alpha))$ so that $O(\lambda^{-1} \log^{-1}(1 + \alpha) \log \Delta) < \epsilon/2$. The algorithm $\text{APPROX-PARTITION-FUNCTION}(\lambda, \epsilon/2)$ in Section 3 provides an estimate \hat{Z} of $\log Z(G, \lambda)$ using $\tilde{O}\left((1/\epsilon)^{\tilde{O}(\sqrt{\Delta/(\epsilon\alpha)})}\right)$ queries, and with high constant probability, \hat{Z} is an $(\epsilon n/2)$ -approximation of $\log Z(G, \lambda)$. From Lemma 11, $\log \hat{Z} - n \log \lambda$ is an ϵn -approximation of $\log \text{PERM}$ with high constant probability. \square

D Maximum matching size

As $\lambda \rightarrow \infty$, $E(G, \lambda)$ tends to the size of a maximum matching, let it be OPT .

Lemma 13. *For any $\epsilon > 0$, taking $\lambda = e^{\frac{\Delta \log 2}{2\epsilon}}$ in $E(G, \lambda)$ leads to an ϵn -approximation of the maximum matching size.*

Proof. Since $m \leq \frac{\Delta n}{2}$ in a bounded degree graph, this lemma follows directly from Lemma 12 in [3], which proves that $E(G, \lambda) \leq \text{OPT} \leq E(G, \lambda) + \frac{m \log 2}{\log \lambda}$. \square

Applying the above value of λ to the $\text{APPROX-MATCHING-STATISTICS}(\lambda, \epsilon)$ algorithm of Section 4 leads to the following proposition.

Proposition 14. *We have an ϵn -approximation algorithm for the maximum matching size with query complexity $\tilde{O}\left((1/\epsilon)^{e^{\tilde{O}(\Delta/\epsilon)}}\right)$.*

Remark. *The query complexity of our algorithm is double exponential in Δ/ϵ , which is outperformed by [33].*

E Independent Sets

Let \mathbb{I} be the set of independent sets of G . The partition function of the system is defined by $Z_I(G, \lambda) = \sum_{I \in \mathbb{I}} \lambda^{|I|}$, and the Gibbs distribution on the space \mathbb{I} is defined by $\pi_{G, \lambda}(I) = \frac{\lambda^{|I|}}{Z_I(G, \lambda)}$. For every $v \in V$, define $p_{G, \lambda}(v) := \pi_{G, \lambda}(v \notin I) = \sum_{I \not\ni v} \pi_{G, \lambda}(I)$, where $I \not\ni v$ is an independent set not containing v .

Notice that $1/Z_I(G, \lambda)$ is exactly the probability of the empty set in Gibbs distribution, which is also equal to $\prod_{v \in V} p_{G_v, \lambda}(v)$, where G_v is defined similarly as in Section 3. Hence we have:

$$\log Z_I(G, \lambda) = \sum_{v \in V} -\log p_{G_v, \lambda}(v).$$

However it is well-known that the correlation decay implying a result similar to Lemma 1 does not hold for all values of λ in the independent set problem. Indeed Weitz in [30] gave a FPTAS for estimating $Z(G, \lambda)$ up to the critical activity λ_c for the uniqueness of the Gibbs measure on the infinite Δ -regular tree. We adapt his approach and get the analogue of Lemma 1.

Lemma 15. (Corollary 2.6 in [30]) *Let λ be such that $0 < \lambda < \lambda_c(\Delta) = \frac{\Delta^\Delta}{(\Delta-1)^{\Delta+1}}$. There exists some decaying rate δ with $\delta(l) = O(e^{-\alpha l})$ for some $\alpha > 0$, such that every graph of maximum degree $\Delta + 1$ with activity λ exhibits strong spatial mixing (see [30]) with rate δ .*

A similar approach as in Section 3 leads to the following result.

Proposition 16. *Let $0 < \lambda < \lambda_c(\Delta)$. We have an ϵn -approximation algorithm for $\log Z_I(G, \lambda)$ with query complexity polynomial in $\frac{1}{\epsilon}$ for any graph G with maximum degree $\Delta + 1$. In addition, any ϵn -approximation algorithm for $\log Z_I(G, \lambda)$ needs $\Omega(1/\epsilon^2)$ queries.*

Proof. The algorithm is the following:

1. take a sampling of $\Theta(1/\epsilon^2)$ vertices;
2. estimate $\log p_{G_v, \lambda}(v)$ with an additive error $\epsilon/2$ for every sampled vertex v ;
3. compute an estimate of $\log Z_I(G, \lambda)$ from these estimates.

Steps 1 and 3 are the same as in Section 3, so we focus on Step 2, i.e., to find two bounds p_0 and p_1 such that $p_0 \leq p_{G_v, \lambda}(v) \leq p_1$ and $\log p_1 - \log p_0 \leq \epsilon/2$. We will do this by exploring a constant-size path-tree (also called self-avoiding-walk in [30]) rooted at v .

By Lemma 15, there exists some constants $\alpha > 0$ such that $\delta(l) = O(e^{-\alpha l})$ is the strong spatial mixing rate. Let $\epsilon' = \frac{\epsilon}{2(1+\lambda)}$. Take $h = (\log \frac{1}{\epsilon'})/\alpha + O(1)$ so that $\delta(h) = \epsilon'$. For the path-tree rooted at v and truncated at level h (resp. level $h + 1$), we compute the marginal p_0 (resp. p_1) at the root v recursively by Equation (1) in [30], which takes $O(\Delta^h)$ (resp. $O(\Delta^{h+1})$) queries. Without loss of generality, we assume that $p_0 \leq p_1$. By the monotonicity argument as for Inequality (5), we have $p_0 \leq p_{G_v, \lambda}(v) \leq p_1$. By the property of strong special mixing, $p_1 - p_0 \leq \delta(h) = \epsilon'$. So $\log p_1 - \log p_0 \leq \log(1 + \frac{\epsilon'}{p_0}) \leq \log(1 + (1 + \lambda)\epsilon') \leq \epsilon/2$, where the second inequality is because $p_0 \geq \frac{1}{1+\lambda}$.

The number of queries in our computation is $O(\Delta^{h+1}/\epsilon^2)$, which is polynomial in $1/\epsilon$. The query lower bound is obtained similarly as Theorem 6. \square

Remark. *If $\Delta < 5$, we have $\lambda_c(\Delta) > 1$. Then we can approximate $Z_I(G, 1)$, which is the number of independent sets in G .*

F Tests on large graphs

We show the performance of our algorithm on the average size of a matching $E(G, 1)$ on large real-world graphs from Stanford large network dataset col-

lection². Our algorithm performs well on both small degree graphs and small average-degree graphs. The tests are based on:

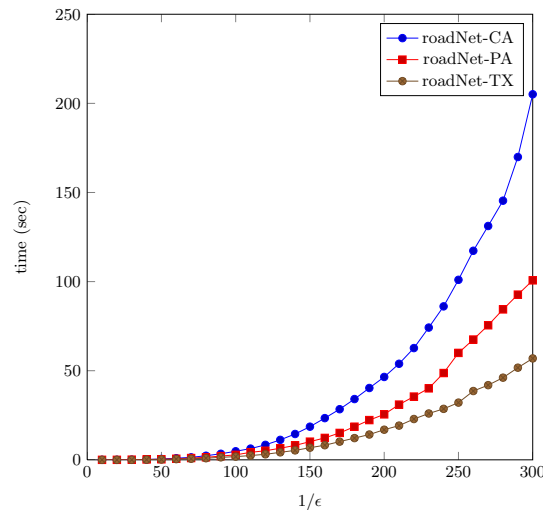
- microprocessor: intel core i5 750 (2.67 GHz, 256KB L2/core, 8MB L3)
- memory: RAM 4 Go
- compiler: g++ version 4.4.3, option -O2
- operating system: Linux Ubuntu 10.04

F.1 Small degree graphs

Consider the three road network graphs from Stanford large network dataset collection, where Δ is at most 12. Intersections and endpoints are represented by nodes and the roads connecting these intersections or road endpoints are represented by undirected edges.

- roadNet-CA : California road network with $n = 1965206$, $\Delta = 12$
- roadNet-PA : Pennsylvania road network with $n = 1088092$, $\Delta = 9$
- roadNet-TX : Texas road network with $n = 1379917$, $\Delta = 12$

We test the APPROX-MATCHING-STATISTICS algorithm in Section 4 for increasing values of $1/\epsilon$. The following figure gives the execution time of our program with respect to $1/\epsilon$, where the three curves correspond to the three graphs above.



² see <http://snap.stanford.edu/data/index.html>

F.2 Small average-degree graphs

Now we extend our algorithm to graphs with small average degree, since these graphs are of much practical interest in the real world. For a large degree node v in such a graph, it takes too many queries for the APPROX-MARGINAL algorithm to output an ϵ -approximation of marginal distribution. The solution is to compute a less accurate approximation of the marginal distribution for large degree nodes. Notice that in a graph with small average degree, the number of large degree nodes is limited. So the approximation performance in general should be good, as shown in the tests later.

In the algorithm, we take a constant number (e.g., 400 in our program) of sampled nodes, and for every node v in the sampling, we limit the time (and thus the number of queries) for estimating $p_{G,1}(v)$ and get an estimate with error possibly larger than ϵ . More precisely, we calculate $x_v^1(v), x_v^2(v), \dots$, and stop if the accumulated time of these calculations exceeds one second. Let $x_v^k(v)$ be the last estimate of $p_{G,1}(v)$ computed by the algorithm. By Inequality (5), the exact value of $p_{G,1}(v)$ is between $x_v^{k-1}(v)$ and $x_v^k(v)$. Thus $x_v^k(v)$ is an estimate of $p_{G,1}(v)$ with error at most $|x_v^k(v) - x_v^{k-1}(v)|$. Let this error be ϵ_v . Define $\bar{\epsilon}$ to be the average value of ϵ_v for all v in the sampling. From the estimates of $p_{G,1}(v)$ on sampled nodes, we get an estimate of $E(G, 1)$, which is an $(\bar{\epsilon} + 0.05)$ -approximation of $E(G, 1)$ with high probability.

Below is the performance of our algorithm on *Brightkite-edges*³ and *CA-CondMat*⁴ from Stanford large network dataset collection.

- *Brightkite-edges*: We get an estimate 32626 of $E(G, 1)$ in 12.9 seconds with $\bar{\epsilon} = 0.0016$.
- *CA-CondMat*: We get an estimate 17952 of $E(G, 1)$ in 287.5 seconds with $\bar{\epsilon} = 0.0291$.

³ This was once a location-based social networking service provider where users shared their locations by checking-in; the friendship network was collected using their public API. In this graph, $n = 58228$, average degree=3.7, $\Delta = 1134$.

⁴ This is a collaboration network of Arxiv Condensed Matter category; there is an edge if authors coauthored at least one paper. In this graph, $n = 23133$, average degree=8.1, $\Delta = 280$.