

Correlation Decay in Random Decision Networks ^{*}

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

April 29, 2013

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 rewards are sampled from a distribution. Our aim is to identify sufficient conditions on the network structure and rewards 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 a near-optimal solution 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 Weight Independent Set problem for the same class of graphs admits a PTAS when the weights are independently and identically distributed with the exponential distribution. Namely, randomization of the reward function turns an NP-hard problem into a tractable one.

Keywords: Optimization, NP-hardness, long-range independence

AMS subject classification: 90B15; 05C85; 68R10; 68W20; 68W25

Contents

1 Introduction and literature review

2

^{*}Preliminary version of this paper will appear in Proceedings of ACM-SIAM Symposium on Discrete Algorithms, 2010, Austin, TX

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

[‡]Research supported by the NSF grant CMMI-0726733

[§]Georgia Institute of Technology, Atlanta, GA, 30332, e-mail: dgoldberg9@isye.gatech.edu

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

2	Model description and notations	6
2.1	Examples	6
2.1.1	Independent Set	6
2.1.2	Graph Coloring	6
2.1.3	MAX-2SAT	6
2.1.4	Edwards-Anderson model	7
2.1.5	Maximum a Posteriori (MAP) estimation	7
2.2	Notations	7
3	Main results	9
3.1	Uniform and Gaussian Distributions	9
3.2	Maximum Weight Independent Set problem	9
4	The bonus recursion	11
4.1	Trees	11
4.2	General graphs	12
4.3	Computation tree and the Cavity Expansion algorithm	16
5	Correlation decay and decentralized optimization	18
5.1	Correlation decay implies near-optimal decentralized decisions	19
5.2	Correlation decay and efficient decentralized optimization	20
6	Establishing the correlation decay property through coupling	21
6.1	Notations	21
6.2	Distance-dependent coupling and correlation decay	23
6.2.1	Proof of Theorem 8	23
6.3	Establishing coupling bounds	27
6.3.1	Coupling Lemma	27
6.3.2	Uniform distribution and proof of Theorem 1	28
6.3.3	Gaussian distribution and proof of Theorem 2	29
7	Maximum Weight Independent Set problem	33
7.1	Cavity expansion and the algorithm	33
7.2	Proof of Theorem 3	35
7.2.1	Correlation decay property	35
7.2.2	Concentration argument	38
7.3	Generalization to higher degrees. Proof of Theorem 4.	39
7.4	Hardness result and proof of Theorem 5	42
8	Conclusion	43

1 Introduction and literature review

A decision network can be thought of as a team of agents working in a networked structure (V, E) , where V is a set of agents, and E the set of connections (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

of actions, and the team incurs a total reward $F(x) = \sum \Phi_v(x_v) + \sum_{u,v \in E} \Phi_{u,v}(x_u, x_v)$, where Φ_v and $\Phi_{u,v}$ are node and edge-based reward functions. The goal of the agents is to choose decisions $(x_v, v \in V)$ so that the total reward F is maximized. In this paper we will primarily focus on the case when these reward functions are random. The problem stated above with random reward functions $\Phi_v, \Phi_{u,v}$ subsumes many models in a variety of fields including economic team theory, statistical inference, combinatorial optimization on graphs, and statistical physics. We now discuss some of these models in high level terms. More details are given in the next sections.

Our first example is the class of so-called graphical models, also known as Markov Random Fields - a common model in the area of statistical inference, Bayesian networks, and coding theory (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 in the framework defined above. The randomness of the reward function is induced from the uncertainty of model parameters.

In 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 is 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 (or any boolean constraint satisfaction problem), 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). Of particular interest in statistical physics are models with random interaction potentials. A classical and one of the most studied examples includes the Sherrington-Kirkpatrick model [Tal10], which in our terms corresponds to a complete graph with Gaussian i.i.d. edge based reward functions. Another model widely studied in statistical physics is the so-called Viana-Bray model, which corresponds to random reward functions taking values in $\{-1, 1\}$. We describe these models in more details in the next section.

The combinatorial nature of the decision problem $\max_x F(x)$ implies that the problem of finding $x^* = \operatorname{argmax}_x F(x)$ is NP-hard in general, 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, the focus has been on 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 the BP algorithm, we provide below a brief summary of known theoretical facts about this algorithm.

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 several cycles [PS88],[WJ08]. 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. This was extended to general convex functions in [MR07]. Other cases where BP produces optimal solutions include Maximum Weight Bipartite Matching [San07, BBCZ08, BSS08] (for matchings), Maximum Weight Independent Sets (MWIS) problems where the LP relaxation is tight ([SSW07]), network flow problems [GSW10], and more generally, optimization problems defined on totally unimodular constraint matrices [Che08]. The problem of finding the maximum likelihood estimation based on Gaussian noise can also be solved by running BP on a loopy graph [FW01],[RR01],[JMW06].

The goal of this paper is to introduce and study a new algorithm for the problem of finding $\max F(x)$ and $x^* = \operatorname{argmax} F(x)$. Our algorithm is called the *Cavity Expansion* (CE) algorithm, and it falls into the framework of message-passing type algorithms. We 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 approximation algorithms for solving certain graph counting problems. Specifically, Bandyopadhyay and Gamarnik [BG08], and Weitz [Wei06] introduced a new class of algorithms for these problems which are based on local (in the graph-theoretic sense) computation. Provided that the model exhibits a form of correlation decay, these algorithms provide provable approximation guarantees. The approach was later extended in Gamarnik and Katz [GK07],[GK10], Bayati et al [BGK⁺07], and Jung and Shah [JS07]. The present work develops a similar approach, but for *optimization* problems. The description of the CE algorithm begins by introducing the notion of a *bonus* $B_v(x)$ for each node/decision pair (v, x) . We note that the notion of bonus was heavily used recently in several papers devoted to the local weak convergence theory [Ald01],[AS03],[GNS06]. Also, the nearly identical notion of cavity has been used recently in the statistical physics literature [MP03, RBMM04]. $B_v(x)$ 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 bonus $B_v(x)$ in terms of bonuses of the neighbors of v in suitably modified sub-networks of the underlying network. The algorithm then proceeds by expanding this recursion, in a breadth-first search manner, for some fixed number of steps t , thus constructing an associated computation tree with depth t . At the initialization point the bonus 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 bonus 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 bonus $B_v(x)$ 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 was shown that some optimization problems on locally tree-like graphs with random rewards 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 r , 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 $\Phi = (\Phi_v, \Phi_{u,v})$. 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 assume that the rewards are independent across different nodes and edges. This is consistent with all of the models with random rewards described above. For the simplicity of analysis we further assume that the reward values are independent for different decisions. Although both the algorithm and analysis can be extended to some cases of dependent rewards as well, we leave the general case of dependent rewards for future research. While the Gaussian distribution is motivated by the spin glass models described above, the case of the uniform distribution is treated simply in order to illustrate that our approach does not hinge on particular distributional assumptions (e.g. Gaussian reward functions).

Finally, we also consider exponentially distributed (with parameter 1) weights for the MWIS problem. In this setting, our results have a particularly interesting implication for the theory of average-case analysis of combinatorial optimization problems. It is known that finding the size of a maximum independent set 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. But 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, as shown in Berman and Karpinski [BK99]. In contrast, we show that when $\Delta \leq 3$ and the node weights are independently generated from a parameter 1 exponential distribution, the problem of finding the maximum weight independent set admits a PTAS. Thus, surprisingly, introducing random weights turns 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 bonus recursion, an exact recursion for computing the bonus 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 bonus 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 monotonicity. We present concluding thoughts 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 $x_u \in \chi \triangleq \{0, 1, \dots, T-1\}$. For every $v \in V$, a function $\Phi_v : \chi \rightarrow \mathbb{R}$ is given. Also for every edge $e = (u, v)$ a function $\Phi_e : \chi^2 \rightarrow \mathbb{R} \cup \{-\infty\}$ 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 $\mathbf{x} = (x_1, x_2, \dots, x_{|V|})$ 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 $\mathbf{X} = (X_1, \dots, X_n)$ 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 observed values $\Phi_v(x)$ and $\Phi_e(x, y)$).

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 (MWIS) 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 $\{1, \dots, q\}$ 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 $(v, x) \in V \times \{1, \dots, q\}$ is equipped with a weight $W_{v,x} \geq 0$. The (weight) 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 $\Phi_{v,u}(x, x) = -\infty$, $\Phi_{v,u}(x, y) = 0, \forall x \neq y \in \chi = \{1, \dots, q\}$, $(v, u) \in E$ and $\Phi_v(x) = W_{v,x}, \forall v \in V, x \in \chi$.

2.1.3 MAX-2SAT

Let (Z_1, \dots, Z_n) be a set of boolean variables. Let (C_1, \dots, C_m) be a list of clauses of the form $(Z_i \vee Z_j)$, $(Z_i \vee \bar{Z}_j)$, $(\bar{Z}_i \vee Z_j)$ or $(\bar{Z}_i \vee \bar{Z}_j)$. The MAX-2SAT problem consists of finding an assignment for binary variables Z_i which maximizes the number of satisfied clauses C_j . In terms of a decision network, take $V = \{1, \dots, n\}$, $E = \{(i, j) : Z_i \text{ and } Z_j \text{ appear in a common clause}\}$, and for any k , let $\Phi_k(x, y)$ 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 . Often a random instance of the MAX-2SAT problem is considered where

the edges are generated according to some random law and the clauses are generated by negating each participating variable equiprobably and independently. This corresponds in our setting to a decision network where both the edge set E and reward functions Φ_k are random.

2.1.4 Edwards-Anderson model

The following model is widely studied in the theory of spin glasses [MPV87]. Using our notations, $T = 1$ and $\Phi_v(x) = 0$ for all v, x and $\Phi_{v,u}(x, y) = J_{v,u}(2x - 1)(2y - 1)$, where $J_{v,u}$ are i.i.d. standard normal Gaussian random variables. Typically the alphabet $\{-1, 1\}$ is used for convenience instead of $\{0, 1\}$, leading to $\Phi_{v,u}(x, y) = Jxy$. Furthermore a lot of focus is on the case when (V, E) is a subgraph induced by the lattice \mathbb{Z}^d . A similar model assumes $J_{u,v}$ are symmetric i.i.d. Bernoulli random variables with values $\{-1, 1\}$.

2.1.5 Maximum a Posteriori (MAP) estimation

This example is motivated by statistical inference. Consider a graph (V, E) with $|V| = n$ and $|E| = m$, a set of real numbers $\mathbf{p} = (p_1, \dots, p_n) \in [0, 1]^n$, and a family (f_1, \dots, f_m) of functions such that for each $(i, j) \in E$, $f_{i,j} = f_{i,j}(o, x, y) : \mathbb{R} \times \{0, 1\}^2 \rightarrow \mathbb{R}_+$ 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’ 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

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

and 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 $\Phi_{i,j}(x, y)$ depends on the observation o , which is random itself.

2.2 Notations

For any two nodes u, v in V , let $d(u, v)$ be the length (number of edges) of a 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 $\mathcal{N}_{\mathcal{G}}(u) \triangleq \mathcal{B}(u, 1) \setminus \{u\}$ 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 (degree) of u in \mathcal{G} . Let $\Delta_{\mathcal{G}}$ be the maximum degree of the 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 $\mathbf{v} = (v_1, \dots, v_k)$, and $\mathbf{x} = (x_1, \dots, x_k) \in \chi^k$, let $J_{\mathcal{G}, \mathbf{v}}(\mathbf{x})$ be the optimal value when the actions of nodes v_1, \dots, v_k are fixed to be x_1, \dots, 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 *bonus* 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 bonus 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 bonus 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 Δ and the size T of the action space both bounded by a constant, for all practical purposes we can think of $|V|$ as the size of the instance. When we say that an algorithm is polynomial time, 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 an 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 reward 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 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 $|V|$ and $1/\epsilon$. For our purposes another relevant class of algorithms is EPTAS. This is the class of algorithms which produce ϵ -approximation (either additive or multiplicative) in time $O(|V|^{O(1)}g(\epsilon))$, where $g(\epsilon)$ is some function independent from $|V|$. 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 $|V|$. 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 provides probabilistic guarantees, one may wonder whether FPRAS (Fully Polynomial Randomized Approximation Scheme) would be a more appropriate framework. The typical setting for FPRAS is, however, to have a deterministic problem input; the randomization relates to the algorithm design, not the problem instance. To the contrary, in our setting the instance is random due to random reward functions, though our algorithms, with the exception of our algorithm for the MWIS problem, will be 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 rewards, 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

Our first main result concerns decision networks with uniformly distributed awards.

Theorem 1. *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 $I_1, I_2 > 0$. Let $\beta = \frac{5I_2}{2I_1}$. If $\beta(\Delta - 1)^2 < 1$, then there exists an additive FPTAS with high probability for the problem of finding $J_{\mathcal{G}}$.*

The value I_1 quantifies the ‘bias’ each agent has towards one action or another, while I_2 quantifies the strength of interactions between agents. The intuition behind this result is as follows. The ratio β measures the *relative* strength of interactions between the actions of neighboring agents. When the relative strength is sufficiently small compared to the maximum degree, the actions between the agents at a sufficient distance from each other are asymptotically independent. This can be utilized for designing an efficient nearly optimal decision algorithm. The complementary side of this intuition will be established later in the context of the MWIS problem. Namely, we will show that when β (appropriately redefined) is relatively large compared to the degree Δ , the problem of finding a nearly optimal decision is (in an appropriate sense) intractable.

Now we turn to the case of Gaussian rewards.

Theorem 2. *Suppose for every edge $e = (u, v)$ and any pair of actions $(x, y) \in \{0, 1\}^2$, $\Phi_{u,v}(x, y)$ is a Gaussian random variable with mean 0 and standard deviation σ_e . Suppose for every node $v \in V$, $\Phi_v(1) = 0$ and $\Phi_v(0)$ 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 . 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 rewards, we have obtained a more general result which incorporates the case of correlated edge rewards. It is given as Proposition 6 in Section 6.

The intuition behind Theorem 2 is the same as above. β measures the relative strength of interactions between the agents, which is now measured in terms of the ratio of the variances. We will show that when the relative strength of interactions is sufficiently small, the actions of agents far apart decorrelate, and this can be utilized for designing a fast approximately optimal decision algorithm.

3.2 Maximum Weight Independent Set problem

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 weight 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 with high probability for the problem of finding $J_{\mathcal{G}}$. The algorithm runs in time $O(|V|2^{O(\epsilon^{-2} \log(1/\epsilon))})$.*

As we discussed in the introduction, 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 by 3 [BK99, Tre01a], 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. Yet, it is natural to ask if the above result can be generalized, and in particular whether it is possible to find for each Δ a distribution for which our approach works for graphs with degree bounded by Δ . This is indeed possible as we now demonstrate. 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 an FPTAS with high probability for the problem of finding $J_{\mathcal{G}}$. The algorithm runs in time $O(|V|(\frac{1}{\epsilon})^{\Delta})$.*

Note that for the case of 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 Section 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 weight independent set with i.i.d. exponentially distributed weights does not admit a PTAS. We need to keep in mind that since we are dealing with instances which are random (in terms of weights) and worst-case (in terms of the underlying graph) at the same time, we need to be careful as to the notion of hardness we use.

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 $\rho \leq \hat{w}/\mathbb{E}[W(I^*)] \leq 1/\rho$ 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 a similar result for expectations: there exists an EPTAS 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 a graph is sufficiently large, it is impossible to approximate the quantity $E[W(I^*)]$ arbitrarily closely, unless $P=NP$. Specifically,

Theorem 5. *Let the node weights of a graph be i.i.d. exponentially distributed with parameter 1. There exist universal constants 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 [Tre01a], no attempt is made to compute a similar bound either.

4 The bonus recursion

In this section, we introduce the *bonus recursion*, an exact recursion for computing the bonus functions of each node in a general decision network. We first start by giving the bonus recursion for trees (which is already known as the max-sum 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

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

For every action $x \in \chi$, μ is called the partial bonus function. Recall the definition of the bonus from Subsection 2.2.

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

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

Proof. Suppose $C(u) = \{v_1, \dots, v_d\}$. Observe that the subtrees $\mathcal{G}_{v_i}, 1 \leq i \leq d$ are disconnected (see figure 1) Thus,

$$\begin{aligned} B_u(x) &= \Phi_u(x) + \max_{x_1, \dots, x_d} \left\{ \sum_{j=1}^d \Phi_{u,v_j}(x, x_j) + J_{\mathcal{G}_{v_j, v_j}}(x_j) \right\} \\ &\quad - \Phi_u(0) - \max_{x_1, \dots, x_d} \left\{ \sum_{j=1}^d \Phi_{u,v_j}(0, x_j) + J_{\mathcal{G}_{v_j, v_j}}(x_j) \right\} \\ &= \Phi_u(x) - \Phi_u(0) \\ &\quad + \sum_{j=1}^d \left\{ \max_y (\Phi_{u,v_j}(x, y) + J_{\mathcal{G}_{v_j, v_j}}(y)) - \max_y (\Phi_{u,v_j}(0, y) + J_{\mathcal{G}_{v_j, v_j}}(y)) \right\}. \end{aligned}$$

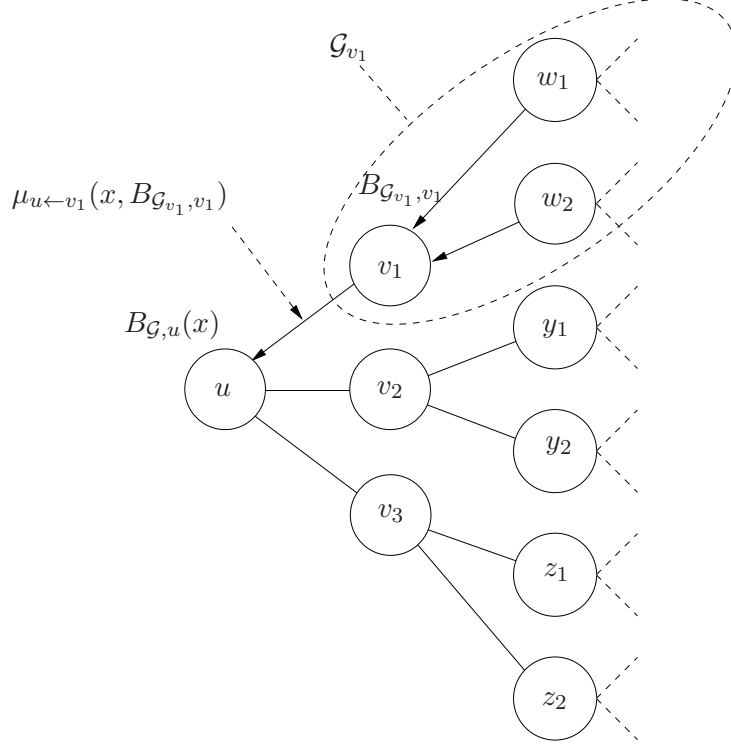


Figure 1: Bonus recursion for trees, equivalent to the BP algorithm.

For every j ,

$$\begin{aligned} & \max_y (\Phi_{u,v_j}(x, y) + J_{\mathcal{G}_{v_j}, v_j}(y)) - \max_y (\Phi_{u,v_j}(0, y) + J_{\mathcal{G}_{v_j}, v_j}(y)) = \\ & \max_y (\Phi_{u,v_j}(x, y) + J_{\mathcal{G}_{v_j}, v_j}(y) - J_{\mathcal{G}_{v_j}, v_j}(0)) - \max_y (\Phi_{u,v_j}(0, y) + J_{\mathcal{G}_{v_j}, v_j}(y) - J_{\mathcal{G}_{v_j}, v_j}(0)). \end{aligned}$$

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

Iteration (2) constitutes what is known as (max-sum) belief propagation. Proposition 1 is the restatement of the well-known fact that BP finds an optimal solution on a tree [PS88]. BP 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 $\mathcal{N}(u) = \{v_1, \dots, v_d\}$. For every $j = 1, \dots, d$ 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, $E' = E \setminus \{(u, v_1), \dots, (u, v_d)\}$. Also $\Phi'_e = \Phi_e$ for all e in E' ,

and the potential functions Φ'_v are defined as follows. For any $v \in V \setminus \{u, v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_d\}$, $\Phi'_v = \Phi_v$, and

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

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

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

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

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

The first step of the proof consists in considering the following telescoping sum (see figure 2):

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

and the k^{th} difference

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

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

$$\begin{aligned}\max_{\mathbf{z}} \left\{ \sum_{j=1}^d \Phi_{u,v_j}(x_{j,k}, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\} &= \\ \max_{z_k} \left(\Phi_{u,v_k}(x, z_k) + \max_{\mathbf{z}_{-k}} \left\{ \sum_{j \leq k-1} \Phi_{u,v_j}(x, z_j) + \sum_{j \geq k+1} \Phi_{u,v_j}(0, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\} \right).\end{aligned}\quad (7)$$

Similarly,

$$\begin{aligned}\max_{\mathbf{z}} \left\{ \sum_{j=1}^d \Phi_{u,v_j}(x_{j,k-1}, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\} &= \\ \max_{z_k} \left(\Phi_{u,v_k}(0, z_k) + \max_{\mathbf{z}_{-k}} \left\{ \sum_{j \leq k-1} \Phi_{u,v_j}(x, z_j) + \sum_{j \geq k+1} \Phi_{u,v_j}(0, z_j) + J_{\mathcal{G} \setminus \{u\}, \mathbf{v}}(\mathbf{z}) \right\} \right).\end{aligned}\quad (8)$$

$$\begin{aligned}
B_u \left(\begin{array}{c} \text{graph} \end{array} \right) &= \Phi_u(x) - \Phi_u(0) + J \left(\begin{array}{c} \text{graph with } u \text{ black} \end{array} \right) - J \left(\begin{array}{c} \text{graph with } u \text{ gray} \end{array} \right) \\
&= \Phi_u(x) - \Phi_u(0) + J \left(\begin{array}{c} \text{graph with } \Phi_{u,v_1} \text{ and } u \text{ black} \end{array} \right) - J \left(\begin{array}{c} \text{graph with } \Phi_{u,v_1} \text{ and } u \text{ gray} \end{array} \right) \\
&= \Phi_u(x) - \Phi_u(0) + J \left(\begin{array}{c} \text{graph with } \Phi_{u,v_1} \text{ and } u \text{ black} \end{array} \right) - J \left(\begin{array}{c} \text{graph with } \Phi_{u,v_1} \text{ and } u \text{ gray} \end{array} \right) \\
&\quad + J \left(\begin{array}{c} \text{graph with } \Phi_{u,v_1} \text{ and } u \text{ gray} \end{array} \right) - J \left(\begin{array}{c} \text{graph with } \Phi_{u,v_1} \text{ and } u \text{ black} \end{array} \right) \\
&\quad + J \left(\begin{array}{c} \text{graph with } \Phi_{u,v_1} \text{ and } u \text{ gray} \end{array} \right) - J \left(\begin{array}{c} \text{graph with } \Phi_{u,v_1} \text{ and } u \text{ black} \end{array} \right)
\end{aligned}$$

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.

$$\begin{aligned}
& J \left(\begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right) - J \left(\begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} \right) = J \left(\begin{array}{c} \text{Diagram 5} \\ \text{Diagram 6} \end{array} \right) - J \left(\begin{array}{c} \text{Diagram 7} \\ \text{Diagram 8} \end{array} \right) \\
& = J \left(\begin{array}{c} \text{Diagram 9} \\ \text{Diagram 10} \end{array} \right) - J \left(\begin{array}{c} \text{Diagram 11} \\ \text{Diagram 12} \end{array} \right) = \mu_{u \leftarrow v_2}(x, B_{\mathcal{G}(u,2,x),v_2});
\end{aligned}$$

with

$$\begin{aligned}
\tilde{\Phi}_{v_1}(z) &= \Phi_{v_1}(z) + \Phi_{u,v_1}(0, z), \\
\tilde{\Phi}_{v_3}(z) &= \Phi_{v_3}(z) + \Phi_{u,v_3}(x, z).
\end{aligned}$$

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.

For each z_k , we have (see figure 3):

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

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

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

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

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

□

4.3 Computation tree and the Cavity Expansion algorithm

Given a decision network \mathcal{G} , for every node $u \in V$ with $\mathcal{N}_u = \{v_1, \dots, v_d\}$, and every $r \in \mathbb{Z}_+$, 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. $\text{CE}[\mathcal{G}, u, 0, x] = 0$ for all x .
2. For every $r = 1, 2, \dots$, every $u \in V$ and every $x \in \chi$,

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

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

Cavity Expansion: $\text{CE}[\mathcal{G}, u, r, x]$

INPUT: A network \mathcal{G} , a node u in \mathcal{G} , an action x and a computation depth $r \geq 0$

BEGIN

If $r = 0$ return 0

else do

Find neighbors $\mathcal{N}(u) = \{v_1, v_2, \dots, v_d\}$ of u in \mathcal{G} .

If $\mathcal{N}(u) = \emptyset$, return $\Phi_u(x) - \Phi_u(0)$.

Else

For each $j = 1, \dots, d$, construct the network $\mathcal{G}(u, j, x)$.

For each $j = 1, \dots, d$, and $y \in \chi$, compute $CE[\mathcal{G}(u, j, x), v_j, r - 1, y]$
 For each $j = 1, \dots, d$, compute $\mu_{u \leftarrow v_j}(x, CE[\mathcal{G}(u, j, x), v_j, r - 1])$
 Return $\Phi_u(x) - \Phi_u(0) + \sum_{1 \leq j \leq d} \mu_{u \leftarrow v_j}(x, CE[\mathcal{G}(u, j, x), v_j, r - 1])$ as $CE[\mathcal{G}, u, r, x]$.

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 $(\mathcal{G}_i, v_i, x_i)_{1 \leq i \leq m}$ 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 vector $\mathcal{C} = (\mathcal{C}_i)_{1 \leq i \leq m}$, which is an arbitrary set of values, we call a *boundary condition*. The boundary condition initializes our recursion based CE algorithm. We denote by $CE[\mathcal{G}, u, r, x, \mathcal{C}]$ the output of the CE 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 bonus $B_{\mathcal{G}, u}(x)$ 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 $\mathcal{N}(u) = \{v_1, \dots, v_d\}$, suppose for every $j = 1, \dots, d$ 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 CE algorithm can be visualized as a computation on a tree, due to its recursive nature. This has some similarity with the 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 $\mathcal{G}(u, j, x)$ (similar to [Wei06, JS07, BGK⁺07, GK10, GK07]). 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 bonus recursion, the reward function of an edge reward is statistically independent from the subtree below that edge.

Proposition 3. *Given u, x and $\mathcal{N}(v) = \{v_1, \dots, v_d\}$, for every $r, j = 1, \dots, d$ 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). \square

Our last proposition analyzes the complexity of running the CE algorithm.

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

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 for any subnetwork \mathcal{G}' of \mathcal{G} , $\text{CE}[\mathcal{G}', u, r, x]$ can be computed in time bounded by $O(r(\Delta T)^r)$. The values for $r = 1$ can be computed in time bounded by $O(\Delta T)$, by the observation above. For $r > 1$, the computations of $\text{CE}[\mathcal{G}', u, r, x]$ requires a fixed reward of $O(\Delta T)$, as well as (ΔT) calls to CE with depth $(r - 1)$. The total reward is therefore bounded by $O(\Delta T + (\Delta T)(r - 1)(\Delta T)^{r-1})$, which is $O(r(\Delta T)^r)$. \square

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 algorithms. When a network exhibits the correlation decay property, the bonus functions of faraway nodes are weakly related, implying a weak dependence between their optimal decisions as well. One can then take advantage of this in order to build a good decentralized decision scheme.

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

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

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

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

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

This observation is the key for designing our approximation algorithms. The following assumptions will be frequently used in later sections.

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 be able to verify this assumption in many special cases of interest. We will also frequently assume the reward functions have finite second moments. Thus we let

$$K_\Phi \triangleq \left(\sum_{x,y \in \chi} \mathbb{E} \Phi_e^2(x, y) \right)^{1/2}, \quad (10)$$

and primarily focus on the case $K_\Phi < \infty$.

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 $x_v^r = \operatorname{argmax}_x CE[\mathcal{G}, v, r, x]$, and $\mathbf{x}^r = (x_v^r)$. The main relation between the correlation decay property, CE 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,*

$$\mathbb{P}(x_u^r \neq x_u) \leq 2T^2 \sqrt{2g\rho(r)}, \quad \forall u \in V, r \geq 1. \quad (11)$$

Proof. For simplicity, let $B_u^r(x)$ denote $CE[\mathcal{G}, u, r, x]$. We will first prove that for every $\epsilon > 0$

$$\mathbb{P}(x_u^r \neq x_u) \leq T^2 \left(g\epsilon + \frac{2\rho(r)}{\epsilon} \right). \quad (12)$$

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 z and z' , if $|z| > \epsilon$ and $|z - z'| \leq \epsilon$, then $zz' > 0$. Applying this to $z = B_u(x) - B_u(y)$ and $z' = B_u^r(x) - B_u^r(y)$, 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 $x \neq y$,

$$\begin{aligned} \mathbb{P}\left((B_u(x) - B_u(y))(B_u^r(x) - B_u^r(y)) \leq 0\right) &\leq \mathbb{P}(|B_u(x) - B_u(y)| \leq \epsilon) + \mathbb{P}(|B_u(x) - B_u^r(x)| \geq \epsilon/2) \\ &\quad + \mathbb{P}(|B_u(y) - B_u^r(y)| \geq \epsilon/2). \end{aligned} \quad (13)$$

Now, $\mathbb{P}(|B_u(x) - B_u(y)| \leq \epsilon)$ is at most $2g\epsilon$ by Assumption 1. Using Markov's inequality, we find that the second summand in (13) 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

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

□

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 Assumption 1 holds. Then*

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

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 the logarithm of the accepted error.

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 $\mathbf{x} = (x_u)$ denote the optimal solution for the network \mathcal{G} , and let $\mathbf{x}^r = (x_u^r)$ be the decisions resulting from the CE algorithm with depth r . Let $K_1 = 10K_\Phi T(|V| + |E|)$, and $K_2 = K_1 (gK_c)^{1/4}$, where K_c is defined under the assumption of exponential correlation decay property, when it applies.

Theorem 7. *Suppose a decision network \mathcal{G} satisfies the correlation decay property with rate $\rho(r)$ and Assumption 1 holds. Then, for all $r > 0$,*

$$\mathbb{E}[F(\mathbf{x}) - F(\mathbf{x}^r)] \leq K_1 (g\rho(r))^{1/4}. \quad (14)$$

The theorem above is relevant, for example in the corollary below, only in the case $K_\Phi < \infty$, even though this assumption is not necessary for the proof.

Corollary 2. *Suppose \mathcal{G} exhibits the exponential correlation decay property with rate α_c , Assumption 1 holds, and $K_\Phi < \infty$. For any $\epsilon > 0$, if*

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

then

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

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

$$\mathbb{P}((x_u^r, x_v^r) \neq (x_u, x_v)) \leq 4T^2 \sqrt{2g\rho(r)}.$$

Also,

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

For any $u, v \in V$,

$$\begin{aligned} \mathbb{E}|\Phi_{u,v}(x_u, x_v) - \Phi_{u,v}(x_u^r, x_v^r)| &\leq \mathbb{E}\left[1_{(x_u^r, x_v^r) \neq (x_u, x_v)} \left(|\Phi_{u,v}(x_u, x_v)| + |\Phi_{u,v}(x_u^r, x_v^r)|\right)\right] \\ &\leq 2K_\Phi \mathbb{P}((x_u^r, x_v^r) \neq (x_u, x_v))^{1/2} \\ &\leq 4K_\Phi T (2g\rho(r))^{1/4}, \end{aligned}$$

where the second inequality follows from Schwarz's inequality. Similarly, for any u , we have

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

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 the bound (14) follows.

Corollary 2 is then proved using Markov's inequality. In particular, applying the definition of the exponential correlation decay property to Relation (14), we obtain

$$\mathbb{P}(F(\mathbf{x}) - F(\mathbf{x}^r) \geq \epsilon) \leq \mathbb{E}[F(\mathbf{x}) - F(\mathbf{x}^r)]/\epsilon \leq K_2 \alpha_c^{r/4}/\epsilon.$$

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

6 Establishing the correlation decay property through coupling

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 the correlation decay property holds. Results in this section are for the case $|\chi| = 2$. We note that 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, \dots, 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 $j = 1, \dots, d$, 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. Let $(v_{j1}, \dots, v_{jn_j})$ be the neighbors of v_j in \mathcal{G}_j . For every $1 \leq k \leq n_j$ let $B_{jk}(r-2) = \text{CE}[\mathcal{G}_j(v_j, k, 1), v_{jk}, r-2, 1, \mathcal{C}]$ and $B'_{jk}(r-2) = \text{CE}[\mathcal{G}_j(v_j, k, 1), v_{jk}, r-2, 1, \mathcal{C}']$. 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. Since 1 is the only action different from the reference action 0, for any adjacent nodes u, v , it suffices to consider $\mu_{u \leftarrow v}$ as a function of a scalar B as opposed to a vector. Thus we define $\mu_{u \leftarrow v}(z) \triangleq \mu_{u \leftarrow v}(1, z)$, for any value z . From Equation (1), we find the following alternative expression for $\mu_{u \leftarrow v}(z)$:

$$\begin{aligned} \mu_{u \leftarrow v}(z) = & \Phi_{u,v}(1, 1) - \Phi_{u,v}(0, 1) + \max(\Phi_{u,v}(1, 0) - \Phi_{u,v}(1, 1), z) \\ & - \max(\Phi_{u,v}(0, 0) - \Phi_{u,v}(0, 1), z). \end{aligned} \quad (15)$$

5. For any $\mathbf{z} = (z_1, \dots, z_d)$, let $\mu_u(\mathbf{z}) = \sum_j \mu_{u \leftarrow v_j}(z_j)$.
6. 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 , $\mathbb{E}|Y_e| \leq K_\Phi$ defined by (10). Equation (9) can be rewritten as

$$B(r) = \mu_u(\mathbf{B}(r-1)) + \Phi_u(1) - \Phi_u(0), \quad (16)$$

$$B'(r) = \mu_u(\mathbf{B}'(r-1)) + \Phi_u(1) - \Phi_u(0). \quad (17)$$

Similarly, we have

$$B_j(r-1) = \mu_{v_j}(\mathbf{B}_j(r-2)) + \Phi_{v_j}(1) - \Phi_{v_j}(0), \quad (18)$$

$$B'_j(r-1) = \mu_{v_j}(\mathbf{B}'_j(r-2)) + \Phi_{v_j}(1) - \Phi_{v_j}(0). \quad (19)$$

Finally, Equation (15) can be rewritten as

$$\mu_{u \leftarrow v}(z) = \Phi_{u,v}^3 + \max(\Phi_{u,v}^1, z) - \max(\Phi_{u,v}^2, z). \quad (20)$$

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 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

Our next goal is to identify conditions on the distributions of the potential and interaction functions which lead to the correlation decay property. We achieve this by adopting a coupling approach.

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' :

$$\mathbb{P}\left(\mu_{u\leftarrow v}(x + \Phi_v(1) - \Phi_v(0)) = \mu_{u\leftarrow v}(x' + \Phi_v(1) - \Phi_v(0))\right) \geq (1 - a) - b|x - x'|. \quad (21)$$

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 a sufficient condition for the network to exhibit (a, b) coupling is that for all x, x' ,

$$\mathbb{P}\left(\mu_{u\leftarrow v}(x) = \mu_{u\leftarrow v}(x')\right) \geq (1 - a) - b|x - x'|. \quad (22)$$

It will turn out that the more general form (21) is more applicable to our setting. The form of coupling above 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

$$a(\Delta - 1) + \sqrt{bK_\Phi}(\Delta - 1)^{3/2} < 1, \quad (23)$$

then the exponential correlation decay property holds with $K_c = \Delta^2 K_\Phi$ and

$$\alpha_c = a(\Delta - 1) + \sqrt{bK_\Phi}(\Delta - 1)^{3/2}.$$

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

$$a(\Delta - 1) + bK_Y(\Delta - 1)^2 < 1, \quad (24)$$

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

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'

$$|\mu_{u\leftarrow v}(x) - \mu_{u\leftarrow v}(x')| \leq |x - x'|. \quad (25)$$

Proof. From (15) we obtain

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

Using the relation that for any real numbers z, z', z'' one has that $\max(z, z') - \max(z, z'') \leq \max(0, z' - z'')$, we obtain:

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

The other inequality is proved similarly. □

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

$$|\mu_{u \leftarrow v}(x) - \mu_{u \leftarrow v}(x')| \leq |Y_{u,v}|. \quad (26)$$

Proof. Using (15), we have

$$\begin{aligned} \mu_{u \leftarrow v}(x) - (\Phi_{u,v}(1, 1) - \Phi_{u,v}(0, 1)) &= \max(\Phi_{u,v}(1, 0) - \Phi_{u,v}(1, 1), x) \\ &\quad - \max(\Phi_{u,v}(0, 0) - \Phi_{u,v}(0, 1), x). \end{aligned}$$

By using the relation that for any real numbers z, z', z'' one has that $\max(z, z') - \max(z, z'') \leq \max(0, z' - z'')$ 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 the two inequalities, we obtain

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

The other inequality follows from a similar proof. \square

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

$$\mathbb{E}|B(r) - B'(r)| \leq a \sum_{1 \leq j \leq d} \mathbb{E}|B_j(r-1) - B'_j(r-1)| + b \sum_{1 \leq j \leq d} \mathbb{E}[|B_j(r-1) - B'_j(r-1)|^2]. \quad (27)$$

Proof. Using (9), we obtain:

$$\begin{aligned} \mathbb{E}|B(r) - B'(r)| &= \mathbb{E} \left[\left| \Phi_u(1) - \Phi_u(0) + \sum_j \mu_{u \leftarrow v_j}(B_j(r-1)) - (\Phi_u(1) - \Phi_u(0)) - \sum_j \mu_{u \leftarrow v_j}(B'_j(r-1)) \right| \right] \\ &\leq \sum_j \mathbb{E} |\mu_{u \leftarrow v_j}(B_j(r-1)) - \mu_{u \leftarrow v_j}(B'_j(r-1))| \\ &= \sum_j \mathbb{E} \left[\mathbb{E} [|\mu_{u \leftarrow v_j}(B_j(r-1)) - \mu_{u \leftarrow v_j}(B'_j(r-1))| \mid \mu_{v_j}(\mathbf{B}_j(r-2)), \mu_{v_j}(\mathbf{B}'_j(r-2))] \right]. \end{aligned}$$

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 that from Equations (18) and (19), $|B_j(r-1) - B'_j(r-1)| = |\mu_{v_j}(\mathbf{B}_j(r-2)) - \mu_{v_j}(\mathbf{B}'_j(r-2))|$; 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,

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

is at most

$$|B_j(r-1) - B'_j(r-1)| \times \mathbb{P}(\mu_{u \leftarrow v_j}(B_j(r-1)) \neq \mu_{u \leftarrow v_j}(B'_j(r-1)) \mid \mu_{v_j}(\mathbf{B}_j(r-2)), \mu_{v_j}(\mathbf{B}'_j(r-2))). \quad (28)$$

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 (21)

$$\begin{aligned} & \mathbb{P}(\mu_{u \leftarrow v_j}(B_j(r-1)) \neq \mu_{u \leftarrow v_j}(B'_j(r-1)) \mid \mu_{v_j}(\mathbf{B}_j(r-2)), \mu_{v_j}(\mathbf{B}'_j(r-2))) \\ & \leq a + b|B_j(r-1) - B'_j(r-1)|. \end{aligned}$$

The result then follows. \square

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

$$\mathbb{E}|\text{CE}(\mathcal{G}, u, r, \mathcal{C}) - \text{CE}(\mathcal{G}, u, r, \mathcal{C}')| \leq K\alpha_{\mathcal{C}}^r. \quad (29)$$

We first establish the bound inductively for the case $d \leq \Delta - 1$. Let e_r denote the supremum of the left-hand side of (29), 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 . From the definition it follows that each subnetwork exhibits the (a, b) coupling when the original network does.

Condition (23). Under (23), we claim that

$$e_r \leq a(\Delta - 1)e_{r-1} + b(\Delta - 1)^3 K_{\Phi} e_{r-2}. \quad (30)$$

Indeed, by applying (18) and (19), we have

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

Thus by Jensen's inequality,

$$\begin{aligned} |B_j(r-1) - B'_j(r-1)|^2 & \leq \left(\sum_{1 \leq k \leq d_j} |\mu_{v_j \leftarrow v_{j_k}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{j_k}}(B'_{jk}(r-2))| \right)^2 \\ & \leq d_j \sum_{1 \leq k \leq d_j} |\mu_{v_j \leftarrow v_{j_k}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{j_k}}(B'_{jk}(r-2))|^2. \end{aligned}$$

By Lemmas 1 and 2 we have

$$|\mu_{v_j \leftarrow v_{j_k}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{j_k}}(B'_{jk}(r-2))| \leq |B_{jk}(r-2) - B'_{jk}(r-2)|,$$

and

$$|\mu_{v_j \leftarrow v_{j_k}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{j_k}}(B'_{jk}(r-2))| \leq |Y_{jk}|.$$

Also, $d_j \leq \Delta - 1$. Therefore,

$$|B_j(r-1) - B'_j(r-1)|^2 \leq (\Delta - 1) \sum_{1 \leq k \leq d_j} |B_{jk}(r-2) - B'_{jk}(r-2)| \times |Y_{jk}|. \quad (31)$$

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

$$\begin{aligned} \mathbb{E}|B_j(r-1) - B'_j(r-1)|^2 &\leq (\Delta-1) \sum_{1 \leq k \leq d_j} \mathbb{E}|B_{jk}(r-2) - B'_{jk}(r-2)| \times \mathbb{E}|Y_{jk}| & (32) \\ &\leq (\Delta-1) K_\Phi \sum_{1 \leq k \leq d_j} \mathbb{E}|B_{jk}(r-2) - B'_{jk}(r-2)| \\ &\leq (\Delta-1)^2 K_\Phi e_{r-2}, \end{aligned}$$

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} , $1 \leq k \leq d_j$ 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

$$\begin{aligned} \mathbb{E}|B(r) - B'(r)| &\leq a \sum_{1 \leq j \leq d} \mathbb{E}|B_j(r-1) - B'_j(r-1)| + b \sum_{1 \leq j \leq d} \mathbb{E}[|B_j(r-1) - B'_j(r-1)|^2] \\ &\leq a(\Delta-1)e_{r-1} + b(\Delta-1)^3 K_\Phi e_{r-2}. \end{aligned}$$

This implies (30).

From (30) we obtain that $e_r \leq K\alpha_c^r$ for $K = \Delta K_\Phi$ and α_c given as the largest in absolute value root of the quadratic equation $\alpha_c^2 = a(\Delta-1)\alpha_c + b(\Delta-1)^3 K_\Phi$. We find this root to be

$$\begin{aligned} \alpha_c &= \frac{1}{2}a(\Delta-1) + \frac{1}{2}\sqrt{a^2(\Delta-1)^2 + 4b(\Delta-1)^3 K_\Phi} \\ &\leq a(\Delta-1) + \sqrt{b(\Delta-1)^3 K_\Phi} \\ &< 1, \end{aligned}$$

where the last inequality follows from assumption (23). 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 (16) and (17) we have

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

Again applying Lemma 1, we find that the right-hand side is at most

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

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 (24). Recall from Lemma 3 that for all r , we have:

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

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 , $|\mu_{v_j \leftarrow v_{jk}}(B_{jk}) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk})| \leq |Y_{jk}| \leq K_Y$, where the first inequality follows from Lemma 2, and the second follows 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 before shows that this property implies correlation decay.

6.3 Establishing coupling bounds

6.3.1 Coupling Lemma

Theorem 8 details sufficient conditions 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 prove coupling bounds in our setting. The following simple observation will 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:

$$E_{u \leftarrow v}^+(x, x') = \{\min(x, x') + \Phi_v(1) - \Phi_v(0) \geq \max(\Phi_{u \leftarrow v}^1, \Phi_{u \leftarrow v}^2)\};$$

$$E_{u \leftarrow v}^-(x, x') = \{\max(x, x') + \Phi_v(1) - \Phi_v(0) \leq \min(\Phi_{u \leftarrow v}^1, \Phi_{u \leftarrow v}^2)\};$$

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

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

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

Proof. From representation (20), we have $\mu_{u \leftarrow v}(x) = \Phi_{u, v}^3 + \max(\Phi_{u, v}^1, z) - \max(\Phi_{u, v}^2, z)$. Let x, x' be any two reals. If both x and x' are greater than both $\Phi_{u, v}^1$ and $\Phi_{u, v}^2$, then $\mu_{u \leftarrow v}(x) = \Phi_{u, v}^3 = \mu_{u \leftarrow v}(x')$. If both x and x' are smaller than both $\Phi_{u, v}^1$ and $\Phi_{u, v}^2$, then $\mu_{u \leftarrow v}(x) = \Phi_{u, v}^3 + \Phi_{u, v}^1 - \Phi_{u, v}^2 = \mu_{u \leftarrow v}(x')$. 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 occurring,

$$\mathbb{P}(\mu_{u \leftarrow v}(x + \Phi_v(1) - \Phi_v(0)) \neq \mu_{u \leftarrow v}(x' + \Phi_v(1) - \Phi_v(0))),$$

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 $x \geq x'$,

$$(E(x, x'))^c = \{\min(\Phi^1, \Phi^2) + \Phi_v(0) - \Phi_v(1) < x < \max(\Phi^1, \Phi^2) + \Phi_v(0) - \Phi_v(1) + x - x'\}. \quad (33)$$

6.3.2 Uniform distribution and proof of Theorem 1

In order to prove Theorem 1, we compute the coupling parameters a, b for this distribution and apply the second part 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 $(u, v) \in \mathcal{G}$, $\Phi_{u,v}^1$ and $\Phi_{u,v}^2$ are i.i.d. random variables with a triangular distribution (difference of two independent uniformly distributed random variables) with support $[-2I_2, 2I_2]$. Because $\Phi_{u,v}^1$ and $\Phi_{u,v}^2$ are i.i.d., by symmetry we obtain

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

We have $\mathbb{P}(x' - a_2 < \Phi_v(0) - \Phi_v(1) < x - a_1)$ is at most $\frac{a_2 - a_1 + x - x'}{2I_1}$, since $\Phi_v(0) - \Phi_v(1)$ is uniformly distributed on $[-I_1, I_1]$. We obtain

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

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 $a_1 \geq 0$,

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

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:

$$\begin{aligned} & \int_{-2I_2}^{2I_2} d\mathbb{P}_{\Phi^1}(a_1) \int_{a_1}^{2I_2} d\mathbb{P}_{\Phi^2}(a_2)(a_2 - a_1) \\ &= \frac{1}{4I_2^2} \left(\int_{-2I_2}^0 ((a_1 + 2I_2)(-a_1 + \frac{1}{24I_2^2}(a_1 + 2I_2)^3) d(a_1) + \int_0^{2I_2} \frac{1}{24I_2^2} (2I_2 - a_1)^4 d(a_1) \right) \\ &= \frac{1}{4I_2^2} \left(\frac{24}{15} I_2^3 + \frac{4}{15} I_2^3 \right) = \frac{7}{15} I_2. \end{aligned}$$

We obtain

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

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

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 $\beta(\Delta - 1)^2 < 1$ plus $\Delta - 1 \leq (\Delta - 1)^2$ implies $\frac{I_2}{2I_1}(\Delta - 1) + \frac{4I_2}{2I_1}(\Delta - 1)^2 < 1$. This is exactly condition (24) with a, b as given by Lemma 5 and $K_Y = 4I_2$. It follows that \mathcal{G} exhibits the exponential correlation decay property, and since Assumption 1 holds for the uniform distribution, all of the conditions of Corollary 2 are satisfied, and there exists an additive FPTAS for computing $J_{\mathcal{G}}$.

6.3.3 Gaussian distribution and proof of Theorem 2

In this section, we compute the coupling parameters when the reward functions have a Gaussian distribution. 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 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

$$\begin{aligned} \sigma_1^2 &= S_{10,10} - 2S_{10,11} + S_{11,11} + \sigma_p^2; & \sigma_2^2 &= S_{00,00} - 2S_{00,01} + S_{01,01} + \sigma_p^2; \\ \rho &= (\sigma_1\sigma_2)^{-1}(S_{00,10} - S_{00,11} - S_{01,10} + S_{01,11} + \sigma_p^2); & C &= \frac{\sigma_2^2 - \sigma_1^2}{\sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\rho^2\sigma_1^2\sigma_2^2}}; \\ \sigma_X^2 &= \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2; & \sigma_Y^2 &= \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2. \end{aligned}$$

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

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

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 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

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

$$b = \sqrt{\frac{1}{2\pi}} \frac{1}{\sqrt{\sigma_e^2 + \sigma_p^2}},$$

implying

$$bK_\Phi \leq \sqrt{\frac{2}{\pi}} \beta \leq \beta.$$

□

Combining Corollary 3 and the first part of Theorem 8 yields Theorem 2. The remainder of this section is devoted to proving Proposition 6.

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

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

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 $X = \bar{\Phi}^1 + \bar{\Phi}^2$, $Y = \bar{\Phi}^2 - \bar{\Phi}^1$. 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. Let also $\bar{X} \triangleq X - \mathbb{E}[X]$ and $\bar{Y} \triangleq Y - \mathbb{E}[Y]$ be 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 (33), we have

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

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.

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

□

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

$$\begin{aligned} \mathbb{P}((E)^c(x, x')) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-C^2}} \int_{S(t)} \exp\left(-\frac{1}{2(1-C^2)}\left(\frac{(x-\mu_1-\mu_2+2b)^2}{\sigma_x^2} + \frac{(y-\mu_2+\mu_1)^2}{\sigma_y^2}\right.\right. \\ &\quad \left.\left.- 2C\frac{(x-\mu_1-\mu_2+2b)(y+\mu_2-\mu_1)}{\sigma_x\sigma_y}\right)\right) dx dy \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-C^2}} \int_y \exp\left(-\frac{1}{2(1-C^2)}\frac{(y-\mu_2+\mu_1)^2}{\sigma_y^2}\right) g(y) dy, \end{aligned} \quad (34)$$

where

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

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

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

Now,

$$\tilde{x}_b - C\tilde{y} = \frac{x - \mu_1 - \mu_2 + 2b - C\sigma_x\sigma_y^{-1}(y - \mu_2 + \mu_1)}{\sigma_x}.$$

Recall Anderson's inequality [Dud99]. Namely, 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

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

Applying this bound in Equation (34), we obtain

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

Finally, note that by the triangle inequality, for any α_c we have

$$S(t) \subset S_{\alpha_c}(t) \triangleq \{(x, y) : |y - \alpha_c| \geq |x| - t - |\alpha_c|\}.$$

We obtain

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

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

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

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 ,

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

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

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

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

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

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

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

implying

$$\mathbb{P}((E)^c(x, x')) \leq \left(\frac{1}{\pi} \arctan\left(\frac{\sigma_y}{\sigma_x\sqrt{1-C^2}}\right) + \sqrt{\frac{2}{\pi(1-C^2)} \frac{|\mu_2 - \mu_1|}{\sigma_x}}\right) + \sqrt{\frac{2}{\pi(1-C^2)} \frac{t}{\sigma_x}}, \quad (35)$$

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

7 Maximum Weight Independent Set problem

7.1 Cavity expansion and the algorithm

In this section, we show how the correlation decay framework applies to the Maximum Weight Independent Set (MWIS) problem and prove Theorems 3,4 and 5. In achieving these goals, we face some additional challenges when compared to the models considered in the previous sections. First, the bounded rewards assumption required for the results of Section 5 does not hold for constrained optimization problems, as the underlying problem allows for infinite (negative) rewards. Second, the coupling technique of Section 6 is not readily applicable for MWIS. We therefore develop a different approach.

Similar to the previous sections, we use a three step approach. 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 node weighted graph $\mathcal{G} = (V, E, W)$, where (V, E) is a graph whose nodes are equipped with arbitrary non-negative weights $W_u, u \in V$. No probabilistic assumption on W_u is adopted yet. Note that for the MWIS problem, we have $J_{\mathcal{G}} = W(I^*)$, and for any (u_1, \dots, u_d) , $J_{\mathcal{G},(u_1, \dots, u_d)}(\mathbf{0}) = J_{\mathcal{G} \setminus \{u_1, \dots, u_d\}}$, where $\mathcal{G} \setminus \{u_1, \dots, u_d\}$ is the subgraph induced by nodes $V \setminus \{u_1, \dots, u_d\}$. Throughout this section we let $n = |V|$ denote the number of nodes. Consider a given node $u \in V$ and let $N(u) = \{u_1, \dots, u_d\}$. In light of the fact that again we are dealing with the model with only two decisions per node, we let $\mathcal{G}(u, l)$ stand for $\mathcal{G}(u, l, 1)$ for $1 \leq l \leq d$. From Theorem 6, we have

$$B_{\mathcal{G},u} = J_{\mathcal{G},u}(1) - J_{\mathcal{G},u}(0) = W_u + \sum_{1 \leq l \leq d} \mu_{u \leftarrow u_l}(1, B_{\mathcal{G}(u,l),u_l}). \quad (36)$$

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 $\mu_{i \leftarrow u_l}$, we have

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

Thus,

$$B_{\mathcal{G},u} = W_u - \sum_{l=1}^d \max(B_{\mathcal{G}(u,l),u_l}, 0).$$

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

- For $v \in \{u_1, \dots, u_{l-1}\}$, $\Phi'_v(0) = \phi_v(0) + \Phi_{u,v}(1, 0) = 0$, and $\Phi'_v(1) = W_v + \Phi_{u,v}(1, 1) = W_v - \infty = -\infty$. Since the weight of v in the modified graph $\mathcal{G}(u, l)$ is $-\infty$, it is equivalent to removing this node from the graph.
- For $v \in \{u_{l+1}, \dots, u_d\}$, $\Phi'_v(0) = \Phi_v(0) + \Phi_{u,v}(0, 0) = 0$, and $\Phi'_v(1) = W_v + \Phi_{u,v}(0, 1) = W_v$.

Thus $\mathcal{G}(u, l)$ is obtained by removing the nodes $\{u, u_1, \dots, u_{l-1}\}$, while keeping the weights of nodes $\{u_{l+1}, \dots, u_d\}$ intact. Equivalently, we simply have $\mathcal{G}(u, l) = \mathcal{G} \setminus \{u, u_1, \dots, u_{l-1}\}$. Therefore,

we obtain

$$B_{\mathcal{G},u} = W_u - \sum_{l=1}^d \max(B_{\mathcal{G}\setminus\{u,u_1,\dots,u_{l-1}\}}, 0).$$

We further modify this recursion by the following change of variables: for any graph \mathcal{G} and node u , let $C_{\mathcal{G}}(u) = \max(B_{\mathcal{G},u}, 0)$. Note that $C_{\mathcal{G}}(u) = \max(J_{\mathcal{G},u}(1), J_{\mathcal{G},u}(0)) - J_{\mathcal{G},u}(0) = J_{\mathcal{G}} - J_{\mathcal{G}\setminus\{u\}}$. The variables C will be called *cavities*. It turns out that for the MWIS problem, working with cavities C is more convenient than with bonuses B . We obtain the cavity recursion for MWIS.

Proposition 7. *For any $u \in V$, let $N(u) = \{u_1, \dots, u_d\}$. Then*

$$C_{\mathcal{G}}(u) = \max\left(0, W_u - \sum_{1 \leq l \leq d} C_{\mathcal{G}\setminus\{u,u_1,\dots,u_{l-1}\}}(u_l)\right), \quad (37)$$

where $\sum_{1 \leq l \leq d} = 0$ when $N(u) = \emptyset$. If $W_u - \sum_{1 \leq l \leq d} C_{\mathcal{G}\setminus\{u,u_1,\dots,u_{l-1}\}}(u_l) > 0$, namely $C_{\mathcal{G}}(u) > 0$, then every largest weight independent set must contain u . Similarly if $W_u - \sum_{1 \leq l \leq d} C_{\mathcal{G}\setminus\{u,u_1,\dots,u_{l-1}\}}(u_l) < 0$, implying $C_{\mathcal{G}}(u) = 0$, then every largest weight independent set does not contain u .

Remark : The proposition leaves out the ‘‘tie’’ case $W_u - \sum_{1 \leq l \leq d} C_{\mathcal{G}\setminus\{u,u_1,\dots,u_{l-1}\}}(u_l) = 0$. 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 $C_{\mathcal{G}}(u) > 0$ ($C_{\mathcal{G}}(u) = 0$) determines whether u must (must not) belong to the maximum weight independent set.

Using the special form of the cavity recursion (37), 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 u , let $C_{\mathcal{H}}^-(u, r) = \max(0, \text{CE}[\mathcal{H}, u, r])$ with boundary condition $\text{CE}[\mathcal{H}, u, 0] = 0$, and let $C_{\mathcal{H}}^+(u, r)$ be the same quantity for the boundary condition $\text{CE}[\mathcal{H}, u, 0] = W_u$. Here we have allowed for random boundary condition C^+ , as the weights $W_u, u \in V$ are random. The reason for this particular choice of the boundary condition will become clear below. The choice of random vs. deterministic boundary condition should not bother the reader since we will not rely on the results of Section 4.3 in later sections. Instead, we will derive the necessary technical results such as the correlation decay property from scratch for our MWIS context.

C^- and C^+ can be alternatively defined by the following recursions:

$$C_{\mathcal{H}}^-(u, r) = \begin{cases} 0, & r = 0; \\ \max\left(0, W_u - \sum_{1 \leq l \leq d} C_{\mathcal{H}\setminus\{u,u_1,\dots,u_{l-1}\}}^-(u_l, r-1)\right), & r \geq 1. \end{cases} \quad (38)$$

$$C_{\mathcal{H}}^+(u, r) = \begin{cases} W_u, & r = 0; \\ \max\left(0, W_u - \sum_{1 \leq l \leq d} C_{\mathcal{H}\setminus\{u,u_1,\dots,u_{l-1}\}}^+(u_l, r-1)\right), & r \geq 1. \end{cases} \quad (39)$$

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

Lemma 7. *For every even r ,*

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

and for every odd r ,

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

Proof. The proof is by induction on r . The assertion holds by definition of C^-, C^+ for $r = 0$. The induction follows from (37), from definitions of C^-, C^+ and from the fact that the function $x \rightarrow \max(0, W - x)$ is non-increasing. \square

We now describe our algorithm for producing a nearly optimal 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 $\mathcal{G}(\epsilon) = (V_\epsilon, E_\epsilon)$. Also, for any fixed subgraph \mathcal{H} of \mathcal{G} , we denote the corresponding (random) subgraph of $\mathcal{G}(\epsilon)$ by $\mathcal{H}(\epsilon)$. In the second stage we compute $C_{\mathcal{G}(\epsilon)}^-(u, r)$ for every node i for the graph $\mathcal{G}(\epsilon)$ for some target even number of steps r . We set

$$\mathcal{I}(r, \epsilon) = \{i : C_{\mathcal{G}(\epsilon)}^-(u, r) > 0\}.$$

Let I_ϵ^* be the largest weight independent set of $\mathcal{G}(\epsilon)$. It will be straightforward to show that the weight of I_ϵ^* is nearly the weight of $W(I^*)$ as ϵ becomes small. Our goal is to show that $\mathcal{I}(r, \epsilon)$ is an independent set with weight close to $W(I_\epsilon^*)$ when ϵ is small and r is large. Finding $\mathcal{I}(r, \epsilon)$ and computing its weight constitutes our algorithm for solving the MWIS problem.

Lemma 8. $\mathcal{I}(r, \epsilon)$ is an independent set. In particular, $\mathcal{I}(r, \epsilon) \subset I_\epsilon^*$.

Proof. By Lemma 7, if $C_{\mathcal{G}(\epsilon)}^-(u, r) > 0$ then $C_{\mathcal{G}(\epsilon)} > 0$, and therefore $\mathcal{I}(r, \epsilon) \subset I_\epsilon^*$. Thus our algorithm produces an independent set in $\mathcal{G}(\epsilon)$ and therefore in \mathcal{G} . \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 then be straightforward to show that $W(I_\epsilon^*)$ is close to $W(I^*)$, which will finally give us the desired result, Theorem 3.

For an arbitrary induced subgraph \mathcal{H} of \mathcal{G} , and any node u in \mathcal{H} , introduce $M_{\mathcal{H}}(u) = \mathbb{E}[\exp(-C_{\mathcal{H}}(u))]$, $M_{\mathcal{H}}^-(u, r) = \mathbb{E}[\exp(-C_{\mathcal{H}}^-(u, r))]$, $M_{\mathcal{H}}^+(u, r) = \mathbb{E}[\exp(-C_{\mathcal{H}}^+(u, r))]$. We note that sometimes, \mathcal{H} will itself be a *random* subgraph, e.g. $\mathcal{G}(\epsilon)$. In this case, the above should be interpreted as conditional expectations, i.e. the relevant expectations are taken over only the randomness in the weights themselves, not the randomness associated with the random deletion of nodes.

The key correlation decay property is established in the following result.

Proposition 8. For every node u in $\mathcal{G}(\epsilon)$ and every r ,

$$\mathbb{P}(C_{\mathcal{G}(\epsilon)}(u) = 0, C_{\mathcal{G}(\epsilon)}^+(u, 2r) > 0) \leq 3(1 - \epsilon^2/16)^{2r}, \quad (40)$$

and

$$\mathbb{P}(C_{\mathcal{G}(\epsilon)}(u) > 0, C_{\mathcal{G}(\epsilon)}^-(u, 2r) = 0) \leq 3(1 - \epsilon^2/16)^{2r}. \quad (41)$$

Proof. Consider a subgraph \mathcal{H} of \mathcal{G} , and a node $u \in \mathcal{H}$ with neighbors $\mathcal{N}_{\mathcal{H}}(u) = \{u_1, \dots, u_d\}$. Examining the recursion (37) we observe that all the randomness in terms $C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}(u_l)$ comes from the subgraph $\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}$, and thus W_u is independent from the vector $(C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}(u_l), 1 \leq l \leq d)$. A similar assertion applies when we replace $C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}(u_l)$ with $C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^-(u_l, r)$ and $C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^+(u_l, r)$ for every r . Using the memoryless property of the exponential distribution, and denoting by W a rate-1 exponential random variable, we obtain:

$$\begin{aligned}
\mathbb{E}[\exp(-C_{\mathcal{H}}(u)) \mid \sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}(u_l) = x] \\
&= \mathbb{P}(W_u \leq x) \mathbb{E}[\exp(0)] + \mathbb{E}[\exp(-(W_u - x)) \mid W_u > x] \mathbb{P}(W_u > x) \\
&= (1 - \mathbb{P}(W_u > x)) + \mathbb{E}[\exp(-W)] \mathbb{P}(W_u > x) \\
&= (1 - \mathbb{P}(W_u > x)) + (1/2) \mathbb{P}(W_u > x) \\
&= 1 - (1/2) \mathbb{P}(W_u > x) \\
&= 1 - (1/2) \exp(-x).
\end{aligned} \tag{42}$$

It follows that

$$\mathbb{E}[\exp(-C_{\mathcal{H}}(u))] = 1 - (1/2) \mathbb{E} \exp \left(- \sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}(u_l) \right).$$

Similarly, we obtain

$$\begin{aligned}
\mathbb{E}[\exp(-C_{\mathcal{H}}^-(u, r))] &= 1 - (1/2) \mathbb{E} \exp \left(- \sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^-(u_l, r - 1) \right); \\
\mathbb{E}[\exp(-C_{\mathcal{H}}^+(u, r))] &= 1 - (1/2) \mathbb{E} \exp \left(- \sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^+(u_l, r - 1) \right).
\end{aligned}$$

As a notational convenience, let us define $u_0 = u$. We now prove that

$$\left| M_{\mathcal{H}}^+(u, r) - M_{\mathcal{H}}^-(u, r) \right| \leq \sum_{i=0}^{|\mathcal{N}_{\mathcal{H}}(u)|-1} \frac{1}{2} \left| M_{\mathcal{H} \cup_{j=0}^i u_j}^+(u_{i+1}, r - 1) - M_{\mathcal{H} \cup_{j=0}^i u_j}^-(u_{i+1}, r - 1) \right|. \tag{43}$$

We treat the cases $d = 0, 1, 2, 3$ separately. For $d = 0$, we have trivially $M_{\mathcal{H}}(u) = M_{\mathcal{H}}^-(u) = M_{\mathcal{H}}^+(u) = 1/2$, and (43) follows. Suppose $d = 1$ and $\mathcal{N}_{\mathcal{H}}(u) = \{u_1\}$. Then,

$$\begin{aligned}
M_{\mathcal{H}}^-(u, r) - M_{\mathcal{H}}^+(u, r) &= (1/2) \left(\mathbb{E}[\exp(-C_{\mathcal{H} \setminus \{u\}}^+(u_1, r - 1))] - \mathbb{E}[\exp(-C_{\mathcal{H} \setminus \{u\}}^-(u_1, r - 1))] \right) \\
&= (1/2) \left(M_{\mathcal{H} \setminus \{u\}}^+(u_1, r - 1) - M_{\mathcal{H} \setminus \{u\}}^-(u_1, r - 1) \right),
\end{aligned} \tag{44}$$

and (43) follows. Suppose $d = 2$, and $\mathcal{N}(u) = \{u_1, u_2\}$. Then

$$\begin{aligned}
M_{\mathcal{H}}^-(u, r) - M_{\mathcal{H}}^+(u, r) \\
&= (1/2) \mathbb{E}[\exp(-C_{\mathcal{H} \setminus \{u\}}^+(u_1, r - 1) - C_{\mathcal{H} \setminus \{u, u_1\}}^+(u_2, r - 1))] \\
&\quad - (1/2) \mathbb{E}[\exp(-C_{\mathcal{H} \setminus \{u\}}^-(u_1, r - 1) - C_{\mathcal{H} \setminus \{u, u_1\}}^-(u_2, r - 1))] \\
&= (1/2) \mathbb{E}[\exp(-C_{\mathcal{H} \setminus \{u\}}^+(u_1, r - 1)) (\exp(-C_{\mathcal{H} \setminus \{u, u_1\}}^+(u_2, r - 1)) - \exp(-C_{\mathcal{H} \setminus \{u, u_1\}}^-(u_2, r - 1)))] \\
&\quad + (1/2) \mathbb{E}[\exp(-C_{\mathcal{H} \setminus \{u, u_1\}}^-(u_2, r - 1)) (\exp(-C_{\mathcal{H} \setminus \{u\}}^+(u_1, r - 1)) - \exp(-C_{\mathcal{H} \setminus \{u\}}^-(u_1, r - 1))].
\end{aligned}$$

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

$$\begin{aligned}
0 &\leq M_{\mathcal{H}}^+(u, r) - M_{\mathcal{H}}^-(u, r) \leq (1/2)\mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{u, u_1\}}^-(u_2, r-1)) - \exp(-C_{\mathcal{H}\setminus\{u, u_1\}}^+(u_2, r-1))] \\
&\quad + (1/2)\mathbb{E}[\exp(-C_{\mathcal{H}\setminus\{u\}}^-(u_1, r-1)) - \exp(-C_{\mathcal{H}\setminus\{u\}}^+(u_1, r-1))], \\
&= (1/2)(M_{\mathcal{H}\setminus\{u, u_1\}}^-(u_2, r-1) - M_{\mathcal{H}\setminus\{u, u_1\}}^+(u_2, r-1)) \\
&\quad + (1/2)(M_{\mathcal{H}\setminus\{u\}}^-(u_1, r-1) - M_{\mathcal{H}\setminus\{u\}}^+(u_1, r-1)), \tag{45}
\end{aligned}$$

and for even r

$$\begin{aligned}
0 &\leq M_{\mathcal{H}}^-(u, r) - M_{\mathcal{H}}^+(u, r) \leq (1/2)(M_{\mathcal{H}\setminus\{u, u_1\}}^+(u_2, r-1) - M_{\mathcal{H}\setminus\{u, u_1\}}^-(u_2, r-1)) \\
&\quad + (1/2)(M_{\mathcal{H}\setminus\{u\}}^+(u_1, r-1) - M_{\mathcal{H}\setminus\{u\}}^-(u_1, r-1)), \tag{46}
\end{aligned}$$

and again (43) follows. Performing a nearly identical argument for the case $d = 3$, and combining with the above, completes the proof of (43). It then follows from a straightforward conditioning argument that

$$\mathbb{E}\left[\left|M_{\mathcal{H}(\epsilon)}^+(u, r) - M_{\mathcal{H}(\epsilon)}^-(u, r)\right|\middle|u \in \mathcal{G}(\epsilon)\right]$$

is at most

$$\sum_{i=0}^{|\mathcal{N}_{\mathcal{H}}(u)|-1} \frac{1}{2}\left(1 - \frac{\epsilon^2}{16}\right)\mathbb{E}\left[\left|M_{\mathcal{H}(\epsilon)\setminus\bigcup_{j=0}^i u_j}^+(u_{i+1}, r-1) - M_{\mathcal{H}(\epsilon)\setminus\bigcup_{j=0}^i u_j}^-(u_{i+1}, r-1)\right|\middle|u_{i+1} \in \mathcal{G}(\epsilon)\right], \tag{47}$$

with all expectations taken w.r.t. the random deletion of nodes. Note that the r.h.s. of (47) is the sum of $|\mathcal{N}_{\mathcal{H}}(u)|$ terms, with each term of the form

$$\frac{1}{2}\left(1 - \frac{\epsilon^2}{16}\right)\mathbb{E}\left[\left|M_{\mathcal{H}'(\epsilon)}^+(u', r-1) - M_{\mathcal{H}'(\epsilon)}^-(u', r-1)\right|\middle|u' \in \mathcal{G}(\epsilon)\right],$$

where u' is a node of degree *at most* 2 in \mathcal{H}' , and the event $\{u' \in \mathcal{G}(\epsilon)\}$ is independent of the random deletion of all nodes belonging to $\mathcal{H}'(\epsilon)$, other than u' itself. Thus we may apply the same argument to each summand on the r.h.s. of (47), and repeat inductively r times (i.e. to the summands appearing in the bounds for those summands, etc.) to conclude that

$$\mathbb{E}\left[\left|M_{\mathcal{H}(\epsilon)}^+(u, r) - M_{\mathcal{H}(\epsilon)}^-(u, r)\right|\middle|u \in \mathcal{G}(\epsilon)\right] \leq 3/2\left(1 - \frac{\epsilon^2}{16}\right)^r.$$

Summarizing, for every r

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

where the relevant expectation is taken w.r.t. both the random weights and the random deletion of nodes. Recalling (42) we have

$$\mathbb{E}[\exp(-C_{\mathcal{G}(\epsilon)}(u))] = 1 - (1/2)\mathbb{P}(W > \sum_{1 \leq l \leq d} C_{\mathcal{G}(\epsilon)\setminus\{u, u_1, \dots, u_{l-1}\}}(u_l)) = 1 - (1/2)\mathbb{P}(C_{\mathcal{G}(\epsilon)}(u) > 0).$$

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

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

Again applying Lemma 7, we obtain

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

and

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

This completes the proof of the proposition. \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 $W(I_\epsilon^* \setminus \mathcal{I}(r, \epsilon))$ 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, $\mathbb{E}[W(\Delta V_\epsilon)] = \epsilon^2/16n$.

To analyze $W(I_\epsilon^* \setminus \mathcal{I}(r, \epsilon))$, observe that by the second part of Proposition 8, for every node $u, \mathbb{P}(u \in I_\epsilon^* \setminus \mathcal{I}(r, \epsilon)) \leq 3(1 - \epsilon^2/16)^r \triangleq \delta_1$. Thus $\mathbb{E}|I_\epsilon^* \setminus \mathcal{I}(r, \epsilon)| \leq \delta_1 n$. In order to obtain a bound on $W(I_\epsilon^* \setminus \mathcal{I}(r, \epsilon))$ we derive a crude bound on the largest weight of a subset with cardinality $\delta_1 n$. Fix a constant C and consider the set V_C of all nodes in $\mathcal{G}(\epsilon)$ with weight greater than C . We have $\mathbb{E}[W(V_C)] \leq (C + \mathbb{E}[W - C | W > C]) \exp(-C)n = (C + 1) \exp(-C)n$. The remaining nodes have weight at most C . Therefore,

$$\begin{aligned} \mathbb{E}[W(I_\epsilon^* \setminus \mathcal{I}(r, \epsilon))] &\leq \mathbb{E}[W((I_\epsilon^* \setminus \mathcal{I}(r, \epsilon)) \cap V_C^c \cup V_C)] \leq C \mathbb{E}[|I_\epsilon^* \setminus \mathcal{I}(r, \epsilon)|] + \mathbb{E}[W(V_C)] \\ &\leq C \delta_1 n + (C + 1) \exp(-C)n. \end{aligned}$$

We conclude that

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

Now we obtain a lower bound on $W(I^*)$. Consider the standard greedy algorithm for generating an independent set: take an arbitrary node, remove its 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.

$$\begin{aligned}
\mathbb{P}\left(\frac{W(\mathcal{I}(r, \epsilon))}{W(I^*)} \leq 1 - \epsilon\right) &\leq \mathbb{P}\left(\frac{W(\mathcal{I}(r, \epsilon))}{W(I^*)} \leq 1 - \epsilon, W(I^*) \geq n/8\right) + \mathbb{P}(W(I^*) < n/8) \\
&\leq \mathbb{P}\left(\frac{|W(I^*) - W(\mathcal{I}(r, \epsilon))|}{W(I^*)} \geq \epsilon, W(I^*) \geq n/8\right) + 64/n \\
&\leq \mathbb{P}\left(\frac{|W(I^*) - W(\mathcal{I}(r, \epsilon))|}{n/8} \geq \epsilon\right) + 64/n \\
&\leq \frac{\epsilon^2/16 + 3C(1 - \epsilon^2/16)^r + (C + 1)\exp(-C)}{\epsilon/8} + 64/n,
\end{aligned}$$

where we have used Markov's inequality in the last step and $\delta_1 = 3(1 - \epsilon^2/16)^r$. Thus it suffices to arrange C so that the first ratio is at most $2\epsilon/3$ and assuming, without 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 desired result. This completes the proof of Theorem 3. \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 an exponentially distributed random variable with rate α_j .

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

Lemma 9. *Fix any subgraph \mathcal{H} , node $u \in \mathcal{H}$ with $N_{\mathcal{H}}(u) = \{u_1, \dots, u_d\}$. Then*

$$\begin{aligned}
\mathbb{E}[\exp(-\alpha_j C_{\mathcal{H}}(u))] &= 1 - \frac{1}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \mathbb{E}[\exp(-\sum_{1 \leq l \leq d} \alpha_k C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}(u_l))]; \\
\mathbb{E}[\exp(-\alpha_j C_{\mathcal{H}}^+(u, r))] &= 1 - \frac{1}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \mathbb{E}[\exp(-\sum_{1 \leq l \leq d} \alpha_k C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^+(u_l, r - 1))]; \\
\mathbb{E}[\exp(-\alpha_j C_{\mathcal{H}}^-(u, r))] &= 1 - \frac{1}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \mathbb{E}[\exp(-\sum_{1 \leq l \leq d} \alpha_k C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^-(u_l, r - 1))].
\end{aligned}$$

Proof. Let $\alpha(u)$ be the random rate associated with node u . Namely, $\mathbb{P}(\alpha(u) = \alpha_j) = 1/\Delta$. We

condition on the event $\sum_{1 \leq l \leq d} C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}(u_l) = x$. As $C_{\mathcal{H}}(u) = \max(0, W_u - x)$, we obtain:

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

Thus,

$$\mathbb{E}[-\alpha_j C_{\mathcal{H}}(u)] = 1 - \frac{1}{\Delta} \sum_k \frac{\alpha_j}{\alpha_j + \alpha_k} \mathbb{E}[\exp(-\sum_{1 \leq l \leq d} \alpha_k C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}(u_l))].$$

The other equalities follow identically. \square

By taking differences, we obtain

$$\begin{aligned}
M_{\mathcal{H}}^{-,j}(u, r) - M_{\mathcal{H}}^{+,j}(u, r) &= \frac{1}{\Delta} \sum_k \frac{\alpha_j}{\alpha_j + \alpha_k} \left(\mathbb{E}[\prod_{1 \leq l \leq d} \exp(-\alpha_k C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^+(u_l, r-1))] - \right. \\
&\quad \left. - \mathbb{E}[\prod_{1 \leq l \leq d} \exp(-\alpha_k C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^-(u_l, r-1))] \right).
\end{aligned}$$

We now use the identity

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

which further implies that

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

when $\max_l |x_l|, |y_l| < 1$. By applying this inequality with $x_l = \exp(-\alpha_k C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^+(u_l, r-1))$ and $y_l = \exp(-\alpha_k C_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^-(u_l, r-1))$, we obtain

$$\begin{aligned}
|M_{\mathcal{H}}^{-,j}(u, r) - M_{\mathcal{H}}^{+,j}(u, r)| &\leq \frac{1}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \sum_{1 \leq l \leq d} |M_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^{-,k}(u_l, r-1) - M_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^{+,k}(u_l, r-1)|.
\end{aligned}$$

This implies

$$|M_{\mathcal{H}}^{-,j}(u, r) - M_{\mathcal{H}}^{+,j}(u, r)| \tag{49}$$

$$\leq \frac{d}{\Delta} \sum_{1 \leq k \leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \max_{1 \leq l \leq d} |M_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^{-,k}(u_l, r-1) - M_{\mathcal{H} \setminus \{u, u_1, \dots, u_{l-1}\}}^{+,k}(u_l, r-1)|. \tag{50}$$

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

$$e_{r,j} = \sup_{\mathcal{H} \subset \mathcal{G}, u \in \mathcal{H}} |M_{\mathcal{H}}^{-,j}(u, r) - M_{\mathcal{H}}^{+,j}(u, r)|. \tag{51}$$

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

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

For any $t \geq 0$, let \mathbf{e}_r denote the vector $(e_{r,1}, \dots, e_{r,m})$. Let \mathbf{M} denote the matrix such that for all (j, k) , $M_{j,k} = \frac{d}{\Delta} \frac{\alpha_j}{\alpha_j + \alpha_k}$. We finally obtain

$$\mathbf{e}_r \leq M \mathbf{e}_{r-1}.$$

Therefore, if M^r converges to zero exponentially fast in each coordinate, then \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 $\frac{d}{\Delta} \leq 1$. 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 the 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 ,

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

Indeed, it is easy to verify that when $\rho > 25$ (i.e. $\epsilon < 1/25$), one has $(1/2 + 2\frac{\sqrt{\epsilon}}{1-\sqrt{\epsilon}}) < 1$, and thus the above completes the proof. Now, 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

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

This completes the proof of the lemma the theorem. \square

7.4 Hardness result and proof of Theorem 5

We first comment on why our proofs of Theorems 3 and 4 establish that there exists an EPTAS for computing the deterministic quantity $\mathbb{E}[W(I^*)]$, the expected weight of the MWIS in the graph \mathcal{G} considered. It follows from Inequality (48), with $r = O(\log(1/\epsilon)/\epsilon^2)$ and $C = O(\log(1/\epsilon))$, that it suffices to establish the existence of an EPTAS for computing $\mathbb{E}[W_u I(C_{\mathcal{G}(\epsilon)}^-(u, 2r) > 0)]$ for any fixed constant r and any given node u in the graph \mathcal{G} . But since $\Delta_{\mathcal{G}} \leq 3$, it follows from the definition and locality of the CE algorithm, and the fact that there are only a finite (depending on r) number of graphs of maximum degree 3 and diameter at most $2r$, that the computation of $\mathbb{E}[W_u I(C_{\mathcal{G}(\epsilon)}^-(u, 2r) > 0)]$ can be reduced to the computation of an integral which depends only on r (and not $|V|$). The existence of an EPTAS for computing $\mathbb{E}[W(I^*)]$ then follows from the fact that such an integral can be approximated to any desired precision in time which itself depends only on r .

We now show a partial converse to the above, by completing the proof of Theorem 5. Given a graph \mathcal{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 a parameter 1 exponential distribution. We make use of the following result due to Trevisan [Tre01b].

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:*

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

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

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

Observe that

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

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

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

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

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

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

8 Conclusion

We considered an optimization model which encompasses many models from a variety of literatures including graphical models, combinatorial optimization, economics, and statistical physics.

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 reward functions are random, this goal is often achievable. Specifically, we have constructed a general purpose algorithm Cavity Expansion, which relies on local information only, and is thus 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 connections between the fields of complexity of algorithms for combinatorial optimization problems and statistical physics, specifically the cavity method and the notion 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 distributions, even though the maximum cardinality version of the same problem is known to be in-approximable, unless $P=NP$.

It would be interesting to see what weight distributions 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 assumptions 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 practicality of our algorithm is yet another question for future research. In a somewhat related context of solving counting problem, the power of our cavity expansion algorithm was demonstrated by estimating the exponent of monomer-dimer configurations on a lattice graph [GK09], where the proposed algorithm was used to improve state of the art estimates by several orders of magnitude. The practicality of this algorithm in the present context of optimization problems is yet to be explored.

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 procedure representing communication between the agents, one can find (in polynomial time) a 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 useful in approaching these problems and identifying new settings where the solution can be found in polynomial time, in spite of the curse of dimensionality typically exhibited by these models.

Acknowledgements

The authors thank Dirk Oliver Theis for pointing out an error in an earlier version of this manuscript.

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), 452–479.
- [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.
- [BK99] P. Berman and M. Karpinski, *On some tighter inapproximability results*, Lecture notes in computer science ed., Springer Berlin / Heidelberg, 1999.
- [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), 1241–1251.
- [Che08] M. Chertkov, *Exactness of belief propagation for some graphical models with loops*, Journal of Statistical Mechanics: Theory and Experiment (2008), P10016–P10028.
- [Dud99] R.M. Dudley, *Uniform central limit theorems*, Cambridge university press, 1999.
- [FW01] W. Freeman and Y. Weiss, *Correctness of belief propagation in gaussian graphical models of arbitrary topology*, Neural Computation **13** (2001), 2173–2200.
- [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 **19** (2009), 1–25.

- [GK07] D. Gamarnik and D. Katz, *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.
- [GK09] ———, *Sequential cavity method for computing free energy and surface pressure*, Journal of Statistical Physics **137** (2009), 205–232.
- [GK10] ———, *A deterministic approximation algorithm for computing a permanent of a 0,1 matrix*, Journal of Computer and System Sciences **76** (2010), 879883.
- [GNS06] D. Gamarnik, T. Nowicki, and G. Swirszcz, *Maximum weight independent sets and matchings in sparse random graphs. Exact results using the local weak convergence method*, Random Structures and Algorithms **28** (2006), 76–106.
- [GSW10] D. Gamarnik, D. Shah, and Y. Wei, *Belief propagation for min-cost network flow: Convergence & correctness*, Proceedings of 21-st ACM-SIAM Symposium on Discrete Algorithms (SODA), 2010.
- [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.
- [JMW06] J. Johnson, D. Malioutov, and A. Willsky, *Walk-Sum Interpretation and Analysis of Gaussian Belief Propagation*, Advances in Neural Information Processing Systems **18** (2006), 579–586.
- [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.
- [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.
- [MPV87] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin-glass theory and beyond*, vol 9 of *Lecture Notes in Physics*, World Scientific, Singapore, 1987.

- [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] C. C. Moallemi and B. Van Roy, *Convergence of the min-sum message passing algorithm for quadratic optimization*, IEEE Transactions on Information Theory **55** (2009), 2413 – 2423.
- [PS88] J. Pearl and G. Shafer, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, JSTOR, 1988.
- [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 with Gaussian 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, pp. 242–247.
- [SSW07] S. Sanghavi, D. Shah, and A. Willsky., *Message-passing for max-weight independent set*, Advances in Neural Information Processing Systems, MIT Press, 2007.
- [Tal10] M. Talagrand, *Mean field models for spin glasses: Volume I: Basic examples*, Springer, 2010.
- [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.
- [Tre01a] L. Trevisan, *Non-approximability results for optimization problems on bounded degree instances*, ACM Symposium on Theory of Computing, 2001.
- [Tre01b] ———, *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.