MIT Open Access Articles

Correlation Decay in Random Decision Networks

Correlation Decay in Random Decision Networks [*]

David Gamarnik [†‡]   David A. Goldberg [§]   Theophane Weber [¶]

October 30, 2018

Abstract

We consider a decision network on an undirected graph in which each node corresponds to a decision variable, and each node and edge of the graph is associated with a reward function whose value depends only on the variables of the corresponding nodes. The goal is to construct a decision vector which maximizes the total reward. This decision problem encompasses a variety of models, including maximum-likelihood inference in graphical models (Markov Random Fields), combinatorial optimization on graphs, economic team theory and statistical physics. The network is endowed with a probabilistic structure in which costs are sampled from a distribution. Our aim is to identify sufficient conditions on the network structure and cost distributions to guarantee average-case polynomiality of the underlying optimization problem. Additionally, we wish to characterize the efficiency of a decentralized solution generated on the basis of local information.

We construct a new decentralized algorithm called *Cavity Expansion* and establish its theoretical performance for a variety of graph models and reward function distributions. Specifically, for certain classes of models we prove that our algorithm is able to find near optimal solutions with high probability in a decentralized way. The success of the algorithm is based on the network exhibiting a certain correlation decay (long-range independence) property and we prove that this property is indeed exhibited by the models of interest. Our results have the following surprising implications in the area of average case complexity of algorithms. Finding the largest independent (stable) set of a graph is a well known NP-hard optimization problem for which no polynomial time approximation scheme is possible even for graphs with largest connectivity equal to three, unless P=NP. Yet we show that the closely related maximum weighted independent set problem for the same class of graphs admits a PTAS when the weights are independent identically distributed with the exponential distribution. Namely, randomization of the reward function turns an NP-hard problem into a tractable one.

1   Introduction and literature review

We consider a team of agents working in a networked structure (V, E), where V is a set of agents, and E the set of edges of the network, each edge indicating potential local interactions between

agents. Each agent v has to make a decision x_v from a finite set, and the team incurs a total reward $F(x) = \sum \Phi_v(x_v) + \sum_{u,v \in E} \Phi_{u,v}(x_u, x_v)$. The goal of each agent is to choose its decision x_v so that the total reward F is maximized. This model subsumes many models in a variety of fields including economic team theory, statistical inference, combinatorial optimization on graphs and statistical physics.

As an example, common models in the area of statistical inference are graphical models, Bayesian networks, and Markov Random Fields (MRF) (see [WJ08] for an overview of inference techniques for graphical models, and [MM08, HW05] for a comprehensive study of the relations between statistical physics, statistical inference, and combinatorial optimization). One of the key objects in such a model is the state which achieves the mode of the density, namely, the state which maximizes the a priori likelihood. The problem of finding such a state can be cast as a problem defined above.

In the economic team theory (see [Mar55, Rad62, MR72]), an interesting question was raised in [RR01]: what is the cost of decentralization in a chain of agents. In other words, if we assume that each node only receives local information on the network topology and costs, what kind of performance can the team attain? Cast in our framework, this the problem of finding the maximum of $F(x)$ by means of local (decentralized) algorithms.

Combinatorial optimization problems typically involve the task of finding a solution which minimizes or maximizes some objective function subject to various constraints supported by the underlying graph. Examples include the problem of finding a largest independent set, minimum and maximum cut problems, max-KSAT problems, etc. Finding an optimal solution in many such problems is a special case of the problem of finding $\max_x F(x)$ described above.

Finally, a key object in statistical physics models is the so-called ground state – a state which achieves the minimum possible energy. Again, finding such an object reduces to solving the problem described above, namely solving the problem $\max_x F(x)$ ($\min_x -F(x)$ to be more precise).

The combinatorial optimization nature of the decision problem $\max_x F(x)$ implies that the problem of finding $x^* = \text{argmax}_x F(x)$ is generally NP-hard, even for the special case when the decision space for each agent consists only of two elements. This motivates a search for approximate methods which find solutions that theoretically or empirically achieve good proximity to optimality. Such methods usually differ from field to field. In combinatorial optimization the focus has been on developing methods which achieve some provably guaranteed approximation level using a variety of approaches, including linear programming, semi-definite relaxations and purely combinatorial methods [Hoc97]. In the area of graphical models, researchers have been developing new families of distributed inference algorithms. One of the most studied techniques is the Belief Propagation (BP) algorithm [Lau96, Jor04, YFW00]. Since the algorithm proposed in the present paper bears some similarity and is motivated by BP algorithm, we provide below a brief summary of known theoretical facts about BP.

The BP algorithm is known to find an optimal solution x^* when the underlying graph is a tree, but may fail to converge, let alone produce an optimal (or correct) solution when the underlying graph contains cycles. Despite this fact, it often has excellent empirical performance. Also, in some cases, BP can be proven to produce an optimal solution, even when the underlying graph contains cycles. In a framework similar to ours, Moallemi and Van Roy [MR09] show that BP converges and produces an optimal solution when the action space is continuous and the cost functions $\Phi_{u,v}$ and Φ_u are quadratic and convex. Some generalization to generally convex functions are obtained in [MR07]. Other cases where BP produces optimal solutions include Maximum Weighted

Bipartite Matching [San07, BBCZ08, BSS08] (for matchings), Maximum Weighted Independent Sets problems where the LP relaxation is tight ([SSW07]), network flow problems [GSW09], and more generally, optimization problems defined on totally unimodular constraint matrices [Che08].

In this paper, we propose a new message-passing like algorithm for the problem of finding $x^* = \text{argmax } F(x)$, which we call the *Cavity Expansion* (CE) algorithm, and obtain sufficient conditions for the asymptotic optimality of our algorithm based on the so-called correlation decay property. Our algorithm draws upon several recent ideas. On the one hand, we rely on a technique used recently for constructing approximate counting algorithms. Specifically, Bandyopadhyay and Gamarnik [BG08], and Weitz [Wei06] proposed approximate counting algorithms which are based on local (in the graph-theoretic sense) computation. Provided that the model exhibits a form of correlation decay these algorithms are approximate counting algorithms. The approach was later extended in Gamarnik and Katz [GK07b],[GK07a], Bayati et al [BGK$^+$07], Jung and Shah [JS07]. The present work develops a similar approach but for the *optimization* problems. The description of the CE algorithm begins by introducing a notion of a *cavity* $B_v(x)$ for each node/decision pair (v, x) (the notion of cavity was heavily used recently in the statistical physics literature [MP03, RBMM04]). It is also called *bonus* in the relevant papers [Ald01],[AS03],[GNS06]. $B_v(x)$ is defined as the difference between the optimal reward for the entire network when the action in v is x versus the optimal reward when the action in the same node is 0 (any other base action can be taken instead of 0). It is easily shown that knowing $B_v(x)$ is equivalent to solving the original decision problem. We obtain a recursion expressing the cavity $B_v(x)$ in terms of cavities of the neighbors of v in suitably modified sub-networks of the underlying network. The algorithm then proceeds by expanding this recursion in the breadth-first search manner for some designed number of steps t, thus constructing an associated computation tree with depth t. At the initialization point the cavity values are assigned some default value. Then the approximation value $\hat{B}_v(x)$ is computed using this computation tree. If this computation was conducted for t equalling roughly the length L of the longest self-avoiding path of the graph, it would result in exact computation of the cavity values $B_v(x)$. Yet the computation effort associated with this scheme is exponential in L, which itself often grows linearly with the size of the graph.

The key insight of our work is that in many cases, the dependence of the cavity $B_v(x)$ on cavities associated with other nodes in the computation tree dies out exponentially fast as a function of the distance between the nodes. This phenomenon is generally called *correlation decay*. In earlier work [Ald92, Ald01, AS03, GNS06, GG09], it is shown that some optimization problems on locally tree-like graphs with random costs are tractable as they exhibit the correlation decay property. This is precisely our approach: we show that if we compute $B_v(x)$ based on the computation tree with only constant depth t, the resulting error $\hat{B}_v(x) - B_v(x)$ is exponentially small in r. By taking $r = O(\log(1/\epsilon))$ for any target accuracy ϵ, this approach leads to an ϵ-approximation scheme for computing the optimal reward $\max_x F(x)$. Thus, the main associated technical goal is establishing the correlation decay property for the associated computation tree.

We indeed establish that the correlation decay property holds for several classes of decision networks associated with random reward functions $\Phi = (\Phi_v, \Phi_{v,u})$. Specifically, we give concrete results for the cases of uniform and Gaussian distributed functions for unconstrained optimization in networks with bounded connectivity (graph degree) Δ. We also consider exponentially distributed (with parameter 1) weights for the Maximum Weighted Independent Set problem. In this setting, the combination of CE (a message passing style algorithm) and a randomized setting has a particularly interesting implication for the theory of average case analysis of combinatorial optimization.

3

Unlike some other NP-hard problems, finding the MWIS of a graph 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 degree at most 3, no factor 1.0071 approximation algorithm can exist, under the same complexity-theoretic assumption, see Berman and Karpinski [BK98]. In contrast, we show when $\Delta \leq 3$ and the node weights are independently generated with a parameter 1 exponential distribution, the problem of finding the maximum weighted independent set admits a PTAS. Thus, surprisingly, introducing random weights translates a combinatorially intractable problem into a tractable one. We further extend these results to the case $\Delta > 3$, but for different node weight distributions.

The rest of the paper is organized as follows. In section 2, we describe the general model and notations. In section 3, we present our main results. In section 4, we derive the cavity recursion, an exact recursion for computing the cavity of a node in a decision network, and from it develop the Cavity Expansion algorithm. In section 5, we prove that the correlation decay property implies optimality of the cavity recursion and local optimality of the solution. The rest of the paper is devoted to identifying sufficient conditions for correlation decay (and hence, optimality of the CE algorithm): in section 6, we show how a coupling argument can be used to prove the correlation decay property for the case of uniform and Gaussian weight distributions, and in section 7, we establish the correlation decay property for the MWIS problem using a different argument based on monotonocity. Concluding thoughts are in section 8.

2   Model description and notations

Consider a decision network $\mathcal{G} = (V, E, \Phi, \chi)$. Here (V, E) is an undirected simple graph in which each node $u \in V$ represents an agent, and edges $e \in E$ represent a possible interaction between two agents. Each agent makes a decision $x_u \in \chi \triangleq \{0, 1, \ldots, T-1\}$. For every $v \in V$, a function $\Phi_v : \chi \to \mathbb{R}$ is given. Also for every edge $e = (u, v)$ a function $\Phi_e : \chi^2 \to \mathbb{R} \cup \{-\infty\}$ is given. The inclusion of $-\infty$ into the range of Φ_e is needed in order to model the "hard constraints" in the MWIS problem - prohibiting two ends of an edge to belong to an independent set. Functions Φ_v and Φ_e will be called *potential functions* and *interaction functions* respectively. Let $\Phi = ((\Phi_v)_{v \in V}, (\Phi_e)_{e \in E})$. A vector $\mathbf{x} = (x_1, x_2, \ldots, x_{|V|})$ of actions is called a solution for the decision network. The value of solution $\mathbf{x}$ is defined to be $F_\mathcal{G}(\mathbf{x}) = \sum_{(u,v) \in E} \Phi_{u,v}(x_u, x_v) + \sum_v \Phi_v(x_v)$. The quantity $J_\mathcal{G} \triangleq \max_\mathbf{x} F_\mathcal{G}(\mathbf{x})$ is called the (optimal) value of the network $\mathcal{G}$. A decision $\mathbf{x}$ is optimal if $F_\mathcal{G}(\mathbf{x}) = J_\mathcal{G}$.

In a Markov Random Field (MRF), a set of random variables $\mathbf{X} = (X_1, \ldots, X_n)$ is assigned a probability $\mathbb{P}(\mathbf{X} = \mathbf{x})$ proportional to $\exp(F_\mathcal{G}(\mathbf{x}))$. In this context, the quantity $F_\mathcal{G}(\mathbf{x})$ can be considered as the log-likelihood of assignment $\mathbf{x}$, and maximizing it corresponds to finding a maximum a posterior assignment of the MRF defined by $F_\mathcal{G}$.

The main focus of this paper will be on the case where $\Phi_v(x), \Phi_e(x, y)$ are random variables (however, the actual realizations of the random variables are observed by the agents, and their decisions depend on the values $\Phi_v(x)$ and $\Phi_e(x, y)$). While we will usually assume independence of these random variables when v and e vary, we will allow dependence for the same v and e when we vary the decisions x, y. The details will be discussed when we proceed to concrete examples.

2.1 Examples

2.1.1 Independent set

Suppose the nodes of the graph are equipped with weights $W_v \geq 0, v \in V$. A set of nodes $I \subset V$ is an independent set if $(u, v) \notin E$ for every $u, v \in I$. The weight of an (independent) set I is $\sum_{u \in I} W_u$. The maximum weight independent set problem is the problem of finding the independent set I with the largest weight. It can be recast as a decision network problem by setting $\chi = \{0, 1\}, \Phi_e(0, 0) = \Phi_e(0, 1) = \Phi_e(1, 0) = 0, \Phi_e(1, 1) = -\infty, \Phi_v(1) = W_v, \Phi_v(0) = 0$.

2.1.2 Graph Coloring

An assignment ϕ of nodes V to colors $\{1, \ldots, q\}$ is defined to be proper coloring if no monochromatic edges are created. Namely, for every edge (v, u), $\phi(v) \neq \phi(u)$. Suppose each node/color pair $(v, x) \in V \times \{1, \ldots, q\}$ is equipped with a weight $W_{v,x} \geq 0$. The (weighted) coloring problem is the problem of finding a proper coloring ϕ with maximum total weight $\sum_v W_{v,\phi(v)}$. In terms of decision network framework, we have $\Phi_{v,u}(x, x) = -\infty, \Phi_{v,u}(x, y) = 0, \forall x \neq y \in \chi = \{1, \ldots, q\}, (v, u) \in E$ and $\Phi_v(x) = W_{v,x}, \forall v \in V, x \in \chi$.

2.1.3 MAX 2-SAT

Let $(Z_1, \ldots, Z_n)$ be a set of boolean variables. Let $(C_1, \ldots, C_m)$ be a list of clauses of the form $(Z_i \vee Z_j), (Z_i \vee \overline{Z_j}), (\overline{Z_i} \vee Z_j)$ or $(\overline{Z_i} \vee \overline{Z_j})$. The MAX-2SAT problem consists in finding an assignment for binary variables Z_i which maximizes the number of satisfied clauses C_j. In terms of a decision network, take $V = \{1, \ldots, n\}$, $E = \{(i, j) : Z_i \text{ and } Z_j \text{ appear in a common clause}\}$, and for any k, let $\Phi_k(x, y)$ to be 1 if the clause C_k is satisfied when $(Z_i, Z_j) = (x, y)$ and 0 otherwise. Let $\Phi_v(x) = 0$ for all v, x.

2.1.4 MAP estimation

In this example, we see a situation in which the reward functions are naturally randomized. Consider a graph (V, E) with $|V| = n$ and $|E| = m$, a set of real numbers $\mathbf{p} = (p_1, \ldots, p_n) \in [0, 1]^n$, and a family $(f_1, \ldots, f_m)$ of functions such that for each $(i, j) \in E$, $f_{i,j} = f_{i,j}(o, x, y) : \mathbb{R} \times \{0, 1\}^2 \to \mathbb{R}_+$ where $o \in \mathbb{R}$ and $x, y \in \{0, 1\}$. Assume that for each (x, y), $f_{i,j}(\cdot, x, y)$ is a probability density function. Consider two sets $\mathbf{C} = (C_i)_{1 \leq i \leq n}$ and $\mathbf{O} = (O_j)_{1 \leq j \leq m}$ of random variables, with joint probability density

$$P(\mathbf{O}, \mathbf{C}) = \prod_i p_i^{c_i} (1 - p_i)^{1 - c_i} \prod_{(i,j) \in E} f_{i,j}(o_{i,j}, c_i, c_j)$$

$\mathbf{C}$ is a set of Bernoulli random variables ("causes") with probability $P(C_i = 1) = p_i$, and $\mathbf{O}$ is a set of continuous "observation" random variables. Conditional on the cause variables $\mathbf{C}$, the observation variables $\mathbf{O}$ are independent, and each $O_{i,j}$ has density $f_{i,j}(o, c_i, c_j)$. Assume the variables $\mathbf{O}$ represent observed measurements used to infer on hidden causes $\mathbf{C}$. Using Bayes's formula, given observations $\mathbf{O}$, the log posterior probability of the causes variables $\mathbf{C}$ is equal to:

$$\log P(\mathbf{C} = \mathbf{c} \mid \mathbf{O} = \mathbf{o}) = K + \sum_i \Phi_i(c_i) + \sum_{i,j \in E} \Phi_{i,j}(c_i, c_j)$$

where

$$\Phi_i(c_i) = \log(p_i/(1 - p_i))c_i$$
$$\Phi_{i,j}(c_i, c_j) = \log(f_{i,j}(o_{i,j}, c_i, c_j))$$

where K is a random number which does not depend on $\mathbf{c}$. Finding the maximum a posteriori values of $\mathbf{C}$ given $\mathbf{O}$ is equivalent to finding the optimal solution of the decision network $\mathcal{G} = (V, E, \Phi, \{0, 1\})$. Note that the interaction functions $\Phi_{i,j}$ are naturally randomized, since $\Phi_{i,j}(x, y)$ is a continuous random variable with distribution

$$d\mathbb{P}(\Phi_{i,j}(x, y) = t) = e^t \sum_{x', y' \in \{0,1\}} d\mathbb{P}(f_{i,j}(o, x', y') = e^t)$$

2.2 Notations

For any two nodes u,v in V, let $d(u, v)$ be the length (number of edges) of the shortest path between u and v. Given a node u and integer $r \geq 0$, let $\mathcal{B}_{\mathcal{G}}(u, r) \triangleq \{v \in V : d(u, v) \leq r\}$ and $\mathcal{N}_{\mathcal{G}}(u) \triangleq \mathcal{B}(u, 1) \backslash \{u\}$ be the set of neighbors of u. For any node u, let $\Delta_{\mathcal{G}}(u) \triangleq |\mathcal{N}_{\mathcal{G}}(u)|$ be the number of neighbors of u in $\mathcal{G}$. Let $\Delta_{\mathcal{G}}$ be the maximum degree of graph (V, E); namely, $\Delta_{\mathcal{G}} = \max_v |\mathcal{N}(v)|$. Often we will omit the reference to the network $\mathcal{G}$ when it is obvious from the context.

For any subgraph (V', E') of (V, E) (i.e. $V' \subset V$, $E' \subset E \cap V'^2$), the subnetwork $\mathcal{G}'$ induced by (V, E) is the network (V', E', Φ', χ), where $\Phi' = ((\Phi_v)_{v \in V'}, (\Phi_e)_{e \in E'})$.

Given a subset of nodes $\mathbf{v} = (v_1, \ldots, v_k)$, and $\mathbf{x} = (x_1, \ldots, x_k) \in \chi^k$, let $J_{\mathcal{G},\mathbf{v}}(\mathbf{x})$ be the optimal value when the actions of nodes $v_1, \ldots, v_k$ are fixed to be $x_1, \ldots, x_k$ respectively: $J_{\mathcal{G},\mathbf{v}}(\mathbf{x}) = \max_{\mathbf{x}:x_{v_i}=x_i, 1 \leq i \leq k} F_{\mathcal{G}}(\mathbf{x})$. Given $v \in V$ and $x \in \chi$, the quantity $B_{\mathcal{G},v}(x) \triangleq J_{\mathcal{G},v}(x) - J_{\mathcal{G},v}(0)$ is called the *cavity* of action x at node v. Namely it is the difference of optimal values when the decision at node v is set to x and 0 respectively (the choice of 0 is arbitrary). The cavity function of v is $B_{\mathcal{G},v} = (B_{\mathcal{G},v}(x))_{x \in \chi}$. Since $B_{\mathcal{G},v}(0) = 0$, $B_{\mathcal{G},v}$ can be thought of as element of $\mathbb{R}^{T-1}$. In the important special case $\chi = \{0, 1\}$, the cavity function is a scalar $B_{\mathcal{G},v} = J_{\mathcal{G},v}(1) - J_{\mathcal{G},v}(0)$. In this case, if $B_{\mathcal{G},v} > 0$ (resp. $B_{\mathcal{G},v} < 0$) then $J_{\mathcal{G},v}(1) > J_{\mathcal{G},v}(0)$ and action 1 (resp. action 0) is optimal for v. When $B_{\mathcal{G},v} = 0$ there are optimal decisions consistent both with $x_v = 0$ and $x_v = 1$. Again, when $\mathcal{G}$ is obvious from the context, it will be omitted from the notation.

For any network $\mathcal{G}$, we call $M(\mathcal{G}) = \max(|V|, |E|, |\chi|)$ the size of the network. Since we will exclusively consider graphs with degree bounded by a constant, for all practical purposes we can think of $|V|$ 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 $|V|$. An algorithm $\mathcal{A}$ is said to be an ϵ-loss additive approximation algorithm for the problem of finding the optimal decision if for any network $\mathcal{G}$ it produces in polynomial time a decision $\hat{x}$ such that $J_{\mathcal{G}} - F(\hat{x}) < \epsilon$. If all cost functions are positive, the algorithm $\mathcal{A}$ is said to be an $(1 + \epsilon)$-factor multiplicative approximation algorithm if it outputs a solution $\hat{x}$ such that $J_{\mathcal{G}}/F(\hat{x}) < 1 + \epsilon$. We call such an algorithm an additive (resp. multiplicative) PTAS (Polynomial Time Approximation Scheme) if it is an ϵ-loss (resp. $(1 + \epsilon)$-factor) additive (resp. multiplicative) approximation factor algorithm for every $\epsilon > 0$ and runs in time which is polynomial in $|V|$. An algorithm is called an FPTAS (Fully Polynomial

Time Approximation Scheme) if it runs in time which is polynomial in n and $1/\epsilon$. For our purposes another relevant class of algorithms is EPTAS. This is the class of algorithms which produces ϵ approximation in time $O(|V|^{O(1)}g(\epsilon))$, 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. Finally, in our context, since the input is random, we will say that an algorithm is an additive (resp. multiplicative) PTAS with high probability if for all $\epsilon > 0$ it outputs in time polynomial in $|V|$ a solution $\hat{x}$ such that $\mathbb{P}(J_{\mathcal{G}} - F(\hat{x}) > \epsilon) < \epsilon$ (resp. $\mathbb{P}(F(\hat{x})/J_{\mathcal{G}} > 1 + \epsilon) \leq \epsilon$); FPTAS and EPTAS w.h.p. are similarly defined. Since our algorithm provide probabilistic guarantee, one may wonder whether FPRAS (Fully Polynomial Randomized Approximtion Scheme) would be a more appropriate framework. The typical setting for FPRAS is, however, a deterministic problem input and the randomization is associated purely with algorithm. In our setting however, the setting itself is random, though the algorithms, with the exception of MWIS, are deterministic.

3 Main results

In this section we state our main results. The first two results relate to decision networks with uniformly and normally distributed costs, respectively, without any combinatorial constraints on the decisions. The last set of results corresponds to the MWIS problem, which does incorporate the combinatorial constraint of the independence property.

3.1 Uniform and Gaussian Distributions

Given $\mathcal{G} = (V, E, \Phi, \{0, 1\})$, suppose that for all $u \in V$, $\Phi_u(1)$ is uniformly distributed on $[-I_1, I_1]$, $\Phi_u(0) = 0$, and that for every $e \in E$, $\Phi_e(0, 0), \Phi_e(1, 0), \Phi_e(0, 1)$ and $\Phi_e(1, 1)$ are all independent and uniformly distributed on $[-I_2, I_2]$, for some $I_1, I_2 > 0$. Intuitively, I_1 quantifies the 'bias' each agent has towards one action or another, while I_2 quantifies the strength of interactions between agents.

Theorem 1. *Let $\beta = \frac{5I_2}{2I_1}$. If $\beta(\Delta - 1)^2 < 1$, then there exists an additive FPTAS for finding $J_{\mathcal{G}}$ with high probability.*

Now we turn to the case of Gaussian costs. Assume that for any edge $e = (u, v)$ and any pair of action $(x, y) \in \{0, 1\}^2$, $\Phi_{u,v}(x, y)$ is a Gaussian random variable with mean 0 and standard deviation σ_e. For every node $v \in V$, suppose $\Phi_v(1) = 0$ and that $\Phi_v(0)$ is a Gaussian random variable with mean 0 and standard deviation σ_p. Assume that all rewards $\Phi_e(x, y)$ and $\Phi_v(x)$ are independent for all choices of v, e, x, y.

Theorem 2. *Let $\beta = \sqrt{\frac{\sigma_e^2}{\sigma_e^2 + \sigma_p^2}}$. If $\beta(\Delta - 1) + \sqrt{\beta(\Delta - 1)^3} < 1$, then there exists an additive FPTAS for finding $J_{\mathcal{G}}$ with high probability.*

While our main result was stated for the case of independent costs, we have obtained a more general result which incorporates the case of correlated edge costs. It is given as Proposition 6 in Section 6.

3.2 Maximum Weight Independent Sets

Here, we consider a variation of the MWIS problem where the nodes of the graph are equipped with random weights $W_i, i \in V$, drawn independently from a common distribution $F(t) = \mathbb{P}(W \leq t), t \geq 0$. Let $I^* = I^*(\mathcal{G})$ 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 almost surely unique when F is a continuous distribution.

Theorem 3. *If $\Delta_{\mathcal{G}} \leq 3$ and the weights are exponentially distributed with parameter 1, then there exists a multiplicative EPTAS for finding $J_{\mathcal{G}}$ with high probability. The algorithm runs in time* $O\left(|V|2^{O(\epsilon^{-2}\log(1/\epsilon))}\right)$.

An interesting implication of Theorem 3 is that while the Maximum (cardinality) Independent Set problem admits neither a polynomial time algorithm nor a PTAS (unless P=NP), even when the degree is bounded to 3 [BK98, Tre01], the problem of finding the maximum weight independent set becomes tractable for certain distributions F, in the PTAS sense.

The exponential distribution is not the only distribution which can be analyzed in this framework, it is just the easiest to work with. For any phase-type distribution, we can characterize correlation decay and identify sufficient conditions for correlation decay to hold. It is natural to ask if the above result can be generalized, and in particular to wonder if it is possible to find for each Δ a distribution which guarantees that the correlation decay property holds for graphs with degree bounded by Δ. It is indeed possible, as we extend Theorem 3, 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 4. *Assume $\Delta_{\mathcal{G}} \leq \Delta$, and that the weights are distributed according to $P(W > t) = \frac{1}{\Delta}\sum_{1\leq j\leq \Delta}\exp(-\alpha_j t)$. Then there exists a FPTAS for finding $J_{\mathcal{G}}$ with high probability. The algorithm runs in time* $O\left(n(\frac{1}{\epsilon})^{\Delta}\right)$.

Note that for the mixture of exponential distributions described above our algorithm is in fact an FPTAS as opposed to an EPTAS for Theorem 3. This is essentially due to the fact that the conditions of Theorem 3 are at the 'boundary' of correlation decay; more technical details are given in 7.

Our final result is a partial converse to the results above; one could conjecture that randomizing the weights makes the problem essentially easy to solve, and that perhaps being able to solve the randomized version does not tell us much about the deterministic version. We show that this is not the case, and that the setting with random weights hits a complexity-theoretic barrier just as the classical cardinality problem does. Specifically, we show that for graphs with sufficiently large degree the problem of finding the largest weighted independent set with i.i.d. exponentially distributed weights does not admit a PTAS. We need to keep in mind that since we 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.

Specifically, for any $\rho < 1$, define an algorithm $\mathcal{A}$ to be a factor-ρ polynomial time approximation algorithm for computing $\mathbb{E}[W(I^*)]$ 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 $\rho \leq \hat{w}/E[W(I^*)] \leq 1/\rho$ in time bounded by $O(n^{O(1)})$. Here the expectation is with respect to the exponential weight distribution and the constant exponent $O(1)$ is allowed to depend on Δ.

En route of Theorems 3 and 4 we establish similar results for expectations: there exists an EPTAS and FPTAS respectively for computing the deterministic quantity $E[W(I^*)]$, the expected weight of the MWIS in the graph $\mathcal{G}$ considered.

However, our next result shows that if the maximum degree of the graph is increased, it is impossible to approximate the quantity $E[W(I^*)]$ arbitrarily closely, unless P=NP. Specifically,

Theorem 5. *There exist Δ_0 and c_1^*, c_2^* such that for all $\Delta \geq \Delta_0$ the problem of computing $\mathbb{E}[W(I^*)]$ to within a multiplicative factor $\rho = \Delta/(c_1^* (\log \Delta) \, 2^{c_2^* \sqrt{\log \Delta}})$ for graphs with degree at most Δ cannot be solved in polynomial time, unless P=NP.*

We could compute a concrete Δ_0 such that for all $\Delta \geq \Delta_0$ the claim of the theorem holds, though such Δ_0 explicitly does not seem to offer much insight. We note that in the related work by Trevisan [Tre01], no attempt is made to compute a similar bound either.

4 The cavity recursion

In this section, we introduce the *cavity recursion*, an exact recursion for computing the cavity functions of each node in a general decision network. We first start by giving the cavity recursion for trees (which is already known as the max-product belief propagation algorithm), and then give a generalization for all networks.

4.1 Trees

Given a decision network $\mathcal{G} = (V, E, \Phi, \chi)$ suppose that (V, E) is a rooted tree with a root u. Using the graph orientation induced by the choice of u as a root, let $\mathcal{G}_v$ be the subtree rooted in node v for any node $v \in V$. In particular, $\mathcal{G} = \mathcal{G}_u$. Denote by $C(u)$ the set of children of u in (V, E). Given a node $u \in V$, a child $v \in C(u)$, and an arbitrary vector $B = (B(x), x \in \chi)$, define

$$\mu_{u \leftarrow v}(x, B) = \max_y (\Phi_{u,v}(x, y) + B(y)) - \max_y (\Phi_{u,v}(0, y) + B(y)) \tag{1}$$

for every action $x \in \chi$. μ is called partial cavity function.

Proposition 1. *For every $u \in V$ and $x \in \chi$,*

$$B_u(x) = \Phi_u(x) - \Phi_u(0) + \sum_{v \in C(u)} \mu_{u \leftarrow v}(x, B_{\mathcal{G}_v, v}) \tag{2}$$

Proof. Suppose $C(u) = \{v_1, \ldots, v_d\}$. Observe that the subtrees $\mathcal{G}_{v_i}, 1 \leq i \leq d$ are disconnected (see
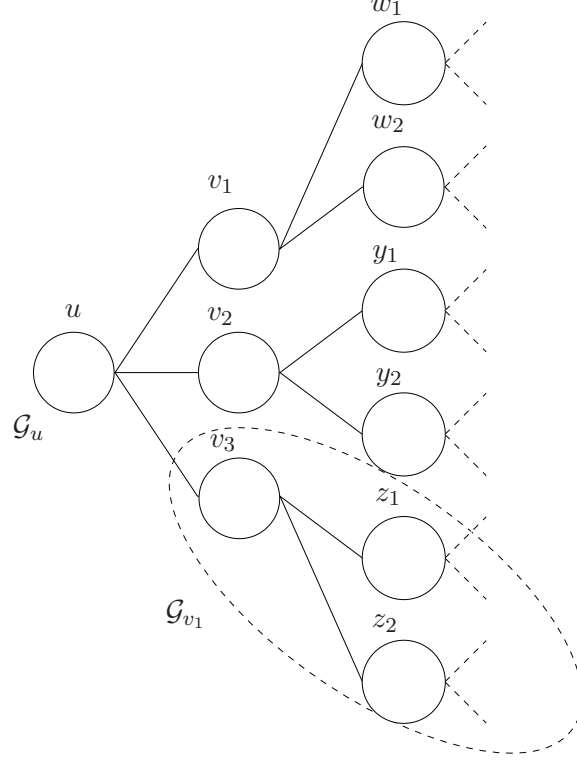
Figure 1: Cavity recursion for trees, equivalent to the BP algorithm

figure 4.1) Thus,

$$B_u(x) = \Phi_u(x) + \max_{x_1,\ldots,x_d} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(x,x_j) + J_{\mathcal{G}_{v_j},v_j}(x_j) \right\}$$

$$- \Phi_u(0) - \max_{x_1,\ldots,x_d} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(0,x_j) + J_{\mathcal{G}_{v_j},v_j}(x_j)) \right\}$$

$$= \Phi_u(x) - \Phi_u(0)$$

$$+ \sum_{j=1}^{d} \left\{ \max_y \left(\Phi_{u,v_j}(x,y) + J_{\mathcal{G}_{v_j},v_j}(y) \right) - \max_y \left(\Phi_{u,v_j}(0,y) + J_{\mathcal{G}_{v_j},v_j}(y) \right) \right\}$$

For every j

$$\max_y \left(\Phi_{u,v_j}(x,y) + J_{\mathcal{G}_{v_j},v_j}(y) \right) - \max_y \left(\Phi_{u,v_j}(0,y) + J_{\mathcal{G}_{v_j},v_j}(y) \right) =$$

$$\max_y \left(\Phi_{u,v_j}(x,y) + J_{\mathcal{G}_{v_j},v_j}(y) - J_{\mathcal{G}_{v_j},j}(0) \right) - \max_y \left(\Phi_{u,v_j}(0,y) + J_{\mathcal{G}_{v_j},v_j}(y) - J_{\mathcal{G}_{v_j},v_j}(0) \right)$$

The quantity above is exactly $\mu_{u \leftarrow v_j}(x, B_{v_j,\mathcal{G}_{v_j}})$. □

Iteration (2) constitutes what is known as (max-product) belief propagation. Proposition 1 is the restatement of the well-known fact that BP finds an optimal solution on a tree (citation). BP

10

can be implemented in non-tree like graphs, but then it is not guaranteed to converge, and even when it does it may produce wrong (suboptimal) solutions. In the following section we construct a generalization of BP which is guaranteed to converge to an optimal decision.

4.2 General graphs

The goal of this subsection is to construct a generalization of identity (2) for an arbitrary network $\mathcal{G}$. This can be achieved by building a sequence of certain auxiliary decision networks $\mathcal{G}(u, j, x)$ constructed as follows.

Given a decision network $\mathcal{G} = (V, E, \Phi, \chi)$ where the underlying graph is arbitrary, fix any node u and action x and let $\mathcal{N}(u) = \{v_1, \ldots, v_d\}$. For every $j = 1, \ldots, d$ let $\mathcal{G}(u, j, x)$ be the decision network (V', E', Φ', χ) on the same decision set χ constructed as follows. (V', E') is the subgraph induced by $V' = V \setminus \{u\}$. Namely, $E' = E \setminus \{(u, v_1), \ldots, (u, v_d)\}$. Also $\Phi'_e = \Phi_e$ for all e in E' and the potential functions Φ'_v are defined as follows. For any $v \in V \setminus \{u, v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_d\}$, $\Phi'_v = \Phi_v$, and

$$
\begin{aligned}
\Phi'_v(y) &= \Phi_v(y) + \Phi_{u,v}(x, y) && \text{for } v \in \{v_1, \ldots, v_{j-1}\} \\
\Phi'_v(y) &= \Phi_v(y) + \Phi_{u,v}(0, y) && \text{for } v \in \{v_{j+1}, \ldots, v_d\}
\end{aligned}
\tag{3}
$$

Theorem 6 (Cavity Recursion). *For every $x \in \chi$,*

$$
B_u(x) = \Phi_u(x) - \Phi_u(0) + \sum_{j=1}^{d} \mu_{u \leftarrow v_j}(x, B_{\mathcal{G}(u,j,x),v_j})
\tag{4}
$$

Proof. For every $k = 0, 1, \ldots, d$, let $x_{j,k} = x$ when $j \le k$ and $= 0$ otherwise. Let $\mathbf{v} = (v_1, \ldots, v_d)$, and $\mathbf{z} = (z_1, \ldots, z_d) \in \chi^d$. We have

$$
\begin{aligned}
B_u(x) = \Phi_u(x) - \Phi_u(0) &+ \max_{\mathbf{z}} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(x, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\} \\
&- \max_{\mathbf{z}} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(0, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\}.
\end{aligned}
$$

The first step of the proof consists in considering the following telescoping sum (see figure 4.2):

$$
\begin{aligned}
B_u(x) = \Phi_u(x) - \Phi_u(0) + \sum_{k=1}^{d} &\left[\max_{\mathbf{z}} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(x_{j,k}, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\} \right. \\
&\left. - \max_{\mathbf{z}} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(x_{j,k-1}, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\} \right]
\end{aligned}
\tag{5}
$$

and the kth difference:

$$
\max_{\mathbf{z}} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(x_{j,k}, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\} - \max_{\mathbf{z}} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(x_{j,k-1}, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\}
\tag{6}
$$

11

Figure 2: First step: building the telescoping sum; black nodes indicate decision x, gray node decision 0; solid circles indicate neighbors of u, dotted circles indicate other nodes

with

$$\tilde{\Phi}_{v_1}(z) = \Phi_{v_1}(z) + \Phi_{u,v_1}(0, z)$$
$$\tilde{\Phi}_{v_3}(z) = \Phi_{v_3}(z) + \Phi_{u,v_3}(x, z)$$

Figure 3: Second step: build the modified subnetworks (here $\mathcal{G}(u, 2, x)$); arrow represents modification of the potential function by incorporating interaction function into them

Let $\mathbf{z}_{-k} = (z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_d)$. Then,

$$\max_{\mathbf{z}} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(x_{j,k}, z_j) + J_{\mathcal{G}\setminus\{u\},\mathbf{v}}(\mathbf{z}) \right\} =$$

$$\max_{z_k} \left(\Phi_{u,v_k}(x, z_k) + \max_{\mathbf{z}_{-k}} \left\{ \sum_{j \leq k-1} \Phi_{u,v_j}(x, z_j) + \sum_{j \geq k+1} \Phi_{u,v_j}(0, z_j) + J_{\mathcal{G}\setminus\{u\},\mathbf{v}}(\mathbf{z}) \right\} \right) \qquad (7)$$

Similarly,

$$\max_{\mathbf{z}} \left\{ \sum_{j=1}^{d} \Phi_{u,v_j}(x_{j,k-1}, z_j) + J_{\mathcal{G}\setminus\{u\},\mathbf{v}}(\mathbf{z}) \right\} =$$

$$\max_{z_k} \left(\Phi_{u,v_k}(0, z_k) + \max_{\mathbf{z}_{-k}} \left\{ \sum_{j \leq k-1} \Phi_{u,v_j}(x, z_j) + \sum_{j \geq k+1} \Phi_{u,v_j}(0, z_j) + J_{\mathcal{G}\setminus\{u\},\mathbf{v}}(\mathbf{z}) \right\} \right) \qquad (8)$$

For each z_k, we have (see figure 4.2):

$$\max_{\mathbf{z}_{-k}} \left\{ \sum_{j \leq k-1} \Phi_{u,v_j}(x, z_j) + \sum_{j \geq k+1} \Phi_{u,v_j}(0, z_j) + J_{\mathcal{G}\setminus\{u\},\mathbf{v}}(\mathbf{z}) \right\} = J_{\mathcal{G}(u,k,x),v_k}(z_k)$$

By adding and substrating $J_{\mathcal{G}(u,k,x),v_k}(0)$, expression (6) can therefore be rewritten as

$$\max_y (\Phi_{u,v_k}(x, y) + B_{\mathcal{G}(u,k,x)}(y)) - \max_y (\Phi_{u,v_k}(0, y) + B_{\mathcal{G}(u,k,x)}(y))$$

which is exactly $\mu_{u \leftarrow v_k}(x, B_{\mathcal{G}(u,k,x)})$. Finally, we obtain:

$$B_u(x) = \Phi_u(x) - \Phi_u(0) + \sum_{k=1}^{d} \mu_{u \leftarrow v_k}(x, B_{\mathcal{G}(u,k,x),v_k})$$

$\square$

4.3 Computation tree and the Cavity Expansion algorithm

Given a decision network $\mathcal{G}$, a node $u \in V$ with $\mathcal{N}_u = \{v_1, \ldots, v_d\}$, and $r \in \mathbb{Z}_+$, introduce a vector $\text{CE}[\mathcal{G}, u, r] = (\text{CE}[\mathcal{G}, u, r, x], x \in \chi) \in \mathbb{R}^T$ defined recursively as follows.

1. $\text{CE}[\mathcal{G}, u, 0, x] = 0$

2. For every $r = 1, 2, \ldots$, and every $x \in \chi$,

$$\text{CE}[\mathcal{G}, u, r, x] = \Phi_u(x) - \Phi_u(0) + \sum_{j=1}^{d} \mu_{u \leftarrow v_j}\left(x, \text{CE}[\mathcal{G}(u, j, x), v_j, r - 1] \right), \qquad (9)$$

where $\mathcal{G}(u, k, x)$ is defined in Subsection 4.2, and the sum $\sum_{j=1}^{d}$ is equal to 0 when $\mathcal{N}_u = \emptyset$. Note that from the definition of $\mathcal{G}(u, k, x)$, the definition and output of $\text{CE}[\mathcal{G}, u, r]$ depend on the order in which the neighbors v_j of u are considered. $\text{CE}[\mathcal{G}, u, r]$ serves as an r-step approximation, in

14

some appropriate sense to be explained later, of the cavity vector $B_{\mathcal{G},u}$. The motivation for this definition is relation (4) of Theorem 6. The local cavity approximation can be computed using an algorithm described below, which we call *Cavity Expansion (CE)* algorithm.

```
    Cavity Expansion:   CE[𝒢, u, r, x]
INPUT: A network 𝒢, a node u in 𝒢, an action x and a computation depth r ≥ 0
BEGIN
If r = 0 return 0
else do
Find neighbors 𝒩(u) = {v₁, v₂, ..., v_d} of u in 𝒢.
If 𝒩(u) = ∅, return Φ_u(x) − Φ_u(0).
Else
For each j = 1, ..., d, construct the network 𝒢(u, j, x).
For each j = 1, ..., d, and y ∈ χ, compute CE[𝒢(u, j, x), v_j, r − 1, y]
For each j = 1, ..., d, compute μ_{u←v_j}(x, CE[𝒢(u, j, x), v_j, r − 1, y])
Return Φ_u(x) − Φ_u(0) + ∑_{1≤j≤d} μ_{u←v_j}(x, CE[𝒢(u, j, x), v_j, r − 1, y]) as CE[𝒢, u, r, x].
```

The algorithm above terminates because r decreases by one at each recursive call of the algorithm. As a result, an initial call to CE$[\mathcal{G}, u, r, x]$ will result in a finite number of recursive calls to some CE$[\mathcal{G}_j, u_j, k_j, x_j]$, where $k_j < r$. Let $(\mathcal{G}_i, v_i, x_i)_{1 \le i \le m}$ be the subset of arguments for the calls used in computing CE$[\mathcal{G}, u, r, x]$ for which $k_i = 0$. In the algorithm above, the values returned for $r = 0$ are 0, but it can be generalized by choosing a value $\mathcal{C}_i$ for the call CE$[\mathcal{G}_i, v_i, 0, x_i]$.

The set of values $\mathcal{C} = (\mathcal{C}_i)_{1 \le i \le m}$ will be called a *boundary condition*. We denote by CE$[\mathcal{G}, u, r, x, \mathcal{C}]$ the output of the cavity algorithm with boundary condition $\mathcal{C}$. The interpretation of CE$[\mathcal{G}, u, r, x, \mathcal{C}]$ is that it is an estimate of the cavity $B_{\mathcal{G},u}(x)$ via r steps of recursion (2) when the recursion is initialized by setting CE$[\mathcal{G}_i, u_i, 0, x_i] = \mathcal{C}_i$ and is run r steps. We will sometimes omit $\mathcal{C}$ from the notation when such specification is not necessary. Call $\mathcal{C}^* = (\mathcal{C}_i^*) \triangleq (B_{\mathcal{G}_i, v_i}(x_i))$ the "true boundary condition". The justification comes from the following proposition, the proof of which follows directly from Theorem 6.

Proposition 2. *Given node u and $\mathcal{N}(u) = \{v_1, \ldots, v_d\}$, suppose for every $j = 1, \ldots, d$ and $y \in \chi$, $CE[\mathcal{G}(u,j,x), v_j, r-1, y] = B_{\mathcal{G}(u,j,x), v_j}(y)$; then, $CE[\mathcal{G}, u, r, x] = B_{\mathcal{G},u}(x)$.*

As a result, if $\mathcal{C}$ is the "correct" boundary condition, then CE$[\mathcal{G}, u, r, x, \mathcal{C}] = B_{\mathcal{G},u}(x)$ for every u, r, x. The execution of the Cavity Expansion algorithm can be visualized as a computation on a tree, due to its recursive nature. This has some similarity with a computation tree associated with the performance of the Belief Propagation algorithm, [TJ02, SSW07, BSS08]. The important difference with [TJ02] is that the presence of cycles is incorporated via the construction $\mathcal{G}(u, j, x)$ (similarly to [Wei06, JS07, BGK⁺07, GK07a, GK07b]. As a result, the computation tree of the CE is finite (though often extremely large), as opposed to the BP computation tree.

An important lemma, which we will use frequently in the rest of the paper, states that in the computation tree of the cavity recursion, the cost function of an edge cost is statistically independent from the subtree below that edge.

15

Proposition 3. *Given u, x and $\mathcal{N}(v) = \{v_1, \ldots, v_d\}$, for every $r, j = 1, \ldots, d$ and $y \in \chi$, $CE[\mathcal{G}(u, j, x), v_j, r - 1, y]$ and Φ_{u,v_j} are independent.*

Note however that Φ_{u,v_j} and $CE[\mathcal{G}(u, k, x), v_k, r - 1, y]$ are generally dependent when $j \neq k$

Proof. The proposition follows from the fact that for any j, the interaction function Φ_{u,v_j} does not appear in $\mathcal{G}(u, j, x)$, because node u does not belong to $\mathcal{G}(u, j, x)$), and does not modify the potential functions of $\mathcal{G}(u, j, x)$ in the step (3). □

Our last proposition analyzes the complexity of running the Cavity Expansion algorithm.

Proposition 4. *For every $\mathcal{G}, u, r, x$, the value $CE[\mathcal{G}, u, r, x]$ can be computed in time $O\big(r(\Delta T)^r\big)$.*

Proof. The computation time required to construct the networks $\mathcal{G}(u, j, x)$, compute the messages $\mu_{u \leftarrow v_j}(x, \hat{B}_{v_j})$, and return $\Phi_u(x) - \Phi_u(0) + \sum_{1 \leq j \leq d} \mu_{u \leftarrow v_j}(x, \hat{B}_{v_j})$, is $O(\Delta T)$. Let us prove by induction that that for any subnetwork $\mathcal{G}'$ of $\mathcal{G}$, $\bar{\mathrm{CE}}[\mathcal{G}', u, r, x]$ can be computed in time bounded by $O(r(\Delta T)^r)$. The values for $r = 1$ can be computed in time bounded by M, since $\mathcal{G}'$ is a subnet of $\mathcal{G}$ and therefore of smaller size. For $r > 1$, the computations of $\mathrm{CE}[\mathcal{G}', u, r, x]$ requires a fixed cost of $O(\Delta T)$, as well as (ΔT) calls to CE with depth $(r - 1)$. The total cost is therefore bounded by $O(\Delta T + (\Delta T)(r - 1)(\Delta T)^{r-1})$, which is $O(r(\Delta T)^r)$. □

5   Correlation decay and decentralized optimization

In this section, we investigate the relations between the correlation decay phenomenon and the existence of near-optimal decentralized decisions. When a network exhibits the correlation decay property, the cavity functions of faraway nodes are weakly related, implying a weak dependence between their optimal decisions as well. Thus one can expect that good decentralized decisions exist. We will show that this is indeed the case.

Definition 1. *Given a function $\rho(r) \geq 0, r \in \mathbb{Z}_+$ such that $\lim_{r \to \infty} \rho(r) = 0$, a decision network $\mathcal{G}$ is said to satisfy the correlation decay property with rate ρ if for every two boundary conditions $\mathcal{C}$, $\mathcal{C}'$*

$$\max_{u,x} \mathbb{E} \big| CE[\mathcal{G}, u, r, x, \mathcal{C}] - CE[\mathcal{G}, u, r, x, \mathcal{C}'] \big| \leq \rho(r).$$

If there exists $K_c > 0$ and $\alpha_c < 1$ such that $\rho(r) \leq K_c \alpha_c^r$ for all r, then we say that $\mathcal{G}$ satisfies the exponential correlation decay property with rate α_c.

The correlation decay property implies that for every u, x,

$$\mathbb{E} \big| CE[\mathcal{G}, u, r, x] - B_{\mathcal{G},u}(x) \big| \leq \rho(r).$$

The following assumptions will be frequently used in future.

Assumption 1. *For all $v \in V, x, y \in \chi$, $B_v(x) - B_v(y)$ is a continuous random variable with density bounded above by a constant $g > 0$.*

We will also assume the costs functions are bounded in L_2 norm:

Assumption 2. *There exists K_Φ such that for any $e \in E$, $\left(\sum_{x,y \in \chi} \mathbb{E}|\Phi_e(x,y)|^2 \right)^{1/2} \leq K_\Phi$ and for any $v \in V$, $\left(\sum_{x \in \chi} \mathbb{E}|\Phi_v(x)|^2 \right)^{1/2} \leq K_\Phi$*

Assumption 1 is designed to lead to the following two properties: (a) There is a unique optimal action in every node with probability 1. (b) The suboptimality gap between the optimal action and the second best action is large enough so that there is a "clear winner" among actions.

5.1 Correlation decay implies near-optimal decentralized decisions

Under Assumption 1 let $\mathbf{x} = (x_v)_{v \in V}$ be the unique (with probability one) optimal solution for the network $\mathcal{G}$. For every $v \in V$, $x \in \chi$, let $x_v^r = \mathrm{argmax}_x CE[\mathcal{G}, v, r, x]$, ties broken arbitrarily, and $\mathbf{x}^r = (x_v^r)$. The main relation between correlation decay property, Cavity Expansion algorithm and the optimization problem is given by the following result.

Proposition 5. *Suppose $\mathcal{G}$ exhibits the correlation decay property with rate $\rho(r)$ and that Assumption 1 holds. Then,*

$$\mathbb{P}(x_u^r \neq x_u) \leq 2T^2 \sqrt{2g\rho(r)}, \quad \forall u \in V, r \geq 1. \tag{10}$$

Proof. For simplicity, let $B_u^r(x)$ denote $CE[\mathcal{G}, u, r, x]$. We will first prove that

$$\mathbb{P}(x_u^r \neq x_u) \leq T^2 \left(g\epsilon + \frac{2\rho(r)}{\epsilon} \right) \tag{11}$$

The proposition will follow by choosing $\epsilon = \sqrt{2\rho(r)g^{-1}}$. Consider a node u, and notice that if

$$(B_u(x) - B_u(y))(B_u^r(x) - B_u^r(y)) > 0, \quad \forall x \neq y,$$

then $x_u^r = x_u$. Indeed, since $B_u(x_u) - B_u(y) > 0$ for all $y \neq x_u$, the property implies the same for B_u^r, and the assertion holds. Thus, the event $\{x_u^r \neq x_u\}$ implies the event

$$\{\exists(x,y), y \neq x : (B_u(x) - B_u(y))(B_u^r(x) - B_u^r(y)) \leq 0\}$$

Fix $\epsilon > 0$ and note that for two real numbers r and s, if $|r| > \epsilon$ and $|r-s| \leq \epsilon$, then $rs > 0$. Applying this to $r = B_u(x) - B_u(y)$ and $s = B_u^r(x) - B_u^r(y)$, we find that the events $|B_u(x) - B_u(y)| > \epsilon$ and

$$(|B_u(x) - B_u^r(x)| < \epsilon/2) \cap (|B_u(y) - B_u^r(y)| < \epsilon/2)$$

jointly imply

$$(B_u(x) - B_u(y))(B_u^r(x) - B_u^r(y)) > 0$$

Therefore, the event $(B_u(x) - B_u(y))(B_u^r(x) - B_u^r(y)) \leq 0$ implies

$$\{|B_u(x) - B_u(y)| \leq \epsilon\} \cup \{|B_u(x) - B_u^r(x)| \geq \epsilon/2\} \cup \{|B_u(y) - B_u^r(y)| \geq \epsilon/2\}$$

Applying the union bound, for any two actions $x \neq y$,

$$\mathbb{P}\Big((B_u(x) - B_u(y))(B_u^r(x) - B_u^r(y)) \leq 0\Big) \leq \mathbb{P}(|B_u(x) - B_u(y)| \leq \epsilon) + \mathbb{P}(|B_u(x) - B_u^r(x)| \geq \epsilon/2)$$
$$+ \mathbb{P}(|B_u(y) - B_u^r(y)| \geq \epsilon/2). \tag{12}$$

Now $\mathbb{P}(|B_u(x) - B_u(y)| \leq \epsilon)$ is at most $2g\epsilon$ by Assumption 1. Using the Markov inequality, we find that the second summand in (12) is at most $2\mathbb{E}|B_u(x) - B_u^r(x)|/\epsilon \leq 2\rho(r)/\epsilon$. The same bound applies to the third summand. Finally, noting there are $T(T-1)/2$ different pairs (x, y) with $x \neq y$ and applying the union bound, we obtain:

$$\begin{aligned} \mathbb{P}(x_u^r \neq x_u) &\leq (T(T-1)/2)(2g\epsilon + 4\rho(r)/\epsilon) \\ &\leq T^2(g\epsilon + \frac{2\rho(r)}{\epsilon}). \end{aligned}$$

$\square$

For the special case of exponential correlation decay, we obtain the following result, the proof of which immediately follows from Proposition 5.

Corollary 1. *Suppose $\mathcal{G}$ exhibits the exponential correlation decay property with rate α_c, and suppose Assumption 1 holds. Then*

$$\mathbb{P}(x_u^r \neq x_u) \leq 2T^2\sqrt{2gK_c}\alpha_c^{r/2}, \qquad \forall u \in V, r \geq 1.$$

In particular, for any $\epsilon > 0$, if

$$r \geq 2\frac{|\log K_c'| + |\log \epsilon|}{|\log(\alpha_c)|}$$

then

$$\mathbb{P}(x_u^r \neq x_u) \leq \epsilon$$

where $K_c' = 2T^2\sqrt{2gK_c}$

In summary, correlation decay - and in particular fast (i.e. exponential) correlation decay - implies that the optimal action in a node depends with high probability only on the structure of the network in a small radius around the node. As in [RR03], we call such a property *decentralization* of optimal actions. Note that the radius required to achieve an ϵ error does not depend on the size of the entire network; moreover, for exponential correlation decay, it grows only as a logarithm of the accepted error.

The main caveat of Proposition 5 is that the Assumption 1 does not necessarily hold. For instance, it definitely does not apply to models with discrete random variables Φ_u and $\Phi_{u,v}$. In fact, assumption 1 is not really necessary, as we show in an online appendix that a regularization technique allows to relax this assumption. Note that Assumption 2 is not needed for Proposition 5 to hold.

5.2 Correlation decay and efficient decentralized optimization

Proposition 5 illustrates how optimal actions are decentralized under the correlation decay property. In this section, we use this result to show that the resulting optimization algorithm is both near-optimal and computationally efficient.

As before, let before $\mathbf{x} = (x_u)$ denote the optimal solution for the network $\mathcal{G}$, and let $\mathbf{x}^r = (x_u^r)$ be the decisions resulting from the Cavity Expansion algorithm with depth r. Let $\tilde{\mathbf{x}} = (\tilde{x}_u)$ denote (any) optimal solution for the perturbed network $\tilde{\mathcal{G}}$. Let $K_1 = 10K_\Phi T(|V| + |E|)$, and $K_2 = K_1(g\, K_c)^{1/4}$, where K_c is defined under the assumption of exponential correlation decay.

Theorem 7. *Suppose a decision network $\mathcal{G}$ satisfies correlation decay property with rate $\rho(r)$. Then, for all $r > 0$*

$$\mathbb{E}[F(\mathbf{x}) - F(\mathbf{x}^{r,\delta})] \leq K_1(g\rho(r))^{1/4} \tag{13}$$

Corollary 2. *Suppose $\mathcal{G}$ exhibits exponential correlation decay property with rate α_c. Then, for any $\epsilon > 0$, if*

$$r \geq \big(8|\log \epsilon| + 4|\log(K_2)|\big)|\log(\alpha_c)|^{-1}$$

then

$$\mathbb{P}\big(F(\mathbf{x}) - F(\mathbf{x}^r) > \epsilon\big) \leq \epsilon$$

and $\mathbf{x}^r$ can be computed in time polynomial in $|V|$, $1/\epsilon$.

Proof. By applying the union bound on Proposition 5, for every (u,v), we have: $\mathbb{P}\big((x_u^r, x_v^r) \neq (x_u, x_v)\big) \leq 4T^2\sqrt{2g\rho(r)}$. We have

$$\mathbb{E}|F(\mathbf{x}) - F(\mathbf{x}^r)| \leq \sum_{u \in V} \mathbb{E}|\Phi_u(x_u) - \Phi_u(x_u^r)| + \sum_{(u,v) \in E} \mathbb{E}|\Phi_{u,v}(x_u, x_v) - \Phi_{u,v}(x_u^r, x_v^r)|$$

For any $u, v \in V$,

$$\mathbb{E}[\Phi_{u,v}(x_u, x_v) - \Phi_{u,v}(x_u^r, x_v^r)] \leq \mathbb{E}\Big[1_{(x_u^r, x_v^r) \neq (x_u, x_v)} \big(\big|\Phi_{u,v}(x_u, x_v)\big| + \big|\Phi_{u,v}(x_u^r, x_v^r)\big|\big)\Big]$$

$$\leq 2K_\Phi \, \mathbb{P}\big((x_u^r, x_v^r) \neq (x_u, x_v)\big)^{1/2}$$

$$\leq 4K_\Phi T \, (2g\rho(r))^{1/4}$$

where the second inequality follows from Cauchy-Schwarz. Similarly, for any u we have

$$\mathbb{E}|\Phi_u(x_u) - \Phi_u(x_u^r)| \leq \leq 4K_\Phi T \, (2g\rho(r))^{1/4}$$

By summing over all nodes and edges, we get: $\mathbb{E}[F(\mathbf{x}) - F(\mathbf{x}^r) \leq 8K_\Phi \, T(2g\rho(r))^{1/4} \leq K_1(g\rho(r))^{1/4}$, and equation (13) follows. The corollary is then proved using Markov Inequality; injecting the definition of exponential correlation decay into equation (13), we obtain

$$\mathbb{P}(J_\mathcal{G} - F(\hat{x}) \geq \epsilon) \leq E[J_\mathcal{G} - F(\hat{x})]/\epsilon \leq K_2\alpha^{r/4}/\epsilon$$

Since $r \geq (4|\log(K_2)| + 8|\log(\epsilon)|)|\log(\alpha)|^{-1}$, we have $K_2\alpha^{r/4} \leq \epsilon^2$ and the result follows. $\qquad\square$

6 Establishing the correlation decay property. Coupling technique

The previous section motivates the search for conditions implying the correlation decay property. This section is devoted to the study of a coupling argument which can be used to show that correlation decay holds. Results in this section are for the case $|\chi| = 2$. They can be extended to the case $|\chi| \geq 2$ at the expense of heavier notations, but not much additional insight gain. For this special case $\chi = \{0, 1\}$, we introduce a set of simplifying notations as follows.

6.1 Notations

Given $\mathcal{G} = (V, E, \Phi, \{0, 1\})$ and $u \in V$, let $v_1, \ldots, v_d$ be the neighbors of u in V. For any $r > 0$ and boundary conditions $\mathcal{C}, \mathcal{C}'$, define:

1. $B(r) \triangleq \text{CE}[\mathcal{G}, u, r, 1, \mathcal{C}]$ and $B'(r) \triangleq \text{CE}[\mathcal{G}, u, r, 1, \mathcal{C}']$

2. For $j = 1, \ldots d$, let $\mathcal{G}_j = \mathcal{G}(u, j, 1)$, and let $B_j(r - 1) \triangleq \text{CE}[\mathcal{G}_j, v_j, r - 1, 1, \mathcal{C}]$ and $B'_j(r - 1) \triangleq \text{CE}[\mathcal{G}_j, v_j, r - 1, 1, \mathcal{C}']$. Also let $\mathbf{B}(r - 1) = (B_j(r - 1))_{1 \leq j \leq d}$ and $\mathbf{B}'(r - 1) = (B'_j(r - 1))_{1 \leq j \leq d}$

3. For $k = 1, \ldots n_j$, let $(v_{j1}, \ldots, v_{jn_j})$ be the neighbors of v_j in $\mathcal{G}_j$, and let $B_{jk}(r - 2) = \text{CE}[\mathcal{G}_j(v_j, k, 1), v_{jk}, r - 2, 1, \mathcal{C}]$ and $B'_{jk}(r - 2) = \text{CE}[\mathcal{G}_j(v_j, k, 1), v_j, r - 2, 1, \mathcal{C}']$ for all $k = 1 \ldots n_j$. Also let $\mathbf{B_j}(r - 2) = (B_{jk}(r - 2))_{1 \leq k \leq n_j}$ and $\mathbf{B'_j}(r - 2) = (B'_{jk}(r - 2))_{1 \leq k \leq n_j}$.

4. For simplicity, since 1 is the only action different from the reference action 0, we denote $\mu_{u \leftarrow v_j}(z) \triangleq \mu_{u \leftarrow v_j}(1, z)$.
   From equation (1), note the following alternative expression for $\mu_{u \leftarrow v_j}(z)$

$$\mu_{u \leftarrow v_j}(z) = \Phi_{u,v_j}(1, 1) - \Phi_{u,v_j}(0, 1) + \max(\Phi_{u,v_j}(1, 0) - \Phi_{u,v_j}(1, 1), z) \qquad (14)$$
$$- \max(\Phi_{u,v_j}(0, 0) - \Phi_{u,v_j}(0, 1), z)$$

5. Similarly, for any $j = 1 \ldots d$ and $k = 1 \ldots n_j$, let $\mu_{v_j \leftarrow v_{jk}}(z) \triangleq \mu_{v_j \leftarrow v_{jk}}(1, z)$.

6. For any $\mathbf{z} = (z_1, \ldots, z_d)$, let $\mu_u(\mathbf{z}) = \sum_j \mu_{u \leftarrow v_j}(z_j)$. Also, for any j, and any $\mathbf{z} = (z_1, \ldots, z_{n_j})$, let $\mu_{v_j}(\mathbf{z}) = \sum_{1 \leq k \leq n_j} \mu_{v_j \leftarrow v_{jk}}(z_k)$.

7. For any directed edge $e = (u \leftarrow v)$, denote

$$\begin{aligned}
\Phi_e^1 &\triangleq \Phi_{u,v}(1, 0) - \Phi_{u,v}(1, 1) \\
\Phi_e^2 &\triangleq \Phi_{u,v}(0, 0) - \Phi_{u,v}(0, 1) \\
\Phi_e^3 &\triangleq \Phi_{u,v}(1, 1) - \Phi_{u,v}(0, 1) \\
X_e &\triangleq \Phi_e^1 + \Phi_e^2 \\
Y_e &\triangleq \Phi_e^2 - \Phi_e^1 = \Phi_{u,v}(1, 1) - \Phi_{u,v}(1, 0) - \Phi_{u,v}(0, 1) + \Phi_{u,v}(0, 0)
\end{aligned}$$

Note that $Y_{u \leftarrow v} = Y_{v \leftarrow u}$, so we simply denote it $Y_{u,v}$.

Note that for any e, $\mathbb{E}|Y_e| \leq K_\Phi$ (see Assumption 2). Equation (9) can be rewritten as

$$\begin{aligned}
B(r) &= \mu_u(\mathbf{B}(r - 1)) + \Phi_u(1) - \Phi_u(0) & (15) \\
B'(r) &= \mu_u(\mathbf{B}'(r - 1)) + \Phi_u(1) - \Phi_u(0) & (16)
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
B_j(r - 1) &= \mu_{v_j}(\mathbf{B_j}(r - 2)) + \Phi_{v_j}(1) - \Phi_{v_j}(0) & (17) \\
B'_j(r - 1) &= \mu_{v_j}(\mathbf{B'_j}(r - 2)) + \Phi_{v_j}(1) - \Phi_{v_j}(0) & (18)
\end{aligned}$$

20

Finally, equation (14) can be rewritten

$$\mu_{u \leftarrow v}(z) \quad = \quad \Phi^3_{u \leftarrow v} + \max(\Phi^1_{u \leftarrow v}, z) - \max(\Phi^2_{u \leftarrow v}, z) \tag{19}$$

Y_e represents how strongly the interaction function $\Phi_{u,v}(x_u, x_v)$ is "coupling" the variables x_u and x_v. In particular, if Y_e is zero, the interaction function $\Phi_{u,v}(x_u, x_v)$ can be decomposed into a sum of two potential functions $\Phi_u(x_u) + \Phi_v(x_v)$, that is, the edge between u and v is then be superfluous and can be removed. To see why this is the case, take $\Phi_u(0) = 0$, $\Phi_u(1) = \Phi_{u,v}(1,0) - \Phi_{u,v}(0,0)$, $\Phi_v(0) = \Phi_{u,v}(0,0)$ and $\Phi_v(1) = \Phi_{u,v}(0,1)$, which is also equal to $\Phi_{u,v}(1,1) - \Phi_{u,v}(1,0) + \Phi_{u,v}(0,0)$, since $Y_e = 0$.

6.2 Distance-dependent coupling and correlation decay

Definition 2. *A network $\mathcal{G}$ is said to exhibit (a,b)-coupling with parameters (a,b) if for every edge $e = (u,v)$, and every two real values x, x':*

$$\mathbb{P}\Big(\mu_{u \leftarrow v}(x + \Phi_v(1) - \Phi_v(0)) = \mu_{u \leftarrow v}(x' + \Phi_v(1) - \Phi_v(0))\Big) \geq (1 - a) - b|x - x'| \tag{20}$$

The probability above, and hence the coupling parameters, depend on both $\Phi_v(1) - \Phi_v(0)$ and the values $\Phi_{u,v}(x, y)$. Note that if for all x, x'

$$\mathbb{P}\Big(\mu_{u \leftarrow v}(x) = \mu_{u \leftarrow v}(x')\Big) \geq (1 - a) - b|x - x'| \tag{21}$$

then $\mathcal{G}$ exhibits (a,b) coupling, but in general the tightest coupling values found for equation (21) are much weaker than the ones we would find by analyzing condition (20). This form of distance dependent coupling is a useful tool in proving that correlation decay occurs, as illustrated by the following theorem:

Theorem 8. *Suppose $\mathcal{G}$ exhibits (a,b)-coupling. If*

$$a(\Delta - 1) + \sqrt{bK_\Phi}(\Delta - 1)^{3/2} < 1 \tag{22}$$

then the exponential correlation decay property holds with $K = \Delta^2 K_\Phi$ and $\alpha = a(\Delta-1) + \sqrt{bK_\Phi}(\Delta - 1)^{3/2}$.

Suppose $\mathcal{G}$ exhibits (a,b)-coupling and that there exists $K_Y > 0$ such that $|Y_e| \leq K_Y$ with probability 1. If

$$a(\Delta - 1) + bK_Y(\Delta - 1)^2 < 1 \tag{23}$$

then the exponential correlation decay property holds with $\alpha = a(\Delta - 1) + bK_Y(\Delta - 1)^2$

6.2.1 Proof of Theorem 8

We begin by proving several useful lemmas.

Lemma 1. *For every (u,v), and every two real values x, x'*

$$|\mu_{u \leftarrow v}(x) - \mu_{u \leftarrow v}(x')| \leq |x - x'|. \tag{24}$$

Proof. From (14) we obtain

$$\mu_{u\leftarrow v}(x) - \mu_{u\leftarrow v}(x') = \max\left(\Phi_{u,v}(1,0) - \Phi_{u,v}(1,1), x\right) - \max\left(\Phi_{u,v}(0,0) - \Phi_{u,v}(0,1), x\right)$$
$$- \max\left(\Phi_{u,v}(1,0) - \Phi_{u,v}(1,1), x'\right) + \max\left(\Phi_{u,v}(0,0) - \Phi_{u,v}(0,1), x'\right).$$

Using twice the relation $\max_x f(x) - \max_x g(x) \leq \max_x(f(x) - g(x))$, we obtain:

$$\mu_{u\leftarrow v}(x) - \mu_{u\leftarrow v}(x') \leq \max(0, x - x') + \max(0, x' - x)$$
$$= |x - x'|$$

The other inequality is proved similarly. $\qquad\square$

Lemma 2. *For every $u, v \in V$ and every two real values x, x'*

$$|\mu_{u\leftarrow v}(x) - \mu_{u\leftarrow v}(x')| \leq |Y_{u,v}| \qquad (25)$$

Proof. Using (14) and (16), we have

$$\mu_{u\leftarrow v}(x) - (\Phi_{u,v}(1,1) - \Phi_{u,v}(0,1)) = \max(\Phi_{u,v}(1,0) - \Phi_{u,v}(1,1), x)$$
$$- \max(\Phi_{u,vx}(0,0) - \Phi_{u,v}(0,1), x).$$

By using the relation $\max_x f(x) - \max_x g(x) \leq \max_x(f(x) - g(x))$ on the right hand side, we obtain

$$\mu_{u\leftarrow v}(x) - (\Phi_{u,v}(1,1) - \Phi_{u,v}(0,1)) \leq \max(0, -Y_{u,v}).$$

Similarly

$$-\mu_{u\leftarrow v}(x') + (\Phi_{u,v}(1,1) - \Phi_{u,v}(0,1)) \leq \max(0, Y_{u,v}).$$

Adding up

$$\mu_{u\leftarrow v}(x) - \mu_{u\leftarrow v}(x') \leq |Y_{u,v}|.$$

The other inequality is also proven similarly. $\qquad\square$

Lemma 3. *Suppose (a, b)-coupling holds. Then,*

$$\mathbb{E}|B(r) - B'(r)| \leq a \sum_{1\leq j\leq d} \mathbb{E}|B_j(r-1) - B'_j(r-1)| + b \sum_{1\leq j\leq d} \mathbb{E}\left[|B_j(r-1) - B'_j(r-1)|^2\right]. \quad (26)$$

Proof. Using (9), we obtain:

$$\mathbb{E}|B(r) - B'(r)| = \mathbb{E}\left[\left|\Phi_u(1) - \Phi_u(0) + \sum_j \mu_{u\leftarrow v_j}(B_j(r-1)) - (\Phi_u(1) - \Phi_u(0)) - \sum_j \mu_{u\leftarrow v_j}(B'_j(r-1))\right|\right]$$

$$\leq \sum_j \mathbb{E}\left|\mu_{u\leftarrow v_j}(B_j(r-1)) - \mu_{u\leftarrow v_j}(B'_j(r-1))\right|$$

$$= \sum_j \mathbb{E}\left[\mathbb{E}\left[|\mu_{u\leftarrow v_j}(B_j(r-1)) - \mu_{u\leftarrow v_j}(B'_j(r-1))|\big|\mu_{v_j}(\mathbf{B_j}(r-2), \mu_{v_j}(\mathbf{B'_j}(r-2))\right]\right]$$

By Lemma 1, we have $|\mu_{u \leftarrow v_j}(B_j(r-1)) - \mu_{u \leftarrow v_j}(B'_j(r-1))| \leq |B_j(r-1) - B'_j(r-1)|$. Also note from that from equation (17) and (18), $|B_j(r-1) - B'_j(r-1)| = |\mu_{v_j}(\mathbf{B_j}(r-2)) - \mu_{v_j}(\mathbf{B'_j}(r-2))|$; hence conditional on both $\mu_{v_j}(\mathbf{B_j}(r-2))$ and $\mu_{v_j}(\mathbf{B'_j}(r-2))$, $|B_j(r-1) - B'_j(r-1)|$ is a constant. Therefore,

$$\mathbb{E}\Big[\big|\mu_{u \leftarrow v_j}(B_j(r-1)) - \mu_{u \leftarrow v_j}(B'_j(r-1))\big| \ \Big|\mu_{v_j}(\mathbf{B_j}(r-2), \mu_{v_j}(\mathbf{B'_j}(r-2)\Big]$$

$$\leq |B_j(r-1) - B'_j(r-1)| \, \mathbb{P}(\mu_{u \leftarrow v_j}(B_j(r-1)) \neq \mu_{u \leftarrow v_j}(B'_j(r-1)) \mid \mu_{v_j}(\mathbf{B_j}(r-2), \mu_{v_j}(\mathbf{B'_j}(r-2))$$

$$\tag{27}$$

Note that in the (a,b) coupling definition, the probability is over the values of the functions Φ_{u,v_j}, and Φ_v. By proposition 3, these are independent from $\mu_{v_j}(\mathbf{B_j}(r-2))$ and $\mu_{v_j}(\mathbf{B'_j}(r-2))$. Thus, by the (a,b) coupling assumption, $\mathbb{P}(\mu_{u \leftarrow v_j}(B_j(r-1)) \neq \mu_{u \leftarrow v_j}(B'_j(r-1)) \mid \mu_{v_j}(\mathbf{B_j}(r-2), \mu_{v_j}(\mathbf{B'_j}(r-2)) \leq a + b|B_j(r-1) - B'_j(r-1)|$. The result then follows. □

Fix an arbitrary node u in $\mathcal{G}$. Let $\mathcal{N}(u) = \{v_1, \ldots, v_d\}$. Let $d_j = |\mathcal{N}(v_j)| - 1$ be the number of neighbors of v_j in $\mathcal{G}$ other than u for $j = 1, \ldots, d$. We need to establish that for every two boundary condition $\mathcal{C}, \mathcal{C}'$

$$\mathbb{E}|\mathrm{CE}(\mathcal{G}, u, r, \mathcal{C}) - \mathrm{CE}(\mathcal{G}, u, r, \mathcal{C}')| \leq K\alpha^r \tag{28}$$

We first establish the bound inductively for the case $d \leq \Delta - 1$. Let e_d denote the supremum of the left-hand side of (28), where the supremum is over all networks $\mathcal{G}'$ with degree at most Δ, such that the corresponding constant $K_{\Phi'} \leq K_\Phi$, over all nodes u in $\mathcal{G}$ with degree $|\mathcal{N}(u)| \leq \Delta - 1$ and all over all choices of boundary conditions $\mathcal{C}, \mathcal{C}'$. Each condition corresponds to a different recursive inequality for e_r.

Condition (22)   Under (22), we claim that

$$e_r \leq a(\Delta - 1)e_{r-1} + b(\Delta - 1)^3 K_\Phi e_{r-2} \tag{29}$$

Applying (17) and (18), we have

$$|B_j(r-1) - B'_j(r-1)| \leq \sum_{1 \leq k \leq d_j} |\mu_{v_j \leftarrow v_{jk}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk}(r-2))|$$

Thus,

$$|B_j(r-1) - B'_j(r-1)|^2 \leq \Big(\sum_{1 \leq k \leq d_j} |\mu_{v_j \leftarrow v_{jk}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk}(r-2))| \Big)^2$$

$$\leq d_j \sum_{1 \leq k \leq d_j} |\mu_{v_j \leftarrow v_{jk}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk}(r-2))|^2$$

By Lemmas 1 and 2 we have $|\mu_{v_j \leftarrow v_{jk}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk}(r-2))| \leq |B_{jk}(r-2) - B'_{jk}(r-2)|$ and $|\mu_{v_j \leftarrow v_{jk}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk}(r-2))| \leq |Y_{jk}|$. Also, $d_j \leq \Delta - 1$. Therefore,

$$|B_j(r-1) - B'_j(r-1)|^2 \leq (\Delta - 1) \sum_{1 \leq k \leq d_j} |B_{jk}(r-2) - B'_{jk}(r-2)| \cdot |Y_{jk}| \tag{30}$$

23

By Proposition 3, the random variables $|B_{jk}(r-2) - B'_{jk}(r-2)|$ and $|Y_{jk}|$ are independent. We obtain:

$$\mathbb{E}|B_j(r-1) - B'_j(r-1)|^2 \le (\Delta - 1) \sum_{1 \le k \le d_j} \mathbb{E}|B_{jk}(r-2) - B'_{jk}(r-2)| \cdot \mathbb{E}|Y_{jk}| \tag{31}$$

$$\le (\Delta - 1)K_\Phi \left(\sum_{1 \le k \le d_j} \mathbb{E}|B_{jk}(r-2) - B'_{jk}(r-2)| \right)$$

$$\le (\Delta - 1)^2 K_\Phi e_{r-2}$$

where the second inequality follows from the definition of K_Φ and the third inequality follows from the definition of e_r and the fact that the neighbors v_{jk}, $1 \le k \le d_j$ of v_j have degrees at most $\Delta - 1$ in the corresponding networks for which $B_{jk}(r-2)$ and $B'_{jk}(r-2)$ were defined. Applying Lemma 3 and the definition of e_r, we obtain

$$\mathbb{E}|B(r) - B'(r)| \le a \sum_{1 \le j \le d} \mathbb{E}|B_j(r-1) - B'_j(r-1)| + b \sum_{1 \le j \le d} \mathbb{E}\left[|B_j(r-1) - B'_j(r-1)|^2\right]$$

$$\le a(\Delta - 1)e_{r-1} + b(\Delta - 1)^3 K_\Phi e_{r-2}$$

This implies (29).

From (29) we obtain that $e_r \le K\alpha^r$ for $K = \Delta K_\Phi$ and α given as the largest in absolute value root of the quadratic equation $\alpha^2 = a(\Delta - 1)\alpha + b(\Delta - 1)^3 K_\Phi$. We find this root to be

$$a = \frac{1}{2}(a(\Delta - 1) + \sqrt{a^2(\Delta - 1)^2 + 4b(\Delta - 1)^3 K_\Phi})$$

$$\le a(\Delta - 1) + \sqrt{b(\Delta - 1)^3 K_\Phi}$$

$$< 1$$

where the last inequality follows from assumption (22). This completes the proof for the case that the degree d of u is at most $\Delta - 1$.

Now suppose $d = |\mathcal{N}(u)| = \Delta$. Applying (15) and (16) we have

$$|B(r) - B'(r)| \le \sum_{1 \le j \le d} |\mu_{u \leftarrow v_j}(B_j(r-1)) - \mu_{u \leftarrow v_j}(B'_j(r-1))|$$

Applying again Lemma 1, the right-hand side is at most

$$\sum_{1 \le j \le d} |B_j(r-1) - B'_j(r-1)| \le \Delta e_{r-1}$$

since $B_j(r-1)$ and $B'_j(r-1)$ are defined for v_j in a subnetwork $\mathcal{G}_j = \mathcal{G}(u, j, 1)$, where v_j has degree at most $\Delta - 1$. Thus again the correlation decay property holds for u with ΔK replacing K.

Condition (23)   Recall from lemma 3 that for all r, we have:

$$\mathbb{E}|B(r) - B'(r)| \le a \sum_{1 \le j \le d} \mathbb{E}|B_j(r-1) - B'_j(r-1)| + b \sum_{1 \le j \le d} \mathbb{E}\left[|B_j(r-1) - B'_j(r-1)|^2\right].$$

For all j, $|B_j(r-1) - B'_j(r-1)| = |\sum_k(\mu_{v_j \leftarrow v_{jk}}(B_{jk}) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk}))|$. Moreover, for each j, k, $|\mu_{v_j \leftarrow v_{jk}}(B_{jk}) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk})| \leq |Y_{jk}| \leq K_Y$ (the second inequality follows from Lemma 2, the third by assumption). As a result,

$$|B_j(r-1) - B'_j(r-1)|^2 \leq (\Delta - 1)K_Y|B_j(r-1) - B_j(r-1)|$$

We obtain:

$$e_r \leq (a + bK_Y(\Delta - 1))(\Delta - 1)e_{r-1}$$

Since $a(\Delta - 1) + bK_Y(\Delta - 1)^2 < 1$, e_r goes to zero exponentially fast. The same reasoning as previously shows that this property implies correlation decay.

6.3 Establishing coupling bounds

6.3.1 Coupling Lemma

Theorem 8 details sufficient condition under which the distance-dependent coupling induces correlation decay (and thus efficient decentralized algorithms, vis-à-vis Proposition 4 and Theorem 7). It remains to show how can we prove coupling bounds. The following simple observation can be used to achieve this goal.

For any edge $(u, v) \in \mathcal{G}$, and any two real numbers x, x', consider the following events

$$E^+_{u \leftarrow v}(x, x') = \{\min(x, x') + \Phi_v(1) - \Phi_v(0) \geq \max(\Phi^1_{u \leftarrow v}, \Phi^2_{u \leftarrow v})\}$$

$$E^-_{u \leftarrow v}(x, x') = \{\max(x, x') + \Phi_v(1) - \Phi_v(0) \leq \min(\Phi^1_{u \leftarrow v}, \Phi^2_{u \leftarrow v})\}$$

$$E_{u \leftarrow v}(x, x') = E^+_{u,v}(x, x') \cup E^-_{u,v}(x, x')$$

Lemma 4. *If $E_{u \leftarrow v}(x, x')$ occurs, then $\mu_{u \leftarrow v}(x + \Phi_v(1) - \Phi_v(0)) = \mu_{u \leftarrow v}(x' + \Phi_v(1) - \Phi_v(0))$. Therefore*

$$P(\mu_{u \leftarrow v}(x + \Phi_v(1) - \Phi_v(0)) = \mu_{u \leftarrow v}(x' + \Phi_v(1) - \Phi_v(0)) \geq P(E_{u \leftarrow v}(x, x'))$$

Proof. From representation (19), we have $\mu_{u \leftarrow v}(x) = \Phi^3_{u \leftarrow v} + \max(\Phi^1_{u \leftarrow v}, z) - \max(\Phi^2_{u \leftarrow v}, z)$; let x, x' be any two reals. If both x and x' are greater than both $\Phi^1_{u \leftarrow v}$ and $Phi^2_{u \leftarrow v}$, then $\mu_{u \leftarrow v}(x) = \Phi^3_{u \leftarrow v} = \mu_{u \leftarrow v}(x')$. If both x and x' are smaller than both $\Phi^1_{u \leftarrow v}$ and $Phi^2_{u \leftarrow v}$, then $\mu_{u \leftarrow v}(x) = \Phi^3_{u \leftarrow v} + \Phi^1_{u \leftarrow v} - \Phi^2_{u \leftarrow v} = \mu_{u \leftarrow v}(x')$. The result follows from applying the above observation to $x + \Phi_v(1) - \Phi_v(0)$ and $x' + \Phi_v(1) - \Phi_v(0)$. $\square$

Note that Lemma 4 implies that the probability of coupling not occuring $P(\mu_{u \leftarrow v}(x + \Phi_v(1) - \Phi_v(0)) \neq \mu_{u \leftarrow v}(x' + \Phi_v(1) - \Phi_v(0)))$ is upper bounded by the probability of $(E_{u \leftarrow v}(x, x'))^c$. When obvious from context, we drop the subscript $u \leftarrow v$. We will often use the following description of $(E_{(}x, x'))^c$: for two real values $x \geq x'$,

$$(E(x, x'))^c = \{\min(\Phi^1, \Phi^2) + \Phi_v(0) - \Phi_v(1) < x < \max(\Phi^1, \Phi^2) + \Phi_v(0) - \Phi_v(1) + x - x'\} \quad (32)$$

6.3.2 Uniform Distribution. Proof of Theorem 1

In order to prove Theorem 1, we compute the coupling parameters a, b for this distribution and apply the second form of Theorem 8.

Lemma 5. *The network with uniformly distributed rewards described in section 3.1 exhibits (a, b) coupling with $a = \frac{I_2}{2I_1}$ and $b = \frac{1}{2I_1}$.*

Proof. For any fixed edge $(u, v) \in \mathcal{G}$, $\Phi^1_{u \leftarrow v}$ and $\Phi^2_{u \leftarrow v}$ are i.i.d. random variables with a triangular distribution with support $[-2I_2, 2I_2]$ and mode 0. Because $\Phi^1_{u \leftarrow v}$ and $\Phi^2_{u \leftarrow v}$ are i.i.d., by symmetry we obtain:

$$\mathbb{P}((E(x, x'))^c) =$$
$$= 2 \int_{-2I_2}^{2I_2} d\mathbb{P}_{\Phi^1}(a_1) \int_{a_1}^{2I_2} d\mathbb{P}_{\Phi^2}(a_2) \, P(a_1 + \Phi_v(0) - \Phi_v(1) < x < \Phi_v(0) - \Phi_v(1) + a_2 + x - x')$$
$$= 2 \int_{-2I_2}^{2I_2} d\mathbb{P}_{\Phi^1}(a_1) \int_{a_1}^{2I_2} d\mathbb{P}_{\Phi^2}(a_2) \, P(x' - a_2 < \Phi_v(0) - \Phi_v(1) < x - a_1)$$

$P(x' - a_2 < \Phi_v(0) - \Phi_v(1) < x - a_1)$ can be upper bounded by $\frac{a_2 - a_1 + x - x'}{2I_1}$, and we obtain:

$$P(E(x, x')^c) \leq \frac{x - x'}{2I_1} + \frac{1}{I_1} \int_{-2I_2}^{2I_2} d\mathbb{P}_{\Phi^1}(a_1) \int_{a_1}^{2I_2} d\mathbb{P}_{\Phi^2}(a_2)(a_2 - a_1)$$

Note that $d\mathbb{P}_{\Phi^2}(a_2) = \frac{1}{4I_2^2}(a_2 + 2I_2)d(a_2)$ for $a_2 \leq 0$, and $d\mathbb{P}_{\Phi^2}(a_2) = \frac{1}{4I_2^2}(2I_2 - a_2)d(a_2)$ for $a_2 \geq 0$; identical expressions hold for $d\mathbb{P}_{\Phi^1}(a_1)$. Therefore, for $a_1 \geq 0$,

$$\int_{a_1}^{2I_2} d\mathbb{P}_{\Phi^2}(a_2)(a_2 - a_1) = \frac{1}{4I_2^2} \int_{a_1}^{2I_2} (2I_2 - a_2)(a_2 - a_1) \, d(a_2)$$
$$= \frac{1}{4I_2^2} \left(- \int_{a_1}^{2I_2} (2I_2 - a_2)^2 d(a_2) + (2I_2 - a_1) \int_{a_1}^{2I_2} (2I_2 - a_2)d(a_2) \right)$$
$$= \frac{1}{4I_2^2} \left(- \frac{1}{3}(2I_2 - a_1)^3 + \frac{1}{2}(2I_2 - a_1)^3 \right) = \frac{1}{24I_2^2}(2I_2 - a_1)^3$$

Similarly, for $a_1 \leq 0$,

$$\int_{a_1}^{2I_2} d\mathbb{P}_{\Phi^2}(a_2)(a_2 - a_1) = -a_1 + \frac{1}{24I_2^2}(a_1 + 2I_2)^3$$

The final integral is therefore equal to:

$$\int_{-2I_2}^{2I_2} d\mathbb{P}_{\Phi^1}(a_1) \int_{a_1}^{2I_2} d\mathbb{P}_{\Phi^2}(a_2)(a_2 - a_1)$$
$$= \frac{1}{4I_2^2} \left(\int_{-2I_2}^{0} \left((a_1 + 2I_2)(-a_1 + \frac{1}{24I_2^2}(a_1 + 2I_2)^3)d(a_1) + \int_{0}^{2I_2} \frac{1}{24I_2^2}(2I_2 - a_1)^4 d(a_1) \right)$$
$$= \frac{1}{4I_2^2} \left(\frac{24}{15}I_2^3 + \frac{4}{15}I_2^3 \right) = \frac{7}{15}I_2$$

26

Finally,

$$P((E(x,x')^c) \leq \frac{7I_2}{15I_1} + \frac{|x-x'|}{2I_1} \leq \frac{I_2}{2I_1} + \frac{|x-x'|}{2I_1}$$

Therefore, the system exhibits coupling with parameters $(\frac{I_2}{2I_1}, \frac{1}{2I_1})$. □

We can now finish the proof of Theorem 1. For all $(u,v) \in E$ and $x,y \in \chi$, $|\Phi_{u,v}(x,y)| \leq I_2$. Therefore, for any (u,v), $|Y_{u,v}| = |\Phi_{u,v}(1,1) - \Phi_{u,v}(0,1) - \Phi_{u,v}(1,0) + \Phi_{u,v}(0,0)| \leq 4I_2$.

Note that for all edges, $|Y_e| \leq 4I_2$, so that the condition $\beta(\Delta-1)^2 < 1$ implies $\frac{I_2}{2I_1}(\Delta-1) + \frac{4I_2}{2I_1}(\Delta-1)^2 < 1$. Since $(\Delta-1) \leq (\Delta-1)^2$, if $\beta(\Delta-1)^2 < 1$ we also have $\frac{I_2}{2I_1}(\Delta-1) + \frac{4I_2}{2I_1}(\Delta-1)^2 < 1$. This is exactly condition (23) with a,b as given by Lemma 5 and $K_Y = 4I_2$. It follows that $\mathcal{G}$ exhibits exponential correlation decay, and since Assumptions 1 and 2 hold, all conditions of Corollary 2 are satisfied, and there exists an additive FPTAS for computing $J_{\mathcal{G}}$.

6.3.3 Gaussian distribution. Proof of Theorem 2

In this section, we compute the coupling parameters for Gaussian distributed reward functions. Rather than considering only the assumptions of Theorem 2, we place ourselves in a more general framework. The proof will then follow from the application of Theorem 8 (first condition) and a special case of the computation detailed below (see Corollary 3). Assume that for every edge $e = (u,v)$ the value functions $(\Phi_{u,v}(0,0), \Phi_{u,v}(0,1), \Phi_{u,v}(1,0), \Phi_{u,v}(1,1))$ are independent, identically distributed four-dimensional Gaussian random variables, with mean $\mu = (\mu_i)_{i \in \{00,01,10,11\}}$, and covariance matrix $S = (S_{ij})_{i,j \in \{00,01,10,11\}}$. For every node $v \in V$, suppose $\Phi_v(1) = 0$ and that $\Phi_v(0)$ is a Gaussian random variable with mean μ_p and standard deviation σ_p. Moreover, suppose all the Φ_v and Φ_e are independent for $v \in V$, $e \in E$. Let

$$\sigma_1^2 = S_{10,10} - 2S_{10,11} + S_{11,11} + \sigma_p^2 \qquad \sigma_2^2 = S_{00,00} - 2S_{00,01} + S_{01,01} + \sigma_p^2$$

$$\rho = (\sigma_1\sigma_2)^{-1}(S_{00,10} - S_{00,11} - S_{01,10} + S_{01,11} + \sigma_p^2) \qquad C = \frac{\sigma_2^2 - \sigma_1^2}{\sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\rho^2\sigma_1^2\sigma_2^2}}$$

$$\sigma_X^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 \qquad \sigma_Y^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

Proposition 6. *Assume $C < 1$. Then the network exhibits coupling with parameters (a,b) equal to:*

$$a = \frac{1}{\pi}\arctan\left(\sqrt{\frac{1}{1-C^2}}\frac{\sigma_Y}{\sigma_X}\right) + \sqrt{\frac{2}{\pi}}\frac{|\mu_{00} + \mu_{11} - \mu_{10} - \mu_{01}|}{\sigma_X}$$

$$b = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_X}$$

Corollary 3. *Suppose that for each e, $(\Phi_e(0,0), \Phi_e(0,1), \Phi_e(1,0), \Phi_e(1,1))$ are i.i.d. Gaussian variables with mean 0 and standard deviation σ_e. Let $\beta = \sqrt{\frac{\sigma_e^2}{\sigma_e^2 + \sigma_p^2}}$ Then $a \leq \beta$ and $bK_\Phi \leq \beta$.*

Proof. Under the conditions of corollary 3, we have $\sigma_Y^2 = 4\sigma_e^2$, $\sigma_X^2 = 4\sigma_p^2 + 4\sigma_e^2$, and $C = 0$. Note also that $K_\Phi \leq 2\sigma_e$ By Proposition 6, the network exhibits coupling with parameters

$$a = \frac{1}{\pi} \arctan\left(\sqrt{\frac{\sigma_e^2}{\sigma_e^2 + \sigma_p^2}}\right) \leq \frac{1}{\pi}\beta \leq \beta$$

$$b = \sqrt{\frac{1}{2\pi}} \frac{1}{\sqrt{\sigma_e^2 + \sigma_p^2}} \text{ and so, } bK_\Phi \leq \sqrt{\frac{2}{\pi}}\beta \leq \beta$$

$\square$

Remark that when $\sigma_e \to 0$, $\beta \to 0$ and correlation decay takes place; moreover, combining Corollary 3 and Theorem 8 (condition (22)) directly yields Theorem 2.

Proof of Proposition 6 . Fix an edge (u, v) in E; for simplicity, in the rest of this section denote $\bar{\Phi}^1 = \Phi^1_{u\leftarrow v} + \Phi_v(0) - \Phi_v(1)$ and $\bar{\Phi}^2 = \Phi^2_{u\leftarrow v} + \Phi_v(0) - \Phi_v(1)$. It follows that $(\overline{\Phi}^1, \overline{\Phi}^2)$ follows a bivariate Gaussian distribution with mean (μ_1, μ_2):

$$\mu_1 = \mu_{10} - \mu_{11} + \mu_p \text{ and } \mu_2 = \mu_{00} - \mu_{01} + \mu_p$$

and covariance matrix

$$S_A = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Let $X = \overline{\Phi}^1 + \overline{\Phi}^2$, $Y = \overline{\Phi}^2 - \overline{\Phi}^1$. Then, (X, Y) is a bivariate Gaussian vector with means $\mathbb{E}[X] = \mu_1 + \mu_2$ and $\mathbb{E}[Y] = \mu_2 - \mu_1$, standard deviations σ_X, σ_Y and correlation C as defined previously. Denote also $\overline{X} \overset{\Delta}{=} X - E[X]$ and $\overline{Y} \overset{\Delta}{=} Y - E[Y]$ the centered versions of X and Y. Consider two real numbers $x \geq x'$, and let (b, t) be the two real numbers such that $x = b + t/2$, $x' = b - t/2$. From equation (32), we have

$$(E(x, x'))^c = \{\min(\overline{\Phi}^1, \overline{\Phi}^2) - t/2 < b < \max(\overline{\Phi}^1, \overline{\Phi}^2) + t/2\}$$

The first step of the proof consists in rewriting the event $(E(x, x'))^c$ in terms of the variables X, Y:

Lemma 6.
$$(E(x, x'))^c = \{|Y| \geq |X - 2b| - t\}$$

Proof.

$$\begin{aligned}
(E(x, x'))^c =& \{\min(\overline{\Phi}^1, \overline{\Phi}^2) - t/2 < b < \max(\overline{\Phi}^1, \overline{\Phi}^2) + t/2\} \\
=& \{\overline{\Phi}^1 - t/2 < b < \overline{\Phi}^2 + t/2, \overline{\Phi}^1 \leq \overline{\Phi}^2\} \cup \{\overline{\Phi}^2 - t/2 < b < \overline{\Phi}^1 + t/2, Y \leq 0, \overline{\Phi}^2 \leq \overline{\Phi}^1\} \\
=& \{2\overline{\Phi}^1 - t < 2b < 2\overline{\Phi}^2 + t, \overline{\Phi}^1 \leq \overline{\Phi}^2\} \cup \{2\overline{\Phi}^2 - t < 2b < 2\overline{\Phi}^1 + t, \overline{\Phi}^2 \leq \overline{\Phi}^1\} \\
=& \{X - Y - t < 2b < X + Y + t, Y \geq 0\} \cup \{X + Y - t < 2b < X - Y + t, Y \leq 0\} \\
=& \{(X - 2b) - |Y| - t < 0 < (X - 2b) + |Y| + t\} \\
=& \{|Y| \geq (X - 2b - t)\} \cap \{|Y| \geq (2b - X - t)\} \\
=& \{|Y| \geq |X - 2b| - t\}
\end{aligned}$$

$\square$

For any b and $t \geq 0$, let $S(t) = \{x, y : |y| \geq |x| - t\}$, and for any real x, let $S(t, y) = \{x : |y| \geq |x| - t\}$. Note $S(t, y)$ is symmetric and convex in x for all y. Using the lemma, we obtain:

$$
\begin{aligned}
\mathbb{P}((E)^c(x, x')) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - C^2}} \int_{S(t)} \exp(-\frac{1}{2(1 - C^2)}(\frac{(x - \mu_1 - \mu_2 + 2b)^2}{\sigma_x^2} + \frac{(y - \mu_2 + \mu_1)^2}{\sigma_y^2} \\
&\quad - 2C\frac{(x - \mu_1 - \mu_2 + 2b)(y + \mu_2 - \mu_1)}{\sigma_x\sigma_y}))dxdy \\
&= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - C^2}} \int_y \exp(-\frac{1}{2(1 - C^2)}\frac{(y - \mu_2 + \mu_1)^2}{\sigma_y^2}) \, g(y)dy
\end{aligned}
\tag{33}
$$

where:

$$
g(y) = \int_{x \in S(t,y)} \exp(-\frac{1}{2(1 - C^2)}(\frac{(x - \mu_1 - \mu_2 + 2b)^2}{\sigma_x^2} - 2C\frac{(x - \mu_1 - \mu_2 + 2b)(y - \mu_2 + \mu_1)}{\sigma_x\sigma_y}))dx
$$

Let $\tilde{x}_b = \frac{(x - \mu_1 - \mu_2 + 2b)}{\sigma_x}$ and $\tilde{y} = \frac{(y - \mu_2 + \mu_1)}{\sigma_y}$. Then:

$$
g(y) = \exp(\frac{C^2}{2(1 - C^2)}\tilde{y}^2) \int_{x \in S(t,y)} \exp(-\frac{1}{2(1 - C^2)}(\tilde{x}_b - C\tilde{y})^2)dx
$$

Now, $\tilde{x}_b - C\tilde{y} = \frac{x - \mu_1 - \mu_2 + 2b - \frac{C\sigma_x(y - \mu_2 + \mu_1)}{\sigma_y}}{\sigma_x}$. Recall Anderson's inequality [Dud99]: let γ be a centered Gaussian measure on $\mathbb{R}^k$, and S be a convex, symmetric subset of $\mathbb{R}^k$. Then, for all z, $\gamma(S) \geq \gamma(S + z)$. Since $S(t, y)$ is a convex symmetric subset, by setting $2b = \mu_1 + \mu_2 + \frac{C\sigma_x(y - \mu_2 + \mu_1)}{\sigma_y}$, it follows that

$$
g(y) \leq \exp(\frac{C^2}{2(1 - C^2)}\tilde{y}^2) \int_{x \in S(t,y)} \exp(-\frac{1}{2\sigma_x^2(1 - C^2)}x^2)dx
$$

Injecting that bound in equation (33), we obtain:

$$
\begin{aligned}
\mathbb{P}((E)^c(x, x')) &\leq \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - C^2}} \int_y \exp(-\frac{1}{2(1 - C^2)}\frac{(y - \mu_2 + \mu_1)^2}{\sigma_y^2}) \\
&\quad \left(\exp(\frac{C^2}{2(1 - C^2)}\frac{(y - \mu_2 + \mu_1)^2}{\sigma_y^2}) \int_{x \in S(t,y)} \exp(-\frac{1}{2\sigma_x^2(1 - C^2)}x^2)dx \right) dy \\
&\leq \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - C^2}} \int_{S(t)} \exp(-\frac{1}{2(1 - C^2)}(\frac{x^2}{\sigma_x^2} + (1 - C^2)\frac{(y - \mu_2 + \mu_1)^2}{\sigma_y^2}))dxdy
\end{aligned}
$$

Finally, note that the triangular inequality, for any α we have $S(t) \subset S_\alpha(t) \overset{\Delta}{=} \{(x, y) : |y - \alpha| \geq |x| - t - |\alpha|\}$. We obtain:

$$
\begin{aligned}
\mathbb{P}((E)^c(x, x')) &\leq \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - C^2}} \int_{S_{\mu_2 - \mu_1}(t)} \exp(-\frac{1}{2(1 - C^2)}(\frac{x^2}{\sigma_x^2} + (1 - C^2)\frac{(y - \mu_2 + \mu_1)^2}{\sigma_y^2}))dxdy \\
&\leq \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - C^2}} \int_{S(t + |\mu_2 - \mu_1|)} \exp(-\frac{1}{2(1 - C^2)}(\frac{x^2}{\sigma_x^2} + (1 - C^2)\frac{y^2}{\sigma_y^2}))dxdy
\end{aligned}
$$

where the second inequality follows from a simple change of variable. Let $t' = t + |\mu_2 - \mu_1|$ Finally, we decompose $S(t')$ as the union of two sets: $S(t) = S_{\text{int}}(T) \cup S_{\text{out}}(t)$, where:

$$S_{\text{int}}(t') = \{(X,Y) : |X| < t'\}$$
$$S_{\text{out}}(t') = \{(X,Y) : |X| \geq t' \text{ and } |Y| \geq (|X| - t')\},$$

and note that $S_{\text{int}}(t') \cap S_{\text{out}}(t') = \emptyset$. We have:

$$\mathbb{P}(S_{\text{int}}(t')) \leq \frac{2t'}{\sqrt{2\pi(1-C^2)}\sigma_x}$$

and, by symmetry of $S_{\text{out}}(t')$ in X and Y,

$$\mathbb{P}(S_{\text{out}}(t')) = 4\mathbb{P}(\{(x,y) : x \geq t, y \geq 0, y \geq x - t\})$$
$$= \frac{2}{\pi\sigma_x\sigma_y\sqrt{1-C^2}} \int_{\{(x,y):x \geq t, y \geq 0, y \geq x-t\}} \exp\left(-\frac{1}{2(1-C^2)}\left(\frac{x^2}{\sigma_x^2} + (1-C^2)\frac{y^2}{\sigma_y^2}\right)\right) dxdy$$

Using the change of variables $(x', y') = \left(\frac{x-t}{\sqrt{1-C^2}\sigma_x}, \frac{y}{\sigma_y}\right)$, we get:

$$\mathbb{P}(S_{\text{out}}(t')) = \frac{2}{\pi} \int_{\{(x',y'):x'>0,y'>0,y' \geq \frac{\sigma_x\sqrt{1-C^2}}{\sigma_y}x'\}} \left(\exp\left(-\left(x' + \frac{t'}{\sqrt{1-C^2}\sigma_x}\right)^2 - y'^2\right)\right) dx'dy'$$

Since $\left(x' + \frac{t'}{\sqrt{1-C^2}\sigma_x}\right)^2 \geq x'^2$, it follows that:

$$\mathbb{P}(S_{\text{out}}(t')) \leq \frac{2}{\pi} \int_{\{(x',y'):x'>0,y'>0,y' \geq \frac{\sigma_x\sqrt{1-C^2}}{\sigma_y}x'\}} \left(\exp(-x'^2 - y'^2)\right) dxdy$$

By using a radial change of variables $(x', y') = (r\cos(\theta), r\sin(\theta))$ we can compute exactly the expression above, and find:

$$\mathbb{P}(S_{\text{out}}(t')) \leq \frac{2}{\pi} \int_{\{(r,\theta):r>0,\arctan(\frac{\sigma_x\sqrt{1-C^2}}{\sigma_y}) \leq \theta \leq \frac{\pi}{2}\}} \exp(-r^2) r\, dr\, d\theta$$
$$= \frac{1}{\pi} \arctan\left(\frac{\sigma_y}{\sigma_x\sqrt{1-C^2}}\right)$$

$$\mathbb{P}((E)^c(x,x')) \leq \left(\frac{1}{\pi}\arctan\left(\frac{\sigma_y}{\sigma_x\sqrt{1-C^2}}\right) + \sqrt{\frac{2}{\pi(1-C^2)}}\frac{|\mu_2 - \mu_1|}{\sigma_x}\right) + \sqrt{\frac{2}{\pi(1-C^2)}}\frac{t}{\sigma_x} \qquad (34)$$

which gives us the desired bounds on (a, b). $\qquad \square$

7 Maximum weighted independent sets

7.1 Cavity expansion and the algorithm

In this section, we show how the correlation decay framework also applies to MWIS problems and prove theorems 3, 4, and 5. There are additional challenges in achieving this goal. First,

the bounded costs assumption required for the results of section 5 does not hold for constrained optimization problems, as the underlying problem has infinite costs. Second, the coupling technique of section 6 is not readily applicable for MWIS. We therefore develop a different approach.

As for unconstrained optimization problems, we follow three steps. First, we detail the Cavity Expansion algorithm. Second, we establish the correlation decay property. Finally, we show that the correlation decay property implies that near-optimal, decentralized optimization can be performed in polynomial time.

Consider a general weighted graph $\mathcal{G} = (V, E, W)$, where (V, E) is a graph whose nodes are equipped with arbitrary non-negative weights $W_i, i \in V$; no probabilistic assumption on W_i is adopted yet. Note that for Independent Sets problems, we have $J_\mathcal{G} = W(I^*)$, and for any $(i_1, \ldots, i_d)$, $J_{\mathcal{G},(i_1,\ldots,i_d)}(\mathbf{0}) = J_{\mathcal{G}\setminus\{i_1,\ldots,i_d\}}$, where $\mathcal{G} \setminus \{i_1, \ldots, i_d\}$ is the subgraph induced by nodes $V \setminus \{i_1, \ldots, i_d\}$. Consider a given node $i \in V$ and let $N(i) = \{i_1, \ldots, i_d\}$. From Theorem 6, we have

$$B_{\mathcal{G},i} = J_{\mathcal{G},i}(1) - J_{\mathcal{G},i}(0) = W_i + \sum_l \mu_{i \leftarrow i_l}(1, B_{\mathcal{G}(i,l),i_l}) \tag{35}$$

Recall that for MWIS, we have $\Phi_e(x, y) = -\infty$ for $(x, y) = (1, 1)$ and $\Phi_e(x, y) = 0$, otherwise. Therefore, by definition of $\mu_{i \leftarrow i_l}$, we have

$$\mu_{i \leftarrow i_l}(1, B_{\mathcal{G}(i,l),i_l}) = \max(-\infty + B_{\mathcal{G}(i,l),i_l}, 0) - \max(B_{\mathcal{G}(i,l),i_l}, 0) = -\max(B_{\mathcal{G}(i,l),i_l}, 0)$$

Thus,

$$B_{\mathcal{G},i} = W_i - \sum_{l=1}^{d} \max(B_{\mathcal{G}(i,l),i_l}, 0)$$

Let $l \leq d$; recall the definition of $\mathcal{G}(i, l)$: $\mathcal{G}(i, l)$ is the network $\mathcal{G} \setminus \{i\}$, where the potential functions of the neighbors of i have been modified as follows:

- for $v \in \{i_1, \ldots, i_{l-1}\}$, $\phi'_v(0) = \phi_v(0) + \phi_{i,v}(1, 0) = 0$, and $\phi'_v(1) = W_v + \phi_{i,v}(1, 1) = W_v - \infty = -\infty$. Since the new weight of v is $-\infty$, it is equivalent to removing this node from the graph.

- for $v \in \{i_{l+1}, \ldots, i_d\}$, $\phi'_v(0) = \phi_v(0) + \phi_{i,v}(0, 0) = 0$, and $\phi'_v(1) = W_v + \phi_{i,v}(0, 1) = W_v$

We thus observe that in $\mathcal{G}(i, l)$, the nodes $\{i, i_1, \ldots, i_{l-1}\}$ can be removed, while the weights of nodes $\{i_{l+1}, \ldots, i_d\}$ are unchanged; equivalently, we have $\mathcal{G}(i, l) = \mathcal{G} \setminus \{i, i_1, \ldots, i_{l-1}\}$. Therefore, we obtain

$$B_{\mathcal{G},i} = W_i - \sum_{l=1}^{d} \max(B_{\mathcal{G}\setminus\{i,i_1,\ldots,i_{l-1}\}}, 0)$$

We further modify the cavity recursion by the following change of variable: for any graph $\mathcal{G}$ and node i, let $C_\mathcal{G}(i) = \max(B_{\mathcal{G},i}, 0)$; note we have $C_\mathcal{G}(i) = \max(J_{\mathcal{G},i}(1), J_{\mathcal{G},i}(0)) - J_{\mathcal{G},i}(0) = J_\mathcal{G} - J_{\mathcal{G}\setminus\{i\}}$. The variables C will be called *cavities*. It turns out that in the case of IS problems, working with cavities C is more convenient than with cavities B. We obtain the cavity recursion for MWIS:

Proposition 7. *For any $i \in V$, let $N(i) = \{i_1, \ldots, i_d\}$. Then*

$$C_\mathcal{G}(i) = \max\left(0, W_i - \sum_{1 \leq l \leq d} C_{\mathcal{G}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l)\right), \tag{36}$$

31

where $\sum_{1 \leq l \leq d} = 0$ when $N(i) = \emptyset$. If $W_i - \sum_{1 \leq l \leq d} C_{\mathcal{G} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l) > 0$, namely $C_{\mathcal{G}}(i) > 0$, then every largest weight independent set must contain i. Similarly if $W_i - \sum_{1 \leq l \leq d} C_{\mathcal{G} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l) < 0$, implying $C_{\mathcal{G}}(i) = 0$, then every largest weight independent set does not contain i.

Remark : The proposition leaves out a "fuzzy" case $W_i - \sum_{1 \leq l \leq d} C_{\mathcal{G} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l) = 0$. 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 $C_{\mathcal{G}}(i) > 0$ ($C_{\mathcal{G}}(i) = 0$) determine whether i must (must not) belong to the largest weighted independent set.

Using the special form of the cavity recursion (36), the cavity expansion algorithm for MWIS is very similar as the one defined in section 4.3. For any induced subgraph $\mathcal{H}$ of $\mathcal{G}$ and node i, let $C_{\mathcal{H}}^-(i, r) = \max(0, \text{CE}[\mathcal{H}, i, r])$ with boundary condition $\text{CE}[\mathcal{H}, i, 0] = 0$, and let $C_{\mathcal{H}}^+(i, r)$ be the same quantity for the boundary condition $\text{CE}[\mathcal{H}, i, 0] = W_i$. Alternatively, C^- and C^+ can be defined by the following recursions:

$$C_{\mathcal{H}}^-(i, r) = \begin{cases} 0, & r = 0; \\ \max\left(0, W_i - \sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^-(i_l, r - 1)\right), & r \geq 1. \end{cases} \tag{37}$$

$$C_{\mathcal{H}}^+(i, r) = \begin{cases} W_i, & r = 0; \\ \max\left(0, W_i - \sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^+(i_l, r - 1)\right), & r \geq 1. \end{cases} \tag{38}$$

The two boundaries condition were chosen so that $C_{\mathcal{H}}^-(i, r)$ and $C_{\mathcal{H}}^+(i, r)$ provide valid bounds on the true cavities $C_{\mathcal{H}}(i)$, as detailed by the following Lemma.

Lemma 7. *For every even r*

$$C_{\mathcal{H}}^-(i, r) \leq C_{\mathcal{H}}(i) \leq C_{\mathcal{H}}^+(i, r),$$

and for every odd r

$$C_{\mathcal{H}}^+(i, r) \leq C_{\mathcal{H}}(i) \leq C_{\mathcal{H}}^-(i, r),$$

Proof. The proof is by induction in r. The assertion holds by definition of C^-, C^+ for $t = 0$. The induction follows from (36), definitions of C^-, C^+ and since the function $x \to \max(0, W - x)$ is non-increasing. $\qquad \square$

We now describe our algorithm 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 $\mathcal{G} = (V, E)$ and delete every node (and incident edges) with probability $\epsilon^2/16$, independently for all nodes. We denote the resulting (random) subgraph by $\mathcal{G}(\epsilon)$. In the second stage we compute $C_{\mathcal{G}(\epsilon)}^-(i, r)$ for every node i for the graph $\mathcal{G}(\epsilon)$ for some target even number of steps r. We set $\mathcal{I}(r, \epsilon) = \{i : C_{\mathcal{G}(\epsilon)}^-(i, r) > 0\}$. Let I_ϵ^* be the largest weighted independent set of $\mathcal{G}(\epsilon)$.

Lemma 8. $\mathcal{I}(r, \epsilon)$ *is an independent set.*

Proof. By Lemma 7, if $C_{\mathcal{G}(\epsilon)}^-(i, r) > 0$ then $C_{\mathcal{G}(\epsilon)} > 0$, and therefore $\mathcal{I} \subset I_\epsilon^*$. Thus our algorithm produces an independent set in $\mathcal{G}(\epsilon)$ and therefore in $\mathcal{G}$. $\qquad \square$

We finish this section by mentioning that due to Proposition 4, the complexity of running both stages of the algorithm is $O(nr\Delta^r)$. As it will be apparent from the analysis, we could take $C_{\mathcal{G}(\epsilon)}^+$ instead of $C_{\mathcal{G}(\epsilon)}^-$ and arrive at the same result using an odd number r.

7.2 Proof of Theorem 3

7.2.1 Correlation decay property

The main bulk of the proof of Theorem 3 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 be then 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 $\mathcal{H}$ of $\mathcal{G}(\epsilon)$, and any node i in $\mathcal{H}$, introduce $M_{\mathcal{H}}(i) = \mathbb{E}[\exp(-C_{\mathcal{H}}(i))], M^-_{\mathcal{H}}(i, r) = \mathbb{E}[\exp(-C^-_{\mathcal{H}}(i, r))], M^+_{\mathcal{H}}(i, r) = \mathbb{E}[\exp(-C^+_{\mathcal{H}}(i, r))]$.

Proposition 8. *Let $\mathcal{G}(\epsilon) = (V_\epsilon, E_\epsilon)$ 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 $\delta = \epsilon^2/16$ independently for all nodes). Then, for every node i in $\mathcal{G}(\epsilon)$ and every r*

$$\mathbb{P}(C_{\mathcal{G}(\epsilon)}(i) = 0, C^+_{\mathcal{G}(\epsilon)}(i, 2t) > 0) \leq 3(1 - \epsilon^2/16)^{2r}, \tag{39}$$

and

$$\mathbb{P}(C_{\mathcal{G}(\epsilon)}(i) > 0, C^-_{\mathcal{G}(\epsilon)}(i, 2t) = 0) \leq 3(1 - \epsilon^2/16)^{2r}. \tag{40}$$

Proof. Consider a subgraph $\mathcal{H}$ of $\mathcal{G}$, node $i \in \mathcal{H}$ with neighbors $\mathcal{N}_{\mathcal{H}}(i) = \{i_1, \ldots, i_d\}$, and suppose for now that the number of neighbors of i in $\mathcal{G}$ is less than 2.

Examine the recursion (36) and observe that all the randomness in terms $C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l)$ comes from the subgraph $\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}$, and thus W_j is independent from the vector $(C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l), 1 \leq l \leq d)$. A similar assertion applies when we replace $C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l)$ with $C^-_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l, r)$ and $C^+_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l, r)$ for every r. Using the memoryless property of the exponential distribution, denoting W a standard exponential random variable, we obtain:

$$\mathbb{E}[\exp(-C_{\mathcal{H}}(i))| \sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l) = x] = \mathbb{P}(W_i \leq x)E[\exp(0)]+$$

$$\mathbb{E}[\exp(-(W_i - x)) \,|\, W_i > x]\mathbb{P}(W_i > x)$$
$$= (1 - \mathbb{P}(W_i > x)) + \mathbb{E}[\exp(-W)]\mathbb{P}(W_i > x)$$
$$= (1 - \mathbb{P}(W_i > x)) + (1/2)\mathbb{P}(W_i > x)$$
$$= 1 - (1/2)\mathbb{P}(W_i > x)$$
$$= 1 - (1/2)\exp(-x) \tag{41}$$

It follows

$$\mathbb{E}[\exp(-C_{\mathcal{H}}(i))] = 1 - (1/2)\mathbb{E}\exp\left(-\sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l)\right).$$

Similarly we obtain

$$\mathbb{E}[\exp(-C^-_{\mathcal{H}}(i, r))] = 1 - (1/2)\mathbb{E}\exp\left(-\sum_{1 \leq l \leq d} C^-_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l, r - 1)\right),$$

$$\mathbb{E}[\exp(-C^+_{\mathcal{H}}(i, r))] = 1 - (1/2)\mathbb{E}\exp\left(-\sum_{1 \leq l \leq d} C^+_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l, r - 1)\right).$$

Since i had two neighbors or less in $\mathcal{G}$, it also has two neighbors or less in $\mathcal{H}$. For $d = 0$, we have trivially $M_{\mathcal{H}}(i) = M_{\mathcal{H}}^-(i) = M_{\mathcal{H}}^+(i)$. Suppose $d = 1 : N_{\mathcal{H}}(i) = \{i_1\}$. Then,

$$M_{\mathcal{H}}^-(i,r) - M_{\mathcal{H}}^+(i,r) = (1/2)\Big(\mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{i\}}^+(i_1,r-1))] - \mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{i\}}^-(i_1,r-1))]\Big)$$
$$= (1/2)\Big(M_{\mathcal{H}\setminus\{i\}}^+(i_1,r-1) - M_{\mathcal{H}\setminus\{i\}}^-(i_1,r-1)\Big) \qquad (42)$$

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

$$M_{\mathcal{H}}^-(i,r) - M_{\mathcal{H}}^+(i,r)$$
$$= (1/2)\mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{i\}}^+(i_1,r-1) - C_{\mathcal{H}\setminus\{i,i_1\}}^+(i_2,r-1))]$$
$$- (1/2)\mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{i\}}^-(i_1,r-1) - C_{\mathcal{H}\setminus\{i,i_1\}}^-(i_2,r-1))]$$
$$= (1/2)\mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{i\}}^+(i_1,r-1))(\exp(-C_{\mathcal{H}\setminus\{i,i_1\}}^+(i_2,r-1)) - \exp(-C_{\mathcal{H}\setminus\{i,i_1\}}^-(i_2,r-1)))]$$
$$+ (1/2)\mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{i,i_1\}}^-(i_2,r-1))(\exp(-C_{\mathcal{H}\setminus\{i\}}^+(i_1,r-1)) - \exp(-C_{\mathcal{H}\setminus\{i\}}^-(i_1,r-1)))]$$

Using the non-negativity of C^-, C^+ and applying Lemma 7 we obtain for odd r

$$0 \le M_{\mathcal{H}}^+(i,r) - M_{\mathcal{H}}^-(i,r) \le (1/2)\mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{i,i_1\}}^-(i_2,r-1)) - \exp(-C_{\mathcal{H}\setminus\{i,i_1\}}^+(i_2,r-1))]$$
$$+ (1/2)\mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{i\}}^-(i_1,r-1)) - \exp(-C_{\mathcal{H}\setminus\{i\}}^+(i_1,r-1))],$$
$$= (1/2)\big(M_{\mathcal{H}\setminus\{i,i_1\}}^-(i_2,r-1) - M_{\mathcal{H}\setminus\{i,i_1\}}^+(i_2,r-1)\big)$$
$$+ (1/2)\big(M_{\mathcal{H}\setminus\{i\}}^-(i_1,r-1) - M_{\mathcal{H}\setminus\{i\}}^+(i_1,r-1)\big) \qquad (43)$$

and for even r

$$0 \le M_{\mathcal{H}}^-(i,r) - M_{\mathcal{H}}^+(i,r) \le (1/2)\big(M_{\mathcal{H}\setminus\{i,i_1\}}^+(i_2,r-1) - M_{\mathcal{H}\setminus\{i,i_1\}}^-(i_2,r-1)\big)$$
$$+ (1/2)\big(M_{\mathcal{H}\setminus\{i\}}^+(i_1,r-1)] - M_{\mathcal{H}\setminus\{i\}}^-(i_1,r-1)\big) \qquad (44)$$

Summarizing the three cases we conclude

$$|M_{\mathcal{H}}^+(i,r) - M_{\mathcal{H}}^-(i,r)| \le (d/2)\max_{\mathcal{H}',j}\Big|M_{\mathcal{H}'}^+(j,r-1) - M_{\mathcal{H}'}^-(j,r-1)\Big|, \qquad (45)$$

where the maximum is over subgraphs $\mathcal{H}'$ of $\mathcal{G}$ and nodes $j \in \mathcal{H}'$ with degree at most 2 in $\mathcal{H}'$. The reason for this is that in equations (42),(43), and (44); the moments $M_{\mathcal{H}'}^+(j,r-1)$ in the right hand side are always computed in a node j which has lost at least one of its neighbors (namely, i) in graph $\mathcal{H}$. Since the degree of j was at most 3 in $\mathcal{G}$ and one neighbor at least is removed, j has at most two neighbors in $\mathcal{H}$. By considering $\mathcal{H} \cap \mathcal{G}(\epsilon)$ in all previous equations, equation (45) implies

$$|M_{\mathcal{H}\cap\mathcal{G}(\epsilon)}^+(i,r) - M_{\mathcal{H}\cap\mathcal{G}(\epsilon)}^-(i,r)| \le (d(\epsilon)/2)\max_{\mathcal{H}',j}\Big|M_{\mathcal{H}'\cap\mathcal{G}(\epsilon)}^+(j,r-1) - M_{\mathcal{H}'\cap\mathcal{G}(\epsilon)}^-(j,r-1)\Big|, \qquad (46)$$

where $d(\epsilon)$ denotes the number of neighbors of i in $\mathcal{H} \cap \mathcal{G}(\epsilon)$. By definition of $\mathcal{G}(\epsilon)$, $d(\epsilon)$ is a binomial random variable with d trials and probability of success $(1 - \epsilon^2/16)$, where d is the degree of i in $\mathcal{H}$. Since $d \le 2$, $E[d(\epsilon)] \le 2(1 - \epsilon^2/16)$. Moreover, this randomness is independent from the randomness of the random weights of $\mathcal{H}$. Therefore,

$$\mathbb{E}[|M_{\mathcal{H}\cap\mathcal{G}(\epsilon)}^+(i,r) - M_{\mathcal{H}\cap\mathcal{G}(\epsilon)}^-(i,r)|] \le (1 - \epsilon^2/16)\max_{\mathcal{H},j}\mathbb{E}\Big|M_{\mathcal{H}\cap\mathcal{G}(\epsilon)}^+(j,r-1) - M_{\mathcal{H}\cap\mathcal{G}(\epsilon)}^-(j,r-1)\Big|$$
$$\qquad (47)$$

where the external expectation is w.r.t. the randomness of the first phase of the algorithm (deleted nodes). Let e_{r-1} denote the right-hand side of (47). By taking the max of the left-hand side of (47) over all $(\mathcal{H}, j)$ where j has degree less than or equal to 2 in $\mathcal{H}$, we obtain the inequality $e_r \leq (1 - \epsilon^2/16)e_{r-1}$. Iterating on r and using $0 \leq M \leq 1$, this implies that $e_r \leq (1 - \epsilon^2/16)^r$ for all $r \geq 0$. Finally, it is easy to show using the same techniques that equation (45) holds for $r = 3$ as well. This finally implies that for an arbitrary node i in $\mathcal{G}(\epsilon)$

$$\mathbb{E}[|M^+_{\mathcal{G}(\epsilon)}(i, r) - M^-_{\mathcal{G}(\epsilon)}(i, r)|] \leq 3/2(1 - \epsilon^2/16)^r.$$

Applying Lemma 7, we conclude for every r

$$0 \leq \mathbb{E}[\exp(-C^-_{\mathcal{G}(\epsilon)}(i, 2r)) - \exp(-C^+_{\mathcal{G}(\epsilon)}(i, 2r))] \leq 3/2(1 - \epsilon^2/16)^{2r}.$$

Recalling (41) we have

$$\mathbb{E}[\exp(-C_{\mathcal{G}(\epsilon)}(i))] = 1 - (1/2)\mathbb{P}(W > \sum_{1 \leq l \leq d} C_{\mathcal{G}(\epsilon) \backslash \{i, i_1, \dots, i_{l-1}\}}(i_l)) = 1 - (1/2)\mathbb{P}(C_{\mathcal{G}(\epsilon)}(i) > 0).$$

Similar expressions are valid for $C^-_{\mathcal{G}(\epsilon)}(i, r)), C^+_{\mathcal{G}(\epsilon)}(i, r))$. We obtain

$$0 \leq \mathbb{P}(C^+_{\mathcal{G}(\epsilon)}(i, 2r) = 0) - \mathbb{P}(C^-_{\mathcal{G}(\epsilon)}(i, 2r) = 0) \leq 3(1 - \epsilon^2/16)^{2r}.$$

Again applying Lemma 7, we obtain

$$\mathbb{P}(C_{\mathcal{G}(\epsilon)}(i) = 0, C^+_{\mathcal{G}(\epsilon)}(i, 2r) > 0) \leq \mathbb{P}(C^-_{\mathcal{G}(\epsilon)}(i, 2r) = 0, C^+_{\mathcal{G}(\epsilon)}(i, 2r) > 0) \leq 3(1 - \epsilon^2/16)^{2r},$$

and

$$\mathbb{P}(C_{\mathcal{G}(\epsilon)}(i) > 0, C^-_{\mathcal{G}(\epsilon)}(i, 2r) = 0) \leq \mathbb{P}(C^-_{\mathcal{G}(\epsilon)}(i, 2r) = 0, C^+_{\mathcal{G}(\epsilon)}(i, 2r) > 0) \leq 3(1 - \epsilon^2/16)^{2r}.$$

This completes the proof of the proposition. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

7.2.2 Concentration argument

We can now complete the proof of Theorem 3. We need to bound $|W(I^*) - W(I^*_\epsilon)|$ and $W(I^*_\epsilon \backslash \mathcal{I}(r, \epsilon))$ and show that both quantities are small.

Let ΔV_ϵ be the set of nodes in $\mathcal{G}$ which are not in $\mathcal{G}(\epsilon)$. Trivially, $|W(I^*) - W(I^*_\epsilon)| \leq W(\Delta V_\epsilon)$. We have $\mathbb{E}[\Delta V_\epsilon] = \epsilon^2/16n$, and since the nodes were deleted irrespectively of their weights, then $\mathbb{E}[W(\Delta V_\epsilon)] = \epsilon^2/16n$.

To analyze $W(I^*_\epsilon \backslash \mathcal{I}(r, \epsilon))$, observe that by (second part of) Proposition 8, for every node $i, \mathbb{P}(i \in I^*_\epsilon \backslash \mathcal{I}(r, \epsilon)) \leq 3(1 - \epsilon^2/16)^r \overset{\Delta}{=} \delta_1$. Thus $\mathbb{E}|I^*_\epsilon \backslash \mathcal{I}(r, \epsilon)| \leq \delta_1 n$. In order to obtain a bound on $W(I^*_\epsilon \backslash \mathcal{I}(r, \epsilon))$ 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 $\mathcal{G}(\epsilon)$ with weights greater than C. We have $\mathbb{E}[W(V_C)] \leq (C + E[W - C|W > C]) \exp(-C)n = (C + 1) \exp(-C)n$. The remaining nodes have a weight at most C. Therefore,

$$\mathbb{E}[W(I^*_\epsilon \backslash \mathcal{I}(r, \epsilon))] \leq \mathbb{E}[W\Big(\big((I^*_\epsilon \backslash \mathcal{I}(r, \epsilon)) \cap V_C\big) \cup V^c_C\Big)] \leq C\mathbb{E}[|I^*_\epsilon \backslash \mathcal{I}(r, \epsilon)|] + \mathbb{E}[V_C]$$
$$\leq C\delta_1 n + (C + 1) \exp(-C)n.$$

We conclude

$$\mathbb{E}[|W(I^*) - W(\mathcal{I}(r,\epsilon))|] \leq \epsilon^2/16n + C\delta_1 n + (C+1)\exp(-C)n. \tag{48}$$

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 the algorithm ignores the weights, then also the expected weight of this set is at least $n/4$. The variance of that weight is upper bounded by n. By Chebyshev's inequality

$$\mathbb{P}(W(I^*) < n/8) \leq \frac{n}{(n/4 - n/8)^2} = 64/n.$$

We now summarize the results.

$$\begin{aligned}
\mathbb{P}\left(\frac{W(\mathcal{I}(r,\epsilon))}{W(I^*)} \leq 1 - \epsilon\right) &\leq \mathbb{P}\left(\frac{W(\mathcal{I}(r,\epsilon))}{W(I^*)} \leq 1 - \epsilon, W(I^*) \geq n/8\right) + \mathbb{P}(W(I^*) < n/8) \\
&\leq \mathbb{P}\left(\frac{|W(I^*) - W(\mathcal{I}(r,\epsilon))|}{W(I^*)} \geq \epsilon, W(I^*) \geq n/8\right) + 64/n \\
&\leq \mathbb{P}\left(\frac{|W(I^*) - W(\mathcal{I}(r,\epsilon))|}{n/8} \geq \epsilon\right) + 64/n \\
&\leq \frac{\epsilon^2/16 + 4C(1 - \epsilon^2/16)^r + (C+1)\exp(-C)}{\epsilon/8} + 64/n,
\end{aligned}$$

where we have used Markov's inequality in the last step and $\delta_1 = 4(1 - \epsilon^2/16)^r$. Thus it suffices to arrange C so that the first ratio is at most $2\epsilon/3$ and assuming, without the loss of generality, that $n \geq 192/\epsilon$, we will obtain that the sum is at most ϵ. It is a simple exercise to show that by taking $r = O(\log(1/\epsilon)/\epsilon^2)$ and $C = O(\log(1/\epsilon))$, we obtain the required result. This completes the proof of Theorem 3. $\qquad\square$

7.3 Generalization to higher degrees. Proof of Theorem 4

In this section we present the proof of Theorem 4. 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 exponentially distributed random variable with rate α_j, conditional on the rate being α_j.

For every subgraph $\mathcal{H}$ of $\mathcal{G}$, node i in $\mathcal{H}$ and $j = 1, \ldots, \Delta$, define $M_{\mathcal{H}}^j(i) = \mathbb{E}[\exp(-\alpha_j C_{\mathcal{H}}(i))]$, $M_{\mathcal{H}}^{-,j}(i,r) = \mathbb{E}[\exp(-\alpha_j C_{\mathcal{H}}^-(i,r))]$ and $M_{\mathcal{H}}^{+,j}(i,r) = \mathbb{E}[\exp(-\alpha_j C_{\mathcal{H}}^+(i,r))]$, where $C_{\mathcal{H}}(i), C_{\mathcal{H}}^+(i,r))$ and $C_{\mathcal{H}}^-(i,r))$ are defined as in Section 7.1.

Lemma 9. *Fix any subgraph $\mathcal{H}$, node $i \in \mathcal{H}$ with $N_{\mathcal{H}}(i) = \{i_1, \ldots, i_d\}$. Then*

$$\begin{aligned}
\mathbb{E}[\exp(-\alpha_j C_{\mathcal{H}}(i))] &= 1 - \frac{1}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \mathbb{E}[\exp(-\sum_{1 \leq l \leq d} \alpha_k C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l))] \\
\mathbb{E}[\exp(-\alpha_j C_{\mathcal{H}}^+(i,r))] &= 1 - \frac{1}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \mathbb{E}[\exp(-\sum_{1 \leq l \leq d} \alpha_k C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^+(i_l, r-1))] \\
\mathbb{E}[\exp(-\alpha_j C_{\mathcal{H}}^-(i,r))] &= 1 - \frac{1}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \mathbb{E}[\exp(-\sum_{1 \leq l \leq d} \alpha_k C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^-(i_l, r-1))]
\end{aligned}$$

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 $\sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l) = x$. As $C_{\mathcal{H}}(i) = \max(0, W_i - x)$, we obtain:

$$
\begin{aligned}
\mathbb{E}[-\alpha_j C_{\mathcal{H}}(i)|x] &= \frac{1}{\Delta} \sum_k \mathbb{E}[-\alpha_j C_{\mathcal{H}}(i)|x, \alpha(i) = \alpha_k] \\
&= \frac{1}{\Delta} \sum_k \Big(\mathbb{P}(W_i \leq x|\alpha(i) = \alpha_k) \\
&\quad + \mathbb{P}(W_i > x|\alpha(i) = \alpha_k)\mathbb{E}[\exp(-\alpha_j(W_i - x))|W_i > x, \alpha(i) = \alpha_k]\Big) \\
&= \frac{1}{\Delta} \sum_k \left(1 - \exp(-\alpha_k x) + \exp(-\alpha_k x)\frac{\alpha_k}{\alpha_j + \alpha_k} \right) \\
&= 1 - \frac{1}{\Delta} \sum_k \frac{\alpha_j}{\alpha_j + \alpha_k} \exp(-\alpha_k x)
\end{aligned}
$$

Thus,

$$
\mathbb{E}[-\alpha_j C_{\mathcal{H}}(i)] = 1 - \frac{1}{\Delta} \sum_k \frac{\alpha_j}{\alpha_j + \alpha_k} \mathbb{E}[\exp(- \sum_{1 \leq l \leq d} \alpha_k C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}(i_l))]
$$

The other equalities follow identically. $\qquad\square$

By taking differences, we obtain

$$
M_{\mathcal{H}}^{-,j}(i, r) - M_{\mathcal{H}}^{+,j}(i, r) =
$$
$$
\frac{1}{\Delta} \sum_k \frac{\alpha_j}{\alpha_j + \alpha_k} \left(\mathbb{E}[\prod_{1 \leq l \leq d} \exp(-\alpha_k C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^{+}(i_l, r - 1))] - \mathbb{E}[\prod_{1 \leq l \leq d} \exp(-\alpha_k C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^{-}(i_l, r - 1))] \right)
$$

We now use the identity

$$
\prod_{1 \leq l \leq r} x_l - \prod_{1 \leq l \leq r} y_l = \sum_{1 \leq l \leq r} \left((\prod_{1 \leq k \leq l-1} x_k)(x_l - y_l)(\prod_{l+1 \leq k \leq r} y_k) \right),
$$

which further implies

$$
\Big| \prod_{1 \leq l \leq r} x_l - \prod_{1 \leq l \leq r} y_l \Big| \leq \sum_{1 \leq l \leq r} |x_l - y_l|,
$$

when $\max_l |x_l|, |y_l| < 1$. By applying this inequality with $x_l = \exp(-\alpha_k C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^{+}(i_l, r - 1))$ and $y_l = \exp(-\alpha_k C_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^{-}(i_l, r - 1))$, we obtain

$$
|M_{\mathcal{H}}^{-,j}(i, r) - M_{\mathcal{H}}^{+,j}(i, r)| \leq \frac{1}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \sum_{1 \leq l \leq d} |M_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^{-,k}(i_l, r - 1) - M_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^{+,k}(i_l, r - 1)|.
$$

This implies

$$
|M_{\mathcal{H}}^{-,j}(i, r) - M_{\mathcal{H}}^{+,j}(i, r)| \leq \frac{r}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \max_{1 \leq l \leq d} |M_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^{-,k}(i_l, r - 1) - M_{\mathcal{H} \setminus \{i, i_1, \ldots, i_{l-1}\}}^{+,k}(i_l, r - 1)|. \tag{49}
$$

For any $t \geq 0$ and j, define $e_{r,j}$ as follows

$$e_{r,j} = \sup_{\mathcal{H} \subset \mathcal{G}, i \in \mathcal{H}} |M_{\mathcal{H}}^{-,j}(i,r) - M_{\mathcal{H}}^{+,j}(i,r)| \tag{50}$$

By taking maximum on the right and left hand side successively, inequality (49) implies

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

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

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

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 3, this will imply that for each node, the error of a decision made in $\mathcal{I}(r,0)$ is exponentially small in r. Note that $\frac{r}{\Delta} \leq 1$. Recall that $\alpha_j = \rho^j$. Therefore, for each j,k, we have $M_{j,k} \leq \frac{\rho^j}{\rho^j + \rho^k}$. Define M_Δ to be a $\Delta \times \Delta$ 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 $1 \leq j,k \leq \Delta$. Since $M \leq M_\Delta$, it suffices to show that M_Δ^r converges to zero exponentially fast. Proof of theorem 4 will thus be completed with the proof of the following lemma:

Lemma 10. *Under the condition $\rho > 25$, there exists $\delta = \delta(\rho) < 1$ such that the absolute value of every entry of M_Δ^r is at most $\delta^r(\rho)$.*

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

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

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

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

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

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

This completes the proof of the lemma and of the theorem. □

7.4 Hardness result. Proof of Theorem 5

The main idea of the proof is to show that the weight of a maximum weighted independent set is close to the cardinality of a maximum independent set. 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 9. *There exist Δ_0 and c^* such that for all $\Delta \geq \Delta_0$ 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 the ratio of the expected weight of a maximum weight independent set to the cardinality of a maximum independent set grows as the logarithm of the maximum degree of the graph.

Proposition 9. *Suppose $\Delta \geq 2$. For every graph $\mathbb{G}$ with maximum degree Δ and n large enough, we have:*

$$1 \leq \frac{E[W(I^*)]}{|I^M|} \leq 10 \log \Delta.$$

This in combination with Theorem 9 leads to the desired result.

Proof. Let $W(1) < W(2) < \cdots < W(n)$ be the ordered weights associated with our graph $\mathbb{G}$. Observe that

$$
\begin{aligned}
E[W(I^*)] &= E[\sum_{v \in I^*} W_v] \\
&\leq E[\sum_{n-|I^*|+1}^{n} W(i)] \\
&\leq E[\sum_{n-|I^M|+1}^{n} W(i)].
\end{aligned}
$$

The exponential distribution implies $E[W(j)] = H(n) - H(n-j)$, where $H(k)$ is the harmonic sum $\sum_{1 \leq i \leq k} 1/i$. Thus

$$
\begin{aligned}
\sum_{j=n-|I^M|+1}^{n} E[W(j)] &= \sum_{n-|I^M|+1 \leq j \leq n} (H(n) - H(n-j)) \\
&= |I^M|H(n) - \sum_{j \leq |I^M|-1} H(j).
\end{aligned}
$$

We use the bound $\log(k) \le H(k) - \gamma \le \log(k) + 1$, where $\gamma \approx .57$ is Euler's constant. Then

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

where the bound $|I^M| \ge n/(\Delta+1)$ (obtained by using the greedy algorithm, see Section 7.2.2) is used. Again using the bound $|I^M| \ge n/(\Delta+1)$, we find that $\frac{E[W(I^*)]}{|I^M|} \le \log(\Delta+1)+3+o(1)$. Since $E[W(I^*)] \ge E[W(I^M)] = |I^M|$, it follows that for all sufficiently large n, $1 \le \frac{E[W(I^*)]}{|I^M|} \le \log(\Delta+1)+4$. The proposition follows since for all $\Delta \ge 2$ we have $\log(\Delta+1)+4 \le 10\log\Delta$. $\quad\square$

8 Conclusion

We considered an optimization model which encompasses many models from the literature including graphical models, combinatorial optimization and economics. In our model, cooperating agents within a networked structure choose decisions from a finite set of actions and seek to collectively optimize a global welfare objective function, which can be additively decomposed on the nodes and edges of the network. The main goal is to answer whether it's possible to find near optimal solutions efficiently, and if possible using distributed algorithms relying only on local information. Despite the apparent NP-hardness of such a problem even in the approximation setting, we find that in a framework where cost functions are random, this goal is often achievable. Specifically, we have constructed a general purpose algorithm Cavity Expansion, which relies on the local information only, and thus is distributed. We have established that under the so-called correlation decay property, our algorithm finds a near optimal solution with high probability. We have identified a variety of models which exhibit the correlation decay property and we have proposed general purpose techniques, such as the coupling technique, which we used to prove the correlation decay property.

Our results highlight interesting and intriguing connections between the fields of complexity of algorithms for combinatorial optimization problems and statistical physics, specifically the cavity method and the issues of long-range independence. For example in the special case of the MWIS problem we showed that the problem admits a PTAS, provided by the CE algorithm, for certain node weight distribution, even though the maximum cardinality version of the same problem is known to be non-approximable unless P=NP.

It would be interesting to see what weight distribution are amenable to the approach proposed in this paper. For example, one could consider the case of Bernoulli weights and see whether the correlation decay property breaks down precisely when the approximation becomes NP-hard. Furthermore, it would be interesting to see if the random weights assumption for general decision networks can be substituted with deterministic weights which have some random like properties,

in a fashion similar to the study of pseudo-random graphs. This would move our approach even closer to the worst-case combinatorial optimization setting.

The framework studied here can be further extended in several additional ways. First, we can consider a network of agents who, instead of cooperating, behave selfishly. Using ideas similar to those presented in this paper, we believe it is possible to identify settings where using a distributed procedures representing communication between the agents, one can find in polynomial time Nash equilibrium of the underlying system. Second, one can consider a dynamical setting where agents take repeated actions that affect both their reward and their future state. This class of models, known as factored Markov Decision Processes, has a very large number of applications (supply chain, communication networks, and many others), but optimality bounds have been identified only in very restricted settings. Again, concepts such as correlation decay may be found useful to approach these problems and identify new settings where the solution can be found in polynomial time, despite the curse of dimensionality typically exhibited by these models.

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.

[BBCZ08]   M. Bayati, C. Borgs, J. Chayes, and R. Zecchina, *On the exactness of the cavity method for weighted b-matchings on arbitrary graphs and its relation to linear programs*, Journal of Statistical Mechanics: Theory and Experiment **6001** (2008), 1–10.

[BG08]   A. Bandyopadhyay and D. Gamarnik, *Counting without sampling. Asymptotics of the log-partition function for certain statistical physics models.*, Random Structures and Algorithms **33** (2008), no. 4.

[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.

[BSS08]   M. Bayati, D. Shah, and M. Sharma, *Max-Product for Maximum Weight Matching: Convergence, Correctness, and LP Duality*, IEEE Transactions on Information Theory **54** (2008), no. 3, 1241–1251.

[Che08]   M. Chertkov, *Exactness of belief propagation for some graphical models with loops*, Journal of Statistical Mechanics: Theory and Experiment **2008** (2008), P10016.

[Dud99]   R.M. Dudley, *Uniform central limit theorems*, Cambridge university press, 1999.

[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.

[GSW09]   D. Gamarnik, D. Shah, and Y. Wei, *Belief propagation for min-cost network flow: Convergence and correctness*, To appear in proceedings of the 2010 ACM-SIAM symposium on Discrete algorithm, 2009.

[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.

[Hoc97]   D. Hochbaum, *Approximation algorithms for NP-hard problems*, WS Publishing Company, Boston, MA, 1997.

[HW05]   A. Hartmann and M. Weigt, *Phase transitions in combinatorial optimization problems: basics, algorithms and statistical mechanics*, Vch Verlagsgesellschaft Mbh, 2005.

[Jor04]   M. Jordan, *Graphical models*, Statistical Science (Special Issue on Bayesian Statistics) **19** (2004), 140–155.

[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.

[Lau96]   S.L. Lauritzen, *Graphical models*, Oxford University Press, USA, 1996.

[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.

[Mar55]   J. Marschak, *Elements for a theory of teams*, Management Science **1** (1955), no. 2, 127–137.

[MM08]   M. Mezard and A. Montanari, *Information, physics and computation*, Oxford: Oxford University Press, 2008.

[MP03]   M. Mézard and G. Parisi, *The Cavity Method at Zero Temperature*, Journal of Statistical Physics **111** (2003), no. 1, 1–34.

[MR72]   J. Marschak and R. Radner, *Economic theory of teams*, Yale Univ. Press, 1972.

[MR07]   C. Moallemi and B. Van Roy, *Convergence of the min-sum algorithm for convex optimization*, Arxiv preprint arXiv:0705.4253 (2007).

[MR09]   ———, *Convergence of min-sum message passing for quadratic optimization*, IEEE Transactions on Information Theory **55** (2009), no. 5, 2413–2423.

[Rad62]   R. Radner, *Team decision problems*, The Annals of Mathematical Statistics **33** (1962), no. 3, 857–881.

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

[RR01]   P. Rusmevichientong and B. Van Roy, *An analysis of belief propagation on the turbo decoding graph withGaussian densities*, IEEE Transactions on Information Theory **47** (2001), no. 2, 745–765.

[RR03]   ———, *Decentralized decision-making in a large team with local information*, Games and Economic Behavior **43** (2003), no. 2, 266–295.

[San07]   S. Sanghavi, *Equivalence of LP Relaxation and Max-Product for Weighted Matching in General Graphs*, Information Theory Workshop, 2007, 2007, pp. 242–247.

[SSW07]   S. Sanghavi, D. Shah, and A. Willsky, *Message-passing for Maximum Weight Independent Set*, Advances in Neural Information Processing Systems (2007).

[TJ02]   S. Tatikonda and M. Jordan, *Loopy belief propagation and Gibbs measures*, Proceedings of the 2002 Annual Conference on Uncertainty in Artificial Intelligence, vol. 18, 2002, pp. 493–500.

[Tre01]   L. Trevisan, *Non-approximability results for optimization problems on bounded degree instances*, Proceedings of the thirty-third annual ACM symposium on Theory of computing, 2001, pp. 453–461.

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

[WJ08]   M. Wainwright and M. Jordan, *Graphical models, exponential families, and variational inference.*, Foundations and Trends in Machine Learning **1** (2008), 1–305.

[YFW00]   J. Yedidia, W. Freeman, and Y. Weiss, *Understanding Belief Propagation and its generalizations*, Tech. Report TR-2001-22, Mitsubishi Electric Research Laboratories, 2000.