PTAS for maximum weight independent set problem with random weights in bounded degree graphs

David Gamarnik [*] David Goldberg [†] Theophane Weber [‡]

October 12, 2009

Abstract

Finding the largest independent set in a graph is a notoriously difficult NP-complete combinatorial optimization problem. Moreover, even for graphs with largest degree 3, no polynomial time approximation algorithm exists with a 1.0071-factor approximation guarantee, unless $P = NP$ [BK98].

We consider the related problem of finding the maximum weight independent set in a bounded degree graph, when the node weights are generated i.i.d. from a common distribution. Surprisingly, we discover that the problem becomes tractable for certain distributions. Specifically, we construct a randomized PTAS (Polynomial-Time Approximation Scheme) for the case of exponentially distributed weights and arbitrary graphs with degree at most 3. We extend our result to graphs with larger constant degrees but for distributions which are mixtures of exponential distributions. At the same time, we prove that no PTAS exists for computing the expected size of the maximum weight independent set in the case of exponentially distributed weights for graphs with sufficiently large constant degree, unless P=NP. Our algorithm, *cavity expansion*, is new and is based on the combination of several powerful ideas, including recent deterministic approximation algorithms for counting on graphs and local weak convergence/correlation decay methods.

1 Introduction

The problem of finding the largest independent set of a graph (the decision version of it) is a well-known NP-complete problem. Moreover, unlike some other NP-complete problems, it does not admit a constant factor approximation algorithm for general graphs: Hastad [Has96] showed that for every $0 < \delta < 1$, no $n^{1-\delta}$ approximation algorithm can exist for this problem unless $P = NP$, where n is the number of nodes. Even for the class of graphs with largest degree at most 3 no factor 1.0071 approximation algorithm can exist, under the same complexity-theoretic assumption, see Berman and Karpinski [BK98]. Similar results are established in the same paper for the cases of graphs with maximum degree 4 and 5 with slightly larger constants. Thus the problem does not admit any PTAS (Polynomial-Time Approximation Scheme) even in the least non-trivial class of degree-3 graphs. Observe that the problem can be solved to optimality trivially for graphs with largest degree 2.

The problem has also been studied in several average case settings. Karp and Sipser [KS81] constructed a simple algorithm which finds an asymptotically largest independent set in Erdős-Rényi random graphs with average degree $c \leq e$ (in fact the maximum matching problem is considered instead, but the approach works for the independent set problem as well in this regime). The largest independent set is known to be order $(2 + o(1))\frac{\log c}{c}n$ in average degree c Erdős-Rényi graphs with n nodes [JŁR00]. Similarly, in dense random graphs with n nodes, where each edge is selected with probability 1/2, the largest independent set is of size $2\log n + o(\log n)$, though the largest independent set produced by any known polynomial time algorithm is only $(1+o(1))\log n$. More recently a different average case model was considered in Gamarnik et al [GNS06]: the nodes of an Erdős-Rényi graph are equipped with random weights distributed exponentially. The limiting expression for both the expected cardinality and expected weight of the maximum weight independent set was found in the regime $c \leq 2e$. Similar results were established for r-regular graphs with girth diverging to infinity for the cases $r = 3, 4$.

In this paper we consider the following natural mixture of the worst case/average case assumptions. We consider an arbitrary graph with largest degree at most 3, where the nodes are equipped with random weights, generated i.i.d. from an exponential distribution with

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

[†] Operations Research Center and Sloan School of Management, MIT, Cambridge, MA, 02139, e-mail: dag3141@mit.edu

[‡] Operations Research Center, MIT, Cambridge, MA, 02139, e-mail: theo_w@mit.edu

parameter 1. The optimization problem is to find the largest weighted independent set. Surprisingly, we discover that this is a tractable problem - we construct a randomized PTAS, even though the unit weight version of this problem (maximum cardinality independent set) does not admit any PTAS, as was mentioned above. We extend this result to more general graphs but for distributions which are mixtures of exponential distributions. Our algorithm, which we call *Cavity Expansion*, is new and draws upon several recent ideas. The first such idea is the local weak convergence method. This is a method which takes advantage of the observation that randomness sometimes induces a long-range independence (correlation decay) in the underlying decision problem. For example, as we show in the present paper, random weights with the exponential distribution imply that whether node i belongs to the largest weighted independent set has asymptotically no correlation with whether a node j belongs to the largest weighted independent set, when the distance between i and j is large. This method was introduced by Aldous [Ald92],[Ald01],[AS03] in the context of solving the well-known $\zeta(2)$ conjecture for the random minimal assignment problem. It was further used in Gamarnik et al [GNS06] in a setting described above, namely regular graphs with large girth. It was shown that the correlation decay property indeed holds when $r = 3, 4$ and does not hold when $r > 4$. The notion of bonus was heavily used recently in the statistical physics literature under the name *cavity* [MP03] (see [RBMM04] for the independent sets setting). The local-weak convergence/cavity method has also been used extensively, but only in the setting of random graphs which have a locally-tree like structure. In order to go beyond the locally-tree like restriction on graphs another idea is needed. Such an idea was proposed recently by Weitz [Wei06] and extended in Gamarnik and Katz [GK07b],[GK07a], Bayati et al [BGK$^+$07], Jung and Shah [JS07] in the context of graph counting problems. Weitz showed that the problem of counting on a graph, and the more general problem of computing the partition function of Gibbs measures, can be reduced to the problem of counting on a related exponentially large tree of 'self-avoiding walks'. Then if the correlation decay property can be established on this tree instead, the tree can be truncated at small depth to obtain approximate inference.

In this paper we combine the correlation decay/local weak convergence approach of [GNS06] and the tree of 'self-avoiding walks' approach of Weitz [Wei06] to obtain the stated result. Our approach does not explicitly use the notion of a tree of self-avoiding walks. Instead, the simpler notions of recursive cavity approximation and expansions(refer to quantities $B^+(i,r), B^-(i,r)$ below) are used and the correlation decay property is established. The algorithm is in fact decentralized: only a local constant size neighborhood around each node is used to decide wether the given node should be a part of the maximum weight independent set.

Furthermore, we show that the setting with random weights hits a complexity-theoretic barrier just as the classical cardinality problem does. Specifically, we prove that no PTAS exists for computing the expected cardinality of the maximum weight independent set in the case of exponentially distributed weights for graphs with sufficiently large constant degree, unless P=NP.

This negative result is proven by showing that for large constant-degree graphs, the maximum weighted independent sets are dominated by independents sets with cardinality close to largest possible. Since the latter does not admit a constant factor approximation up to a $\Delta/2^{O(\sqrt{log(\Delta)})}$ multiplicative factor [Tre01], the same will apply to the former case.

Our results further highlight interesting and intriguing connection between the field of complexity of algorithms for combinatorial optimization problems and statistical physics (cavity method, long-range independence). The approach taken in this paper can be extended to other combinatorial optimization problems and is researched in larger generality in a companion paper [GDT]. It would be interesting to see what other weight distributions are amenable to the approach proposed in this paper. Furthermore, it would be interesting to see if the random weights assumptions can be substituted with deterministic weights which have some pseudo-random properties. This would move our approach even closer to the worst-case combinatorial optimization setting.

The remainder of the paper is organized as follows. The definitions, assumptions and the main results are stated in the following section. The algorithm and some preliminary results are presented in Section 3. The proof of our first main result is in Section 4. The remaining proofs are in the Appendix.

2 Model, problem formulation and the main results

Consider a simple undirected graph $\mathbb{G} = (V, E), V \triangleq [n] = \{1, 2, \ldots, n\}$. The quantity $M = M(\mathbb{G}) = \max(n, |E|)$ is called the size of the instance $\mathbb{G}$. Since we will exclusively consider graphs with degree bounded by a constant, for all practical purposes we can think of n as the size of the instance. When we say polynomial time algorithm, we mean that the running time of the algorithm is upper bounded by a polynomial in n. For every node $i \in V$, $N(i)$ denotes the set of neighbors: $N(i) = \{j \in V : (i,j) \in E\}$. Sometimes we will write $N_{\mathbb{G}}(i)$ in order to emphasize the underlying graph. $\Delta(i) \triangleq |N(i)|$ is the degree of the node i, and $\Delta = \Delta_{\mathbb{G}} \triangleq \max_i \Delta(i)$ is the degree of the graph. A set $I \subset V$ is an independent set iff $(i,j) \notin E \ \forall i,j \in I$. The quantity $\alpha = \alpha(\mathbb{G}) = \max_I |I|$ is called the independence number of the graph, where the maximization is over all independent sets. Let $I^* = I^*_{\mathbb{G}}$ denote an independent set with the largest size: $|I^*| = \alpha$. In cases where we have several such independent sets, let I^* be any such independent set. An algorithm $\mathcal{A}$ is said to be a $\rho > 1$-factor approximation algorithm for the problem of computing the size of the largest independent set α, if for every graph instance $\mathbb{G}$ it produces an independent set I in $\mathbb{G}$ s.t. $\alpha/|I| \leq \rho$. Such an algorithm is called a PTAS (Polynomial Time Approximation Scheme) if it is a $\rho = 1+\epsilon$ approximation factor algorithm for every $\epsilon > 0$ and runs in time which is polynomial in n. An algorithm is called an FPTAS (Fully Polynomial Time Approximation Scheme) if it runs in time which is polynomial in n and $1/\epsilon$. It turns out that for our setting the relevant notion is the intermediate class of algorithms, namely EPTAS. This is the class of algorithms which produces a $1+\epsilon$ approximation in time $O(n^{O(1)}g(\epsilon))$, where $g(\epsilon)$ is some function independent from n. Namely, while it is not required that the running time of the algorithm is polynomial in $1/\epsilon$, the $1/\epsilon$ quantity does not appear in the exponent of n.

In this paper we consider a variation of the problem of finding the cardinality of a largest independent set, when the nodes of the graph are equipped with random weights $W_i, i \in V$, drawn independently from a common distribution $F(t) = \mathbb{P}(W \leq t), t \geq 0$. The goal is to find an independent set I with the largest total weight $W(I) \triangleq \sum_{i \in I} W_i$. Naturally, this problem includes the problem of computing $\alpha(\mathbb{G})$ as a special case when $F(t)$ is the deterministic distribution concentrated on 1. Our main result shows that, surprisingly, the problem of finding the maximum weight independent set becomes tractable for certain distributions F, specifically when F is an exponential distribution with parameter 1, $F(t) = 1 - \exp(-t)$, and the graph has degree $\Delta \leq 3$. Let $I^* = I^*(\mathbb{G})$ be the largest weighted independent set, when it is unique, and let $W(I^*)$ be its weight. In our setting it is a random variable. Observe that I^* is indeed unique w.p. 1 when F is a continuous distribution, which is the setting in which we work.

We now state our first main result.

THEOREM 2.1. *There exists a randomized algorithm which for every $\mathbb{G} = (V,E)$ with $\Delta_{\mathbb{G}} \leq 3$ and $\epsilon > 0$ produces a (random) independent set $\mathcal{I}$ such that*

$$\mathbb{P}\left(\frac{W(I^*)}{W(\mathcal{I})} > 1+\epsilon\right) < \epsilon, \tag{2.1}$$

when the nodes' weights are independently and exponentially distributed with parameter 1. The algorithm runs in time $O\Big(n2^{O(\epsilon^{-2}\log(1/\epsilon))}\Big)$, namely it is an EPTAS.

REMARK 2.1. 1. *Our algorithm, as we shall see, uses randomization, independent of the underlying randomness of the instance. Thus the probabilistic statement (2.1) is w.r.t. two sources of randomness: randomness of weights and randomization of the algorithm.*

2. *The choice of parameter 1 in the distribution is without loss of generality, of course: any common parameter leads to the same result. Presumably, our result can be generalized to the case of node-specific parameters, but we do not pursue this direction in this paper.*

3. *Observe that the running time of the algorithm is actually* linear *in the number of nodes n. The dependence on the approximation and accuracy parameter ϵ is exponential, but this exponent does not affect n. In fact our algorithm is local in nature and, as a result, it can be run in a distributed fashion.*

We can extend the result of Theorem 2.1, albeit to the case of mixtures of exponential distributions. Let $\rho > 25$ be an arbitrary constant and let $\alpha_j = \rho^j, j \geq 1$.

THEOREM 2.2. *There exists an algorithm which for every $\mathbb{G} = (V,E)$ with $\Delta_{\mathbb{G}} \leq \Delta$ and $\epsilon > 0$ produces a (random) independent set $\mathcal{I}$ such that*

$$\mathbb{P}\left(\frac{W(I^*)}{W(\mathcal{I})} > 1+\epsilon\right) < \epsilon, \tag{2.2}$$

when the nodes' weights are distributed according to $P(W > t) = \frac{1}{\Delta}\sum_{1 \leq j \leq \Delta} \exp(-\alpha_j t)$. The algorithm runs in time $O\Big(n(\frac{1}{\epsilon})^{\Delta}\Big)$, namely it is an FPTAS.

Note that for the case of the mixture of exponential distributions described above our algorithm is in fact a F(fully)PTAS as opposed to an EPTAS for Theorem 2.1. The reason for this (rather the reason for the weaker EPTAS result) is that in order to establish the correlation decay property for the case of exponential distributions we need, for technical reasons, that the average degree is strictly less than two. Thus our algorithm is preempted by preprocessing consisting of deleting each node with small probability $\delta = \delta(\epsilon)$ independently for all nodes. This makes the correlation decay rate dependent on δ and ultimately leads to an exponential dependence on ϵ. On the other hand, for the case of a mixture of exponential distributions, we will show a correlation decay rate which holds for every degree (by adjusting the weights in the mixture). This way we achieve an FPTAS.

Finally we turn to our third and last result - the hardness of approximating $W(I^*)$ when the weights are exponentially distributed and the degree of the graph is large. We need to keep in mind that since we are dealing with instances which are random (in terms of weights) and worst-case (in terms of the underlying graph) at the same time, we need to be careful as to the notion of hardness we use. In fact we will prove a result using the standard (non-average case) notions of complexity theory. Specifically, define algorithm $\mathcal{A}$ to be a factor-ρ polynomial time approximation algorithm for computing $E[W(I^*)]$ for graphs with degree at most Δ, if given any graph with degree at most Δ, $\mathcal{A}$ produces a value $\hat{w}$ such that $\frac{1}{\rho} \leq \hat{w}/\mathbb{E}[W(I^*)] \leq \rho$ in time bounded by $n^{O(1)}$. Here the expectation is with respect to the exponential weight distribution and the constant exponent $O(1)$ is allowed to depend on Δ.

Theorem 2.3. *There exist Δ_0 and c_1^*, c_2^* such that for all $\Delta \geq \Delta_0$ the problem of computing $E[W(I^*)]$ to within a multiplicative factor $\rho = \Delta/(c_1^* \log \Delta 2^{c_2^* \sqrt{\log \Delta}})$ for graphs with degree at most Δ is NP-complete.*

The main idea of the proof is to show that the difference between the largest weighted independent set and the largest independent set measured by cardinality is diminishing in Δ. A similar proof idea was used in [LV97] for proving the hardness of approximately counting independent sets in sparse graphs.

3 Cavity expansion and the algorithm

We begin by establishing some preliminary results. The main ones are related to the notion of cavity (bonus) associated with the objective function $W(I^*)$. In this section we consider a general graph $\mathbb{G}$, whose nodes are equipped with arbitrary non-negative weights $W_i, i \in V$. Thus no probabilistic assumption on W_i is adopted yet. Let $J_{\mathbb{G}} = W(I^*)$. For every $i_1, \ldots, i_d \in V$ let $J_{\mathbb{G}}(i_1, \ldots, i_d)$ be the largest weight of an independent set I, when nodes $i_1, \ldots, i_d$ are restricted to be *not in* I. Equivalently, $J_{\mathbb{G}}(i_1, \ldots, i_d) = J_{\mathbb{G}\setminus\{i_1,\ldots,i_d\}}$, where $\mathbb{G} \setminus \{i_1, \ldots, i_d\}$ is the subgraph induced by nodes $V \setminus \{i_1, \ldots, i_d\}$. The quantity $B_{\mathbb{G}}(i_1, \ldots, i_d) = J_{\mathbb{G}} - J_{\mathbb{G}}(i_1, \ldots, i_d)$ will be called the *cavity* at the nodes $i_1, \ldots, i_d$. This quantity is also commonly called the *bonus* [Ald01],[AS03],[GNS06],[GG09].

Proposition 3.1. *Given $i \in V$, let $N(i) = \{i_1, \ldots, i_d\}$. Then*

(3.3)
$$B_{\mathbb{G}}(i) = \max\Big(0, W_i - \sum_{1\leq l\leq d} B_{\mathbb{G}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l)\Big),$$

where the inner summation evaluates to zero when $N(i) = \emptyset$. Moreover, if $W_i - \sum_{1\leq l\leq d} B_{\mathbb{G}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l) > 0$, namely $B_{\mathbb{G}}(i) > 0$, then every maximum weight independent set must contain i. Similarly if $W_i - \sum_{1\leq l\leq d} B_{\mathbb{G}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l) < 0$, implying $B_{\mathbb{G}}(i) = 0$, then every maximum weight independent set does not contain i.

Remark 3.1. *Note that the proposition leaves out the "fuzzy" case $W_i - \sum_{1\leq l\leq d} B_{\mathbb{G}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l) = 0$. This will not be a problem in our setting since, due to the continuity of the weight distribution, the probability of this event is zero. Modulo this tie, the event $B_{\mathbb{G}}(i) > 0 (B_{\mathbb{G}}(i) = 0)$ determines whether i must (must not) belong to the maximum weight independent set.*

Proof. Observe

$$J_{\mathbb{G}} = \max\Big(J_{\mathbb{G}\setminus\{i\}}, W_i + J_{\mathbb{G}\setminus\{i,i_1,\ldots,i_d\}}\Big).$$

Subtracting $J_{\mathbb{G}\setminus\{i\}}$ from both sides we obtain

$$B_{\mathbb{G}}(i) = \max\Big(0, W_i - (J_{\mathbb{G}\setminus\{i\}} - J_{\mathbb{G}\setminus\{i,i_1,\ldots,i_d\}})\Big).$$

Observe further,

$$\begin{aligned} &J_{\mathbb{G}\setminus\{i\}} - J_{\mathbb{G}\setminus\{i,i_1,\ldots,i_d\}} \\ &\quad = \sum_{1\leq l\leq d} \left(J_{\mathbb{G}\setminus\{i,i_1,\ldots,i_{l-1}\}} - J_{\mathbb{G}\setminus\{i,i_1,\ldots,i_l\}}\right) \\ &\quad = \sum_{1\leq l\leq d} B_{\mathbb{G}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l). \end{aligned}$$

The proof of the second part follows directly from the analysis above. Q.E.D.

We now construct quantities which provide bounds on the cavity B. For every induced subgraph $\mathbb{H}$ of $\mathbb{G}$, every $r = 0, 1, 2, \ldots$, and every node $i \in \mathbb{H}$, define $B^-_{\mathbb{H}}(i, r)$ recursively as follows. Let $N_{\mathbb{H}}(i) = \{i_1, \ldots, i_d\}$. Then

(3.4)
$$B^-_{\mathbb{H}}(i, r) = \begin{cases} 0, & r = 0; \\ \max\Big(0, W_i - \sum_{1\leq l\leq d} B^-_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l, r-1)\Big), & r \geq 1. \end{cases}$$

Let $B^+_{\mathbb{H}}(i, r)$ denote the similar quantity, except that we set $B^+_{\mathbb{H}}(i, 0) = W_i$ for all $\mathbb{H}, i$. By Proposition 3.1, if it was the case that $B^-_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l, r-1) = B_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l)$ for all l, then $B^-_{\mathbb{H}}(i,r) = B_{\mathbb{H}}(i,r)$. The same applies to $B^+_{\mathbb{H}}(i,r)$. However, this is generally not the case due to our "incorrect" initialization $B^-_{\mathbb{H}}(i,0) = 0$. Our goal is to show that when r is sufficiently large, $B^-_{\mathbb{H}}(i,r)$ is *approximately* correct. Showing this is the subject of the next section. Now we close this section with the complexity analysis of computing the approximate cavity $B^-_{\mathbb{G}}(i,r)$ for the initial graph $\mathbb{G}$. The algorithm is exactly computing the recursion (3.4) for a target number of iterations r and will be called *Cavity Expansion Algorithm (CA)*. Note that we assume that basic arithmetic operations , comparisons, and storage/access can be performed on the node weights (and the values derived from them by applying recursion (3.4)) in constant time.

PROPOSITION 3.2. *For every $\mathbb{G}$ with degree Δ, $i \in V(\mathbb{G})$ and $t \geq 0$, the quantities $B^-_{\mathbb{G}}(i,r), B^+_{\mathbb{G}}(i,r)$ can be computed in time $O(r\Delta^r)$.*

Proof. The proof is by induction on r. Let $\phi(r)$ be a bound on the time required to compute $B^-_{\mathbb{H}}(i,r)$ for a general subgraph $\mathbb{H}$ of $\mathbb{G}$ and node $i \in \mathbb{H}$. Given values $B^-_{\mathbb{H}'}(i', r-1)$ for all (at most) Δ $(\mathbb{H}', i')$ pairs needed to compute $B^-_{\mathbb{H}}(i,r)$ using recursion (3.4), it takes $O(\Delta)$ time to compute $B^-_{\mathbb{H}}(i,r)$ (perform all relevant look-ups , basic arithmetic operations, and storage). Since each of the (at most) Δ relevant $B^-_{\mathbb{H}'}(i', r-1)$ values can be computed in time $\phi(r-1)$ by the induction hypothesis, we find that $\phi(r) = O(\Delta + \Delta\phi(r-1))$ The assertion then follows. The analysis for $B^+_{\mathbb{H}}(i, r-1)$ is similar.

It turns out that $B^-_{\mathbb{H}}(i,r)$ and $B^+_{\mathbb{H}}(i,r)$ provide valid bounds on the true cavities $B_{\mathbb{H}}(i)$.

LEMMA 3.1. *For every even r*

$$B^-_{\mathbb{H}}(i,r) \leq B_{\mathbb{H}}(i) \leq B^+_{\mathbb{H}}(i,r),$$

and for every odd r

$$B^+_{\mathbb{H}}(i,r) \leq B_{\mathbb{H}}(i) \leq B^-_{\mathbb{H}}(i,r).$$

Proof. The proof is by induction on r. The assertion holds by definition of B^-, B^+ for $r = 0$. The induction follows from (3.3), definitions of B^-, B^+ and since the function $x \to \max(0, W - x)$ is non-increasing in x.

We now describe our algorithm $CA(r,\epsilon)$ for producing a large weighted independent set. Our algorithm runs in two stages. Fix $\epsilon > 0$. In the first stage we take an input graph $\mathbb{G} = (V, E)$ and delete every node (and incident edges) with probability $\epsilon^2/2$, independently for all nodes. We denote the resulting (random) subgraph by $\mathbb{G}(\epsilon)$. In the second stage we compute $B^-_{\mathbb{G}(\epsilon)}(i,r)$ for every node i for the graph $\mathbb{G}(\epsilon)$ for some target even number of steps r. We set $\mathcal{I}(r,\epsilon) = \{i : B^-_{\mathbb{G}(\epsilon)}(i,r) > 0\}$. Let I^*_ϵ be the largest weighted independent set of $\mathbb{G}(\epsilon)$.

LEMMA 3.2. *$\mathcal{I}(r,\epsilon)$ is an independent set.*

Proof. By Lemma 3.1, if $B^-_{\mathbb{G}(\epsilon)}(i,r) > 0$ then $B_{\mathbb{G}(\epsilon)} > 0$, and therefore $\mathcal{I} \subset I^*_\epsilon$. Thus our algorithm produces an independent set in $\mathbb{G}(\epsilon)$ and therefore in $\mathbb{G}$.

Due to Proposition 3.2, the complexity of running these stages of $CA(r,\epsilon)$ is $O(n\ r\ \Delta^r)$. We now proceed to the analysis of the Cavity Expansion Algorithm $CA(r,\epsilon)$.

4 Proof of Theorem 2.1

4.1 Correlation decay property The main bulk of the proof of Theorem 2.1 will be to show that $\mathcal{I}(r,\epsilon)$ is close to I^*_ϵ in the set-theoretic sense. We will use this to show that $W(\mathcal{I}(r,\epsilon))$ is close to $W(I^*_\epsilon)$. It will then be straightforward to show that $W(I^*_\epsilon)$ is close to $W(I^*)$, which will finally give us the desired result, theorem 5. The key step therefore consists in proving that the correlation decay property holds. It is the object of our next proposition.

First, we introduce for any arbitrary induced subgraph $\mathbb{H}$ of $\mathbb{G}(\epsilon)$, and any node i in $\mathbb{H}$, $M_{\mathbb{H}}(i) = E[e^{-B_{\mathbb{H}}(i)}], M^-_{\mathbb{H}}(i,r) = E[e^{-B^-_{\mathbb{H}}(i,r)}], M^+_{\mathbb{H}}(i,r) = E[e^{-B^+_{\mathbb{H}}(i,r)}]$.

PROPOSITION 4.1. *Let $\mathbb{G}(\epsilon) = (V_\epsilon, E_\epsilon)$ be the graph obtained from the original underlying graph as a result of the first phase of the algorithm (namely deleting every node with probability $\delta = \epsilon^2/2$ independently for all nodes). Then, for every node i in $\mathbb{G}(\epsilon)$ and every r*

$$\mathbb{P}(B_{\mathbb{G}(\epsilon)}(i) = 0, B^+_{\mathbb{G}(\epsilon)}(i, 2r) > 0) \leq 3(1-\epsilon^2/2)^{2r}, \tag{4.5}$$

and

$$\mathbb{P}(B_{\mathbb{G}(\epsilon)}(i) > 0, B^-_{\mathbb{G}(\epsilon)}(i, 2r) = 0) \leq 3(1-\epsilon^2/2)^{2r}. \tag{4.6}$$

Proof. Consider a subgraph $\mathbb{H}$ of $\mathbb{G}$, node $i \in \mathbb{H}$ with neighbors $\mathcal{N}_{\mathbb{H}}(i) = \{i_1, \ldots, i_d\}$, an integer $r \geq 1$, and suppose for now that the number of neighbors of i in $\mathbb{H}$ is at most 2.

Examine the recursion (3.3) and observe that all the randomness in the terms $B_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l)$ is derived from the subgraph $\mathbb{H}\setminus\{i, i_1, \ldots, i_{l-1}\}$, and thus W_i is independent from the vector $(B_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l), 1 \leq l \leq d)$. A similar assertion applies when we replace $B_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l)$ with $B^-_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l, r)$ and

$B^+_{\mathbb{H}\setminus\{i,i_1,\dots,i_{l-1}\}}(i_l,r)$ for every r. Using the memoryless property of the exponential distribution, letting W denote a standard exponential random variable, we obtain:

$$\begin{aligned}
&E[e^{-B_{\mathbb{H}}(i)}|\sum_{1\le l\le d} B_{\mathbb{H}\setminus\{i,i_1,\dots,i_{l-1}\}}(i_l)=x]\\
&= \mathbb{P}(W_i\le x)\mathbb{E}[e^0]+\\
&\quad E[e^{-(W_i-x)}\,|\,W_i>x]\mathbb{P}(W_i>x)\\
&= (1-\mathbb{P}(W_i>x))+E[e^{-W}]\mathbb{P}(W_i>x)\\
&= (1-\mathbb{P}(W_i>x))+(1/2)\mathbb{P}(W_i>x)\\
&= 1-(1/2)\mathbb{P}(W_i>x)\\
&= 1-(1/2)\exp(-x). \qquad (4.7)
\end{aligned}$$

It follows that

$$\mathbb{E}[e^{-B_{\mathbb{H}}(i)}]=1-\frac{1}{2}\mathbb{E}[e^{-\sum_{1\le l\le d} B_{\mathbb{H}\setminus\{i,i_1,\dots,i_{l-1}\}}(i_l)}];$$

$$\mathbb{E}[e^{-B^-_{\mathbb{H}}(i,r)}]=1-\frac{1}{2}\mathbb{E}[e^{-\sum_{1\le l\le d} B^-_{\mathbb{H}\setminus\{i,i_1,\dots,i_{l-1}\}}(i_l,r-1)}];$$

$$\mathbb{E}[e^{-B^+_{\mathbb{H}}(i,r)}]=1-\frac{1}{2}\mathbb{E}[e^{-\sum_{1\le l\le d} B^+_{\mathbb{H}\setminus\{i,i_1,\dots,i_{l-1}\}}(i_l,r-1)}].$$

By assumption i has two neighbors or less in $\mathbb{H}$, and thus $d\le 2$. For $d=0$, we have trivially $M_{\mathbb{H}}(i)=M^-_{\mathbb{H}}(i,r)=M^+_{\mathbb{H}}(i,r)$. Suppose $d=1: N_{\mathbb{H}}(i)=\{i_1\}$. Then,

$$\begin{aligned}
&M^-_{\mathbb{H}}(i,r)-M^+_{\mathbb{H}}(i,r)\\
&\quad=(1/2)\big(\mathbb{E}[e^{-B^+_{\mathbb{H}\setminus\{i\}}(i_1,r-1)}]-\mathbb{E}[e^{-B^-_{\mathbb{H}\setminus\{i\}}(i_1,r-1)}]\big)\\
&\quad=(1/2)\big(M^+_{\mathbb{H}\setminus\{i\}}(i_1,r-1)-M^-_{\mathbb{H}\setminus\{i\}}(i_1,r-1)\big). \qquad (4.8)
\end{aligned}$$

Finally, suppose $d=2$: $N(i)=\{i_1,i_2\}$. Then

$$\begin{aligned}
&M^-_{\mathbb{H}}(i,r)-M^+_{\mathbb{H}}(i,r)\\
&=\frac{1}{2}\big(\mathbb{E}[e^{-B^+_{\mathbb{H}\setminus\{i\}}(i_1,r-1)-B^+_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)}]\\
&\qquad-\mathbb{E}[e^{-B^-_{\mathbb{H}\setminus\{i\}}(i_1,r-1)-B^-_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)}]\big)\\
&=\frac{1}{2}\mathbb{E}[e^{-B^+_{\mathbb{H}\setminus\{i\}}(i_1,r-1)}\\
&\qquad\big(e^{-B^+_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)}-e^{-B^-_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)}\big)]\\
&+\frac{1}{2}\mathbb{E}[e^{-B^-_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)}\\
&\qquad\big(e^{-B^+_{\mathbb{H}\setminus\{i\}}(i_1,r-1)}-e^{-B^-_{\mathbb{H}\setminus\{i\}}(i_1,r-1)}\big)].
\end{aligned}$$

Using the non-negativity of B^-, B^+ and applying Lemma 3.1 we obtain for even r

$$\begin{aligned}
&0\le M^-_{\mathbb{H}}(i,r)-M^+_{\mathbb{H}}(i,r)\\
&\quad\le\frac{1}{2}\mathbb{E}[e^{-B^+_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)}-e^{-B^-_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)}]\\
&\qquad+\frac{1}{2}\mathbb{E}[e^{-B^+_{\mathbb{H}\setminus\{i\}}(i_1,r-1)}-e^{-B^-_{\mathbb{H}\setminus\{i\}}(i_1,r-1)}]\\
&\quad=\frac{1}{2}\big(M^+_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)-M^-_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)\big)\\
&\qquad+\frac{1}{2}\big(M^+_{\mathbb{H}\setminus\{i\}}(i_1,r-1)-M^-_{\mathbb{H}\setminus\{i\}}(i_1,r-1)\big); \qquad (4.9)
\end{aligned}$$

and for odd r

$$\begin{aligned}
&0\le M^+_{\mathbb{H}}(i,r)-M^-_{\mathbb{H}}(i,r)\\
&\quad\le\frac{1}{2}\big(M^-_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)-M^+_{\mathbb{H}\setminus\{i,i_1\}}(i_2,r-1)\big)\\
&\qquad+\frac{1}{2}\big(M^-_{\mathbb{H}\setminus\{i\}}(i_1,r-1)]-M^+_{\mathbb{H}\setminus\{i\}}(i_1,r-1)\big). \qquad (4.10)
\end{aligned}$$

Summarizing the three cases we conclude that

$$\begin{aligned}
&|M^+_{\mathbb{H}}(i,r)-M^-_{\mathbb{H}}(i,r)|\le \qquad (4.11)\\
&\quad\frac{d}{2}\max_{\mathbb{H}',j}|M^+_{\mathbb{H}'}(j,r-1)-M^-_{\mathbb{H}'}(j,r-1)|;
\end{aligned}$$

where the maximum is over all induced subgraphs $\mathbb{H}'$ of $\mathbb{G}$ and nodes $j\in\mathbb{H}'$ with degree at most 2 in $\mathbb{H}'$. The reason for this is that in equations (4.8),(4.9), and (4.10); the moments $M^+_{\mathbb{H}'}(j,r-1)$ appearing on the right hand side are always computed for a node j which has lost at least one of its neighbors (namely, i) in graph $\mathbb{H}'$. Since the degree of j was at most 3 in $\mathbb{G}$ and at least one neighbor is removed, j has at most two neighbors in $\mathbb{H}'$. By considering $\mathbb{H}\cap\mathbb{G}(\epsilon)$ in all previous equations, equation (4.11) implies

$$\begin{aligned}
&|M^+_{\mathbb{H}\cap\mathbb{G}(\epsilon)}(i,r)-M^-_{\mathbb{H}\cap\mathbb{G}(\epsilon)}(i,r)|\le \qquad (4.12)\\
&\quad\frac{d(\epsilon)}{2}\max_{\mathbb{H}',j}\Big|M^+_{\mathbb{H}'\cap\mathbb{G}(\epsilon)}(j,r-1)-M^-_{\mathbb{H}\cap\mathbb{G}(\epsilon)}(j,r-1)\Big|;
\end{aligned}$$

where $d(\epsilon)$ denotes the number of neighbors of i in $\mathbb{H}\cap\mathbb{G}(\epsilon)$ and again the maximum is over all induced subgraphs $\mathbb{H}'$ of $\mathbb{G}$ and nodes $j\in\mathbb{H}'$ with degree at most 2 in $\mathbb{H}'$. By definition of $\mathbb{G}(\epsilon)$, $d(\epsilon)$ is a binomial random variables with d trials and probability of success $(1-\epsilon^2/2)$, where d is the deree of i in $\mathbb{H}$. Since $d\le 2$, $\mathbb{E}[d(\epsilon)]\le 2(1-\epsilon^2/2)$. Moreover, this randomness is independent from the randomness of the random weights of $\mathbb{H}$. Therefore,

$$\begin{aligned}
&\mathbb{E}[|M^+_{\mathbb{H}\cap\mathbb{G}(\epsilon)}(i,r)-M^-_{\mathbb{H}\cap\mathbb{G}(\epsilon)}(i,r)|]\le \qquad (4.13)\\
&(1-\frac{\epsilon^2}{2})\max_{\mathbb{H}',j}E\Big|M^+_{\mathbb{H}'\cap\mathbb{G}(\epsilon)}(j,r-1)-M^-_{\mathbb{H}'\cap\mathbb{G}(\epsilon)}(j,r-1)\Big|
\end{aligned}$$

where the external expectation is w.r.t. randomness of the first phase of the algorithm (deleted nodes), and again the maximum is over all induced subgraphs $\mathbb{H}'$ of $\mathbb{G}$ and nodes $j \in \mathbb{H}'$ with degree at most 2 in $\mathbb{H}'$. Let e_{r-1} denote the right-hand side of Ineq. (4.13). By taking the max of the left-hand side of (4.13) over all $(\mathbb{H}, j)$ s.t. $\mathbb{H}$ is a subgraph of $\mathbb{G}$ and j is a node in $\mathbb{H}$ with degree less than or equal to 2 in $\mathbb{H}$, we obtain the inequality $e_r \leq (1 - \epsilon^2/2)e_{r-1}$. Iterating on r and using the fact that for any non-negative r.v. X, $0 \leq \mathbb{E}[e^{-X}] \leq 1$, this implies that $e_r \leq (1 - \epsilon^2/2)^r$ for all $r \geq 0$ (since $0 \leq e_0 \leq 1$). Finally, it is easy to show using the same techniques that equation (4.11) holds for a node i with degree $d = 3$ in $\mathbb{H}$ as well. This implies that for an arbitrary node i in $\mathbb{G}(\epsilon)$,

$$E[|M^+_{\mathbb{G}(\epsilon)}(i, r) - M^-_{\mathbb{G}(\epsilon)}(i, r)|] \leq \frac{3}{2}(1 - \epsilon^2/2)^r.$$

Applying Lemma 3.1, we conclude for every $r \geq 0$

$$0 \leq E[e^{-B^-_{\mathbb{G}(\epsilon)}(i,2r)} - e^{-B^+_{\mathbb{G}(\epsilon)}(i,2r)}] \leq \frac{3}{2}(1 - \epsilon^2/2)^{2r}.$$

Recalling (4.7) we have

$$\begin{aligned}
&\mathbb{E}[\exp(-B_{\mathbb{G}(\epsilon)}(i))] \\
&\quad = 1 - (1/2)\mathbb{P}(W > \sum_{1 \leq l \leq d} B_{\mathbb{G}(\epsilon)\setminus\{i, i_1, \ldots, i_{l-1}\}}(i_l)) \\
&\quad = 1 - (1/2)\mathbb{P}(B_{\mathbb{G}(\epsilon)}(i) > 0),
\end{aligned}$$

Similar expressions are valid for $B^-_{\mathbb{G}(\epsilon)}(i, r)), B^+_{\mathbb{G}(\epsilon)}(i, r))$. Applying Lemma 3.1, we obtain

$$\begin{aligned}
&0 \leq \mathbb{P}(B^-_{\mathbb{G}(\epsilon)}(i, 2r) = 0) - \mathbb{P}(B^+_{\mathbb{G}(\epsilon)}(i, 2r) = 0) \\
&\quad \leq 3(1 - \epsilon^2/2)^{2r}.
\end{aligned}$$

Again applying Lemma 3.1, we obtain

$$\begin{aligned}
&\mathbb{P}(B_{\mathbb{G}(\epsilon)}(i) = 0, B^+_{\mathbb{G}(\epsilon)}(i, 2r) > 0) \\
&\quad \leq \mathbb{P}(B^-_{\mathbb{G}(\epsilon)}(i, 2r) = 0, B^+_{\mathbb{G}(\epsilon)}(i, 2r) > 0) \\
&\quad = \mathbb{P}(B^-_{\mathbb{G}(\epsilon)}(i, 2r) = 0) \\
&\qquad - \mathbb{P}(B^-_{\mathbb{G}(\epsilon)}(i, 2r) = 0, B^+_{\mathbb{G}(\epsilon)}(i, 2r) = 0) \\
&\quad = \mathbb{P}(B^-_{\mathbb{G}(\epsilon)}(i, 2r) = 0) - \mathbb{P}(B^+_{\mathbb{G}(\epsilon)}(i, 2r) = 0) \\
&\quad \leq 3(1 - \epsilon^2/2)^{2r};
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(B_{\mathbb{G}(\epsilon)}(i) > 0, B^-_{\mathbb{G}(\epsilon)}(i, 2r) = 0) \\
&\quad \leq \mathbb{P}(B^-_{\mathbb{G}(\epsilon)}(i, 2r) = 0, B^+_{\mathbb{G}(\epsilon)}(i, 2r) > 0) \\
&\quad \leq 3(1 - \epsilon^2/2)^{2r}.
\end{aligned}$$

This completes the proof of the proposition.

4.2 Concentration argument We can now complete the proof of Theorem 2.1. We need to bound $|W(I^*) - W(I^*_\epsilon)|$ and $W(I^*_\epsilon \setminus \mathcal{I}(r, \epsilon))$ and show that both quantities are small.

Let ΔV_ϵ be the set of nodes in $\mathbb{G}$ which are not in $\mathbb{G}(\epsilon)$. Trivially, $|W(I^*) - W(I^*_\epsilon)| \leq W(\Delta V_\epsilon)$. We have $E[|\Delta V_\epsilon|] = \frac{\epsilon^2}{2}n$, and since the nodes were deleted irrespectively of their weights, then $E[W(\Delta V_\epsilon)] = \frac{\epsilon^2}{2}n$.

To analyze $W(I^*_\epsilon \setminus \mathcal{I}(r, \epsilon))$, observe that by (the second part of) Proposition 4.1, for every node $i, \mathbb{P}(i \in I^*_\epsilon \setminus \mathcal{I}(r, \epsilon)) \leq 3(1 - \epsilon^2/2)^r \triangleq \delta_1$. Thus $E|I^*_\epsilon \setminus \mathcal{I}(r, \epsilon)| \leq \delta_1 n$. In order to obtain a bound on $W(I^*_\epsilon \setminus \mathcal{I}(r, \epsilon))$ we derive a crude bound on the largest weight of a subset with cardinality $\delta_1 n$. Fix a constant C and consider the set V_C of all nodes in $\mathbb{G}(\epsilon)$ with weight greater than C. We have $E[W(V_C)] \leq (C + \mathbb{E}[W - C|W > C])e^{-C}n = (C + 1)e^{-C}n$. Each node in $V \setminus V_C$ has weight at most C. Therefore,

$$\begin{aligned}
&\mathbb{E}[W(I^*_\epsilon \setminus \mathcal{I}(r, \epsilon))] \\
&\quad \leq E[W\Big(\big((I^*_\epsilon \setminus \mathcal{I}(r, \epsilon)\big) \cap V^c_C) \cup V_C\Big)] \\
&\quad \leq CE[|I^*_\epsilon \setminus \mathcal{I}(r, \epsilon)|] + E[V_C] \\
&\quad \leq C\delta_1 n + (C + 1)e^{-C}n.
\end{aligned}$$

We conclude

$$\mathbb{E}[|W(I^*) - W(\mathcal{I}(r, \epsilon))|] \leq \frac{\epsilon^2}{2}n + C\delta_1 n + (C + 1)e^{-C}n$$

Now we obtain a lower bound on $W(I^*)$. Consider the standard greedy algorithm for generating an independent set: take arbitrary node, remove neighbors, and repeat. It is well known and simple to see that this algorithm produces an independent set with cardinality at least $n/4$, since the largest degree is at most 3. Since I^* does not depend on the weights W, $E[W(I^*)] = I^* \geq n/4$. Also, $\mathrm{Var}(W(I^*)) = I^* \leq n$. By Chebyshev's inequality,

$$\mathbb{P}(W(I^*) < n/8) \leq \frac{n}{(n/8)^2} = 64/n.$$

We now summarize the results.

$$\begin{aligned}
&\mathbb{P}(\frac{W(\mathcal{I}(r, \epsilon))}{W(I^*)} \leq 1 - \epsilon') \\
&\quad \leq \mathbb{P}(\frac{W(\mathcal{I}(r, \epsilon))}{W(I^*)} \leq 1 - \epsilon', W(I^*) \geq \frac{n}{8}) \\
&\qquad + \mathbb{P}(W(I^*) < n/8) \\
&\quad \leq \mathbb{P}(\frac{|W(I^*) - W(\mathcal{I}(r, \epsilon))|}{W(I^*)} \geq \epsilon', W(I^*) \geq \frac{n}{8}) + \frac{64}{n} \\
&\quad \leq \mathbb{P}(\frac{|W(I^*) - W(\mathcal{I}(r, \epsilon))|}{\frac{n}{8}} \geq \epsilon') + \frac{64}{n} \\
&\quad \leq \frac{\epsilon^2/2 + 3C(1 - \epsilon^2/2)^r + (C + 1)\exp(-C)}{\epsilon'/8} + \frac{64}{n},
\end{aligned}$$

where we have used Markov's inequality in the last step and $\delta_1 = 3(1-\epsilon^2/2)^r$. Thus it suffices to arrange ϵ, C so that the first ratio is at most $\epsilon'/2$ and assuming, without the loss of generality that $n \geq 128/\epsilon'$, we will obtain that the sum is at most ϵ'. It is a simple exercise to show that by taking $\epsilon = \frac{\epsilon'}{64}, r = O(\log(1/\epsilon)/\epsilon^2)$ and $C = O(\log(1/\epsilon))$, we obtain the required result. This completes the proof of Theorem 2.1.

References

[Ald92] D. Aldous, *Asymptotics in the random assignment problem*, Probability Theory and Related Fields **93** (1992), no. 4, 507–534.

[Ald01] ______, *The $\zeta(2)$ limit in the random assignment problem*, Random Structures and Algorithms **18** (2001), 381–418.

[AS03] D. Aldous and J. M. Steele, *The objective method: Probabilistic combinatorial optimization and local weak convergence*, Discrete Combinatorial Probability, H. Kesten Ed., Springer-Verlag, 2003.

[BGK+07] M. Bayati, D. Gamarnik, D. Katz, C. Nair, and P. Tetali, *Simple deterministic approximation algorithms for counting matchings*, Proc. of the 39th annual ACM Symposium on Theory of computing, 2007, pp. 122–127.

[BK98] P. Berman and M. Karpinski, *On some tighter inapproximability results*, Tech. report, Electronic Colloquium on Computational Complexity, 1998.

[GDT] D. Goldberg, D.Gamarnik, and T.Weber, *Correlation decay in random decision networks: Sufficient conditions for efficient decentralized optimization*, In preparation.

[GG09] D. Gamarnik and D.A. Goldberg, *Randomized greedy algorithms for independent sets and matchings in regular graphs: Exact results and finite girth corrections*, Combinatorics, Probability and Computing (2009), 1–25.

[GK07a] D. Gamarnik and D. Katz, *A Deterministic Approximation Algorithm for Computing a Permanent of a 0, 1 matrix*, Arxiv preprint math.CO/0702039 (2007).

[GK07b] ______, *Correlation decay and deterministic FPTAS for counting list-colorings of a graph*, Proc. of the 18th annual ACM-SIAM Symposium On Discrete Algorithms, 2007, pp. 1245–1254.

[GNS06] D. Gamarnik, T. Nowicki, and G. Swirscsz, *Maximum weight independent sets and matchings in sparse random graphs. Exact results using the local weak convergence method*, Random Structures and Algorithms **28** (2006), no. 1, 76–106.

[Has96] J. Hastad, *Clique is hard to approximate within n*, Proceedings of the 37th Annual Symposium on Foundations of Computer Science, 1996, pp. 627–636.

[JŁR00] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, John Wiley New York, 2000.

[JS07] K. Jung and D. Shah, *Inference in Binary Pair-wise Markov Random Fields through Self-Avoiding Walks*, Proceedings of Allerton Conference on Computation, Communication and Control, 2007, p. 8.

[KS81] R. Karp and M. Sipser, *Maximum matching in sparse random graphs*, Proceedings of the 22nd Annual Symposium on the Foundations of Computer Science, 1981, pp. 364–375.

[LV97] M. Luby and E. Vigoda, *Approximately counting up to four*, Proceedings of the 29d Annual ACM Symposium on the Theory of Computing (1997), 682–687.

[MP03] M. Mezard and G. Parisi, *The cavity method at zero temperature*, Journal of Statistical Physics **111** (2003), no. 1-2, 1–34.

[RBMM04] O. Rivoire, G. Biroli, O. C. Martin, and M. Mezard, *Glass models on Bethe lattices*, Eur. Phys. J. B **37** (2004), 55–78.

[Tre01] L. Trevisan, *Non-approximability results for optimization problems on bounded degree instances*, ACM Symposium on Theory of Computing, 2001.

[Wei06] D. Weitz, *Counting independent sets up to the tree threshold*, Proc. 38th Ann. Symposium on the Theory of Computing, 2006.

Appendix

In this section we present proofs of Theorems 2.2,2.3.

Proof. [Proof of Theorem 2.2] The mixture of Δ exponential distributions with rates $\alpha_j, 1 \leq j \leq \Delta$ and equal weights $1/\Delta$ can be viewed as first randomly generating a rate α with the probability law $\mathbb{P}(\alpha = \alpha_j) = 1/\Delta$ and then randomly generating an exponentially distributed random variable with rate α_j, conditional on the rate being α_j.

For every subgraph $\mathbb{H}$ of $\mathbb{G}$, node i in $\mathbb{H}$ and $j = 1, \ldots, \Delta$, define $M^j_{\mathbb{H}}(i) = E[\exp(-\alpha_j\ B_{\mathbb{H}}(i))]$, $M^{-,j}_{\mathbb{H}}(i,r) = E[\exp(-\alpha_j\ B^-_{\mathbb{H}}(i,r))]$ and $M^{+,j}_{\mathbb{H}}(i,r) = E[\exp(-\alpha_j B^+_{\mathbb{H}}(i,r)]$, where $B_{\mathbb{H}}(i), B^+_{\mathbb{H}}(i,r)$ and $B^-_{\mathbb{H}}(i,r)$ are defined as in Section 3.

LEMMA 4.1. *Fix any subgraph $\mathbb{H}$, node $i \in \mathbb{H}$ with $N_{\mathbb{H}}(i) = \{i_1, \ldots, i_d\}$. Then*

$$\mathbb{E}[\exp(-\alpha_j B_{\mathbb{H}}(i))] = 1 - \sum_{1\leq k\leq\Delta} \frac{\alpha_j}{\alpha_j+\alpha_k} \mathbb{E}[e^{-\sum_{1\leq l\leq d} \alpha_k B_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l)}];$$

$$\mathbb{E}[e^{-\alpha_j B^+_{\mathbb{H}}(i,r)}] = 1 - \sum_{1\leq k\leq\Delta} \frac{\alpha_j}{\alpha_j+\alpha_k} \mathbb{E}[e^{-\sum_{1\leq l\leq d} \alpha_k B^+_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1)}];$$

$$E[\exp(-\alpha_j B^-_{\mathbb{H}}(i,r))] = 1 - \sum_{1\leq k\leq\Delta} \frac{\alpha_j}{\alpha_j+\alpha_k} \mathbb{E}[e^{-\sum_{1\leq l\leq d} \alpha_k B^-_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1)}].$$

Proof. Let $\alpha(i)$ be the random rate associated with node i. Namely, $\mathbb{P}(\alpha(i) = \alpha_j) = 1/\Delta$. We condition on the event $\sum_{1\le l\le d} B_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l) = x$. As $B_{\mathbb{H}}(i) = \max(0, W_i - x)$, we obtain:

$$
\begin{aligned}
&E[-\alpha_j B_{\mathbb{H}}(i)|x]\\
&= \frac{1}{\Delta}\sum_k E[-\alpha_j B_{\mathbb{H}}(i)|x, \alpha(i)=\alpha_k]\\
&= \frac{1}{\Delta}\sum_k \Big(\mathbb{P}(W_i \le x|\alpha(i)=\alpha_k)\\
&+\mathbb{P}(W_i > x|\alpha(i)=\alpha_k)E[e^{-\alpha_j(W_i-x)}|W_i>x,\alpha(i)=\alpha_k]\Big)\\
&= \frac{1}{\Delta}\sum_k \left(1 - e^{-\alpha_k x} + e^{-\alpha_k x}\frac{\alpha_k}{\alpha_j+\alpha_k}\right)\\
&= 1 - \frac{1}{\Delta}\sum_k \frac{\alpha_j}{\alpha_j+\alpha_k}e^{-\alpha_k x}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&E[-\alpha_j B_{\mathbb{H}}(i)] =\\
&1 - \frac{1}{\Delta}\sum_k \frac{\alpha_j}{\alpha_j+\alpha_k}E[e^{-\sum_{1\le l\le d}\alpha_k B_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l)}].
\end{aligned}
$$

The other equalities follow identically.

By taking differences, we obtain

$$
\begin{aligned}
&M^{-,j}_{\mathbb{H}}(i,r) - M^{+,j}_{\mathbb{H}}(i,r)\\
&= \frac{1}{\Delta}\sum_k \frac{\alpha_j}{\alpha_j+\alpha_k}\Bigg(E[\prod_{1\le l\le d} e^{-\alpha_k B^{+}_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1)}]\\
&\quad -E[\prod_{1\le l\le d} e^{-\alpha_k B^{-}_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1)}]\Bigg).
\end{aligned}
$$

We now prove by induction that

$$
\Big|\prod_{1\le l\le d} x_l - \prod_{1\le l\le d} y_l\Big| \le \sum_{1\le l\le d}|x_l - y_l|, \tag{4.14}
$$

when $\max_l |x_l|, |y_l| < 1$. The base case $d = 1$ is trivial. Now, suppose Ineq. 4.14 holds for $d \le k$. Note that

$$
\begin{aligned}
&\Big|\prod_{1\le l\le k+1} x_l - \prod_{1\le l\le k+1} y_l\Big|\\
&= \Big|x_{k+1}\prod_{1\le l\le k} x_l - y_{k+1}\prod_{1\le l\le k} y_l\Big|\\
&= \frac{1}{2}\Big|(x_{k+1}-y_{k+1})(\prod_{1\le l\le k} x_l + \prod_{1\le l\le k} y_l)\\
&\quad +(x_{k+1}+y_{k+1})(\prod_{1\le l\le k} x_l - \prod_{1\le l\le k} y_l)\Big|\\
&\le |x_{k+1}-y_{k+1}| + |\prod_{1\le l\le k} x_l - \prod_{1\le l\le k} y_l|\\
&\quad \text{since} \frac{1}{2}|\prod_{1\le l\le k} x_l + \prod_{1\le l\le k} y_l| \le 1, \frac{1}{2}|x_{k+1}+y_{k+1}| \le 1\\
&\quad \le \sum_{l=1}^{k+1}|x_l - y_l| \text{ by the induction hyp.}
\end{aligned}
$$

By applying Ineq. (4.14) with $x_l = e^{-\alpha_k B^{+}_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1)}$ and $y_l = e^{-\alpha_k B^{-}_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1)}$, we obtain

$$
\begin{aligned}
&|M^{-,j}_{\mathbb{H}}(i,r) - M^{+,j}_{\mathbb{H}}(i,r)| \le\\
&\frac{1}{\Delta}\sum_{1\le k\le\Delta}\frac{\alpha_j}{\alpha_j+\alpha_k}\sum_{1\le l\le d}\Big(\\
&\quad |M^{-,k}_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1) - M^{+,k}_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1)|\Big).
\end{aligned}
$$

This implies

$$
\begin{aligned}
&|M^{-,j}_{\mathbb{H}}(i,r) - M^{+,j}_{\mathbb{H}}(i,r)| \le\\
&\frac{d}{\Delta}\sum_{1\le k\le\Delta}\frac{\alpha_j}{\alpha_j+\alpha_k}\max_{1\le l\le d}\Big(\\
&\quad |M^{-,k}_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1) - M^{+,k}_{\mathbb{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l,r-1)|\Big).
\end{aligned} \tag{4.15}
$$

For any $t \ge 0$ and j, let

$$
e_{r,j} \triangleq \sup_{\mathbb{H}\subset\mathbb{G}, i\in\mathbb{H}} |M^{-,j}_{\mathbb{H}}(i,r) - M^{+,j}_{\mathbb{H}}(i,r)|. \tag{4.16}
$$

By taking the maximum on the right and left hand side of Ineq. 4.15 successively (as in the exponential case), we find that

$$
e_{r,j} \le \frac{d}{\Delta}\sum_{1\le k\le\Delta}\frac{\alpha_j}{\alpha_j+\alpha_k}e_{r-1,k}
$$

For any $t \ge 0$, denote by $\mathbf{e_r}$ the vector $(e_{r,1},\ldots,e_{r,\Delta})$. Denote by $\mathbf{M}$ the $\Delta\times\Delta$ matrix such that for all (j,k), $M_{j,k} = \frac{d}{\Delta}\frac{\alpha_j}{\alpha_j+\alpha_k}$. We finally obtain

$$
\mathbf{e_r} \le M\mathbf{e_{r-1}}.
$$

Therefore, if M^r converges to zero exponentially fast in each coordinate, then also $\mathbf{e_r}$ converges exponentially fast to 0. Following the same steps as the proof of theorem 2.1, this will imply that for each node, the error of a decision made by $\mathrm{CE}(r,0)$ is exponentially small in r . Note that $\frac{d}{\Delta} \le 1$. Recall that $\alpha_j = \rho^j$. Therefore, for each j,k, we have $M_{j,k} \le \frac{\rho^j}{\rho^j+\rho^k}$. By a simple bounding argument , if $\rho > 1$ it suffices to show that M^t_Δ converges to zero exponentially fast, where M_Δ is the matrix defined by $M_{j,j} = 1/2, M_{j,k} = 1, j > k$ and $M_{j,k} = (1/\rho)^{k-j}, k > j$, for all $1 \le j,k \le \Delta$.

Theorem 2.2 thus follows from the following lemma:

Lemma 4.2. *Under the condition $\rho > 25$, there exists $\delta = \delta(\rho) < 1$ such that the absolute value of every entry of M^t_Δ is at most $\delta^t(\rho)$.*

Proof. Let $\epsilon = 1/\rho$. Let M' denote the transpose of M_Δ. Since elements of M_Δ are non-negative, it suffices to exhibit a strictly positive vector $x = x(\rho)$ and $0 < \theta = \theta(\rho) < 1$ such that $M'x \le \theta x$. Let x be the vector defined by $x_k = \epsilon^{k/2}, 1 \le k \le \Delta$. We show that for any j,

$$(M'x)_j \le (1/2 + 2\frac{\sqrt{\epsilon}}{1-\sqrt{\epsilon}})x_j.$$

It is easy to verify that when $\rho > 25$, that is $\epsilon < 1/25$, $(1/2 + 2\frac{\sqrt{\epsilon}}{1-\sqrt{\epsilon}}) < 1$, completing the proof. Fix $1 \le j \le \Delta$. Then,

$$\begin{aligned}(M'x)_j &= \sum_{1\le k\le j-1} M_{k,j}\, x_k + 1/2 x_j + \sum_{j+1\le k\le\Delta} M_{k,j}\, x_k \\ &= \sum_{1\le k\le j-1} \epsilon^{j-k}\epsilon^{k/2} + 1/2\epsilon^{j/2} + \sum_{j+1\le k\le\Delta} \epsilon^{k/2}.\end{aligned}$$

Since $x_j = \epsilon^{j/2}$, we have

$$\begin{aligned}\frac{(M'x)_j}{x_j} &\le \sum_{1\le k\le j-1} \epsilon^{(j-k)/2} + 1/2 + \sum_{j+1\le k\le\Delta} \epsilon^{(k-j)/2} \\ &= \frac{1}{2} + \sum_{1\le k\le j-1} \epsilon^{\frac{k}{2}} + \sum_{1\le k\le \Delta-j} \epsilon^{\frac{k}{2}} \\ &\le \frac{1}{2} + \frac{2\epsilon^{\frac{1}{2}}}{1-\epsilon^{\frac{1}{2}}}.\end{aligned}$$

This completes the proofs of both the lemma and Theorem 2.2.

Proof of Theorem 2.3 The main idea of the proof is to show that the difference between the largest weighted independent set and the largest independent set measured by cardinality is diminishing in Δ. A similar proof idea was used in [LV97] for proving the hardness of approximately counting independent sets in sparse graphs.

Given a graph $\mathbb{G}$ with degree bounded by Δ, let I^M denote (any) maximum cardinality independent set, and let I^* denote the unique maximum weight independent set corresponding to i.i.d. weights with $\exp(1)$ distribution. We make use of the following result due to Trevisan [Tre01].

Theorem 4.1. *There exist Δ_0 and c^* such that for all $\Delta \ge \Delta_0$ the problem of approximating the largest independent set in graphs with degree at most Δ to within a factor $\rho = \Delta/2^{c^*\sqrt{\log\Delta}}$ is NP-complete.*

Our main technical result is the following proposition. It states that an independent set of large weight is a large independent up to a factor which grows as the logarithm of the maximum degree of the graph.

Proposition 4.2. *Suppose $\Delta \ge 2$. For every graph $\mathbb{G}$ with n large enough, we have:*

$$1 \le \frac{|I^M|}{E[|I^*|]} \le 20\log\Delta.$$

This in combination with Theorem 4.1 leads to the desired result.

Proof. Let $W(1) < W(2) < \cdots < W(n)$ be the ordered weights associated with our graph $\mathbb{G}$. Fix δ, which we later will choose to be $1/(10\log\Delta)$. let $m = \lceil\delta|I^M|\rceil$. Observe that the event $|I^*| < \delta|I^M|$ implies that $W(I^M) \le W(I^*) \le \sum_{j=n-m+1}^n W(j)$. The exponential distribution implies $E[W(j)] = H(n) - H(n-j)$, where $H(k)$ is the harmonic sum $\sum_{1\le i\le k} 1/i$. Thus

$$\begin{aligned}\sum_{j=n-m+1}^{n} E[W(j)] &= \sum_{n-m+1\le j\le n} (H(n) - H(n-j)) \\ &= mH(n) - \sum_{j\le m-1} H(j).\end{aligned}$$

We use the bound $\log(k) \le H(k) - \gamma \le \log(k) + 1$, where γ is Euler's constant. Then

$$\begin{aligned}&\sum_{j=n-m+1}^{n} E[W(j)] \\ &\le m(H(n)-\gamma) + \log(m) - \sum_{1\le j\le m}\log(j) \\ &\le m(H(n)-\gamma) + \log(m) - \int_1^m \log(t)dt \\ &\le m\log(n) + m + \log(m) - m\log(m) + m \\ &\le (\delta|I^M|+1)(\log\frac{n}{\delta|I^M|} + 2 + \log(m)/m) \\ &\le \delta|I^M|(\log\frac{\Delta+1}{\delta} + 3) + (\log\frac{\Delta+1}{\delta} + 3),\end{aligned}$$

where the bound $|I^M| \geq n/(\Delta+1)$ (obtained by using the greedy algorithm, see Section 4.2) is used. For $\delta = 1/(10\log\Delta)$, we can check that $\delta(\log\frac{\Delta+1}{\delta}+3) < .89$, implying that for all sufficiently large n,

$$\begin{aligned}
&\sum_{j=n-m+1}^{n} E[W(j)] - E[W(I^M)] \\
&\leq |I^M|(\delta(\log\frac{\Delta+1}{\delta}+3)-1) + o(n) \\
&< -0.11|I^M| + o(n) \\
&\leq -\frac{0.1}{\Delta+1}n.
\end{aligned}$$

Our next step is to show that $\sum_{n-m\leq j\leq n} W(j)$ is concentrated around its mean. It is known that for any i, j, n, $W(i)$ and $W(j)$ are positively correlated. As a result,

$$\begin{aligned}
&\mathrm{Var}\left[\sum_{j=n-m+1}^{n} W(j)\right] \\
&= \sum_{j=n-m+1}^{n} \sum_{i=n-m+1}^{n} \mathrm{Cov}(W(i), W(j)) \\
&\leq \sum_{j=1}^{n} \sum_{i=1}^{n} \mathrm{Cov}(W(i), W(j)) \\
&= Var[\sum_i W(i)] \\
&= n.
\end{aligned}$$

On the other hand, $\mathrm{Var}[W(I^M)] = |I^M| \leq n$. Using the fact that for two (possibly dependent) r.v. X, Y we have $Var[X-Y] \leq Var[X-Y] + Var[X+Y] = 2(Var[X]+Var[Y])$, we thus obtain $\mathrm{Var}[\sum_{j=n-m+1}^{n} W(j) - W(I^M)] \leq 4n$. By applying Chebyshev's inequality, we find that:

$$\begin{aligned}
&\mathbb{P}(|I^*| < \delta|I^M|) \\
&\leq \mathbb{P}\big(W(I^M) - \sum_{j=n-m+1}^{n} W(j) \leq 0\big) \\
&\leq \mathbb{P}\Big(\big(W(I^M) - \sum_{j=n-m+1}^{n} W(j)\big) \\
&\quad -\mathbb{E}[\big(W(I^M) - \sum_{j=n-m+1}^{n} W(j)\big)] \leq -\frac{0.1}{\Delta+1}n\Big) \\
&\leq (\frac{2(\Delta+1)}{.1})^2 \frac{1}{n} \\
&\leq \frac{400(\Delta+1)^2}{n}.
\end{aligned}$$

Armed with this inequality, we can finish the proof:

$$\begin{aligned}
\frac{\mathbb{E}[|I^*|]}{|I^M|} &= \mathbb{E}[\frac{|I^*|}{|I^M|}] \geq= \mathbb{P}(\frac{|I^*|}{|I^M|} > \delta)\mathbb{E}[\frac{|I^*|}{|I^M|} \mid \frac{|I^*|}{|I^M|} > \delta] \\
&\geq \delta(1 - \frac{400(\Delta+1)^2}{n}) \geq \frac{1}{10\log\Delta}(1 - \frac{400(\Delta+1)^2}{n})
\end{aligned}$$

For n sufficiently large, $(1 - \frac{400(\Delta+1)^2}{n})$ is greater than $1/2$, and $\frac{\mathbb{E}[|I^*|]}{|I^M|} \geq \frac{1}{20\log\Delta}$.