Online Companion to "Correlation Decay in Random Decision Networks"

A Examples of Decision Networks

A.1 Independent set

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

A.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 (weighted) coloring problem is the problem of finding a proper coloring ϕ with maximum total weight $\sum_v W_{v,\phi(v)}$. In terms of decision network framework, we have $\Phi_{v,u}(x,x) = -\infty, \Phi_{v,u}(x,y) = 0, \forall x \neq y \in \chi = \{1, \dots, q\}, (v,u) \in E$ and $\Phi_v(x) = W_{v,x}, \forall v \in V, x \in \chi$.

A.3 MAP estimation

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

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

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

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

where

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

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

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

B Proof of proposition 5

Proposition 8. *[Proposition 5 in main text] For every $x \neq y$, $\tilde{B}_v(x) - \tilde{B}_v(y)$ is a continuous random variable with density bounded above by $\frac{1}{\sqrt{4\pi}\delta}$. Moreover, for any random vector $\mathbf{z}$, we have:*

$$\mathbb{E}|F(\mathbf{z}) - \tilde{F}(\mathbf{z})] \leq T\sqrt{\frac{2}{\pi}}|V|\delta. \tag{23}$$

Proof. We have:

$$\tilde{B}_u(x) - \tilde{B}_u(y) = B_u(x) - B_u(y) + \delta Z_{u,x} - \delta Z_{u,y}.$$

Let $D = B_u(x) - B_u(y)$ and $\tilde{D} = \tilde{B}_u(x) - \tilde{B}_v(y)$. Since $Z_{u,x} - Z_{u,y}$ is a zero mean Gaussian random variable with variance 2, then for every $t \in \mathbb{R}$ and $h > 0$, by conditioning on $Z_{u,x} - Z_{u,y}$, we obtain:

$$\begin{aligned}\mathbb{P}(t \leq \tilde{D} < t+h) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{4\pi}\delta} e^{-\frac{u^2}{4\delta^2}} \mathbb{P}(t-u \leq D < t+h-u)du \\ &\leq \frac{1}{\sqrt{4\pi}\delta} \int_{-\infty}^{+\infty} \mathbb{P}(t-u \leq D < t+h-u)du \leq \frac{h}{\sqrt{4\pi}\delta},\end{aligned}$$

where the last inequality follows from the fact that, for any random variable X with $\mathbb{E}|X| < +\infty$,

$$\int_{-\infty}^{\infty} \mathbb{P}(x \leq X < x+h)dx \leq h$$

Taking the limit for $h \to 0$, we conclude that D has a density and which is bounded by $\frac{1}{\sqrt{4\pi}\delta}$. Finally, $F(\mathbf{x}) - F(\tilde{\mathbf{x}}) = F(\mathbf{x}) - \tilde{F}(\mathbf{x}) + \tilde{F}(\mathbf{x}) - \tilde{F}(\tilde{\mathbf{x}}) + \tilde{F}(\tilde{\mathbf{x}}) - F(\tilde{\mathbf{x}})$. By optimality of $\tilde{\mathbf{x}}$ for $\tilde{F}$, we have $\tilde{F}(\mathbf{x}) - \tilde{F}(\tilde{\mathbf{x}}) \leq 0$, so $\mathbb{E}[F(\mathbf{x}) - F(\tilde{\mathbf{x}})] \leq \mathbb{E}|F(\mathbf{x}) - \tilde{F}(\mathbf{x})| + \mathbb{E}|\tilde{F}(\tilde{\mathbf{x}}) - F(\tilde{\mathbf{x}})|$.

We have:

$$|F(\mathbf{x}) - \tilde{F}(\mathbf{x})| \leq \delta \sum_v |Z_{v,x}| \leq \delta \sum_v \sum_y |Z_{v,y}|$$

which implies

$$\mathbb{E}|F(\mathbf{x}) - \tilde{F}(\mathbf{x})| \leq 2T\sqrt{\frac{2}{\pi}}\,|V|\,\delta$$

Similarly we have

$$\mathbb{E}|F(\tilde{\mathbf{x}}) - \tilde{F}(\tilde{\mathbf{x}})| \leq \delta \sum_{v} E|Z_{v,\tilde{x}}| \leq 2T\sqrt{\frac{2}{\pi}}\,|V|\,\delta$$

and so the suboptimality gap is bounded by $4T\sqrt{\frac{2}{\pi}}\,|V|\,\delta$ □

C Proof of Theorem 6

Theorem 8. *[Theorem 6 in main text] Suppose a decision network $\mathcal{G}$ satisfies correlation decay property with rate $\rho(r)$. Then, there exists some constants K_1, K_2 such that for all $r > 0$*

$$\mathbb{E}[F(\mathbf{x}) - F(\mathbf{x}^{r,\delta})] \leq K_1\sqrt{\frac{\rho(r)}{\delta}} + K_2\delta. \tag{24}$$

Let $\delta = \delta(r) = \rho^{1/5}(r)$ and $K_3 = K_1 + K_2$. Then,

$$\mathbb{E}[F(\mathbf{x}) - F(\mathbf{x}^{r,\delta(r)})] \leq K_3\delta(r). \tag{25}$$

Finally suppose $\mathcal{G}$ exhibits exponential correlation decay property with rate α_c. For any $\epsilon > 0$, if $r \geq 3\Big(|\log\epsilon| + \log|K_3| + 1/3|\log K_c|\Big)|\log(\alpha_c)|^{-1}$, then $\mathbb{E}[F(\mathbf{x})] - \mathbb{E}[F(\mathbf{x}^{r,\delta(r)})] \leq \epsilon$

Proof. Let $\tilde{\mathbf{x}}$ be an optimal solution for the network $\tilde{\mathcal{G}}$. We first write a telescoping sum

$$F(\mathbf{x}) - F(\mathbf{x}^{r,\delta}) = \Big(F(\mathbf{x}) - \tilde{F}(\mathbf{x})\Big) + \Big(\tilde{F}(\mathbf{x}) - \tilde{F}(\tilde{\mathbf{x}})\Big) + \Big(\tilde{F}(\tilde{\mathbf{x}}) - \tilde{F}(\mathbf{x}^{r,\delta})\Big) + \Big(\tilde{F}(\mathbf{x}^{r,\delta}) - F(\mathbf{x}^{r,\delta})\Big)$$

Since $\tilde{x}$ is optimal for $\tilde{F}$, $\Big(\tilde{F}(\mathbf{x}) - \tilde{F}(\tilde{\mathbf{x}})\Big) \leq 0$, and thus

$$\mathbb{E}\big[F(\mathbf{x}) - F(\mathbf{x}^{r,\delta})\big] \leq \mathbb{E}\big|F(\mathbf{x}) - \tilde{F}(\mathbf{x})\big| + \mathbb{E}\big|\tilde{F}(\tilde{\mathbf{x}}) - \tilde{F}(\mathbf{x}^{r,\delta})\big| + \mathbb{E}\big|\tilde{F}(\mathbf{x}^{r,\delta}) - F(\mathbf{x}^{r,\delta})\big| \tag{26}$$

By proposition 5, we have

$$\mathbb{E}\big|F(\mathbf{x}) - \tilde{F}(\mathbf{x})\big| \leq T\sqrt{\frac{2}{\pi}}\,|V|\,\delta \qquad \text{and} \qquad \mathbb{E}\big|\tilde{F}(\mathbf{x}^{r,\delta}) - F(\mathbf{x}^{r,\delta}) \leq \big|T\sqrt{\frac{2}{\pi}}\,|V|\,\delta \tag{27}$$

We turn our attention to the term $\mathbb{E}\big|\tilde{F}(\tilde{\mathbf{x}}) - \tilde{F}(\mathbf{x}^{r,\delta})\big|$. By Propositions 4 and 5 for every u

$$\mathbb{P}(x_u^{r,\delta} \neq \tilde{x}_u) \leq \frac{2T^2}{\pi^{1/4}}\sqrt{\frac{\rho(r)}{\delta}}$$

and therefore, by applying the union bound, for every (u, v), we obtain:

$$\mathbb{P}\big((x_u^{r,\delta}, x_v^{r,\delta}) \neq (\tilde{x}_u, \tilde{x}_v)\big) \leq \frac{4T^2}{\pi^{1/4}}\sqrt{\frac{\rho(r)}{\delta}}$$

We have

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

For any $u,v \in V$,

$$\begin{aligned}
\mathbb{E}[\Phi_{u,v}(\tilde{x}_u,\tilde{x}_v) - \Phi_{u,v}(x_u^{r,\delta},x_v^{r,\delta})] &\leq \mathbb{E}\Big[1_{(x_u^{r,\delta},x_v^{r,\delta})\neq(\tilde{x}_u,\tilde{x}_v)}\Big(\big|\Phi_{u,v}(\tilde{x}_u,\tilde{x}_v)\big| + \big|\Phi_{u,v}(x_u^{r,\delta},x_v^{r,\delta})\big|\Big)\Big] \\
&\leq 2K_\Phi\, \mathbb{P}\big((x_u^{r,\delta},x_v^{r,\delta}) \neq (\tilde{x}_u,\tilde{x}_v)\big)^{1/2} \\
&\leq K_\Phi \frac{2T}{\pi^{1/8}}\Big(\frac{\rho(r)}{\delta}\Big)^{1/4}
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}|\tilde{\Phi}_u(\tilde{x}_u) - \tilde{\Phi}_u(x_u^{r,\delta})| \leq& \mathbb{E}\Big[1_{x_u^{r,\delta}\neq\tilde{x}_u}\big(|\tilde{\Phi}_u(\tilde{x}_u)| + |\tilde{\Phi}_u(x_u^{r,\delta}|)\Big] \\
\leq& \mathbb{E}\Big[1_{x_u^{r,\delta}\neq\tilde{x}_u}\big(|\Phi_u(\tilde{x}_u)| + |\Phi_u(x_u^{r,\delta}|)\Big] + 2\sum_x \mathbb{E}|Z_{u,x}| \\
\leq& P(x_u^{r,\delta} \neq \tilde{x}_u)^{1/2}\, K_\Phi \;+\; 2\delta T\sqrt{\frac{2}{\pi}} \\
\leq& K_\Phi \frac{2T}{\pi^{1/8}}\Big(\frac{\rho(r)}{\delta}\Big)^{1/4} + 2\delta T\sqrt{\frac{2}{\pi}}
\end{aligned}$$

By summing over all nodes and edges, we get:

$$\mathbb{E}[\tilde{F}(\tilde{\mathbf{x}}) - \tilde{F}(\mathbf{x}^{r,\delta}) \leq K_1\Big(\frac{\rho(r)}{\delta}\Big)^{1/4} + \frac{K_2}{2}\delta. \tag{28}$$

Finally, by injecting equations (27) and (28) into equation (26), equation (24) follows. The bound (25) is obtained by direct substitution of δ. The last part of the theorem follows immediately from (25). □

D Establishing the correlation decay property. Coupling technique

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

D.1 Notations

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

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

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

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

4. For simplicity, since 1 is the only action different from the reference action 0, we denote $\mu_{u\leftarrow v_j}(z) \triangleq \mu_{u\leftarrow v_j}(1, z)$.
From equation (3), note the following alternative expression for $\mu_{u\leftarrow v_j}(z)$

$$\begin{aligned}\mu_{u\leftarrow v_j}(z) = \Phi_{u,v_j}(1,1) - \Phi_{u,v_j}(0,1) &+ \max(\Phi_{u,v_j}(1,0) - \Phi_{u,v_j}(1,1), z) \\ &- \max(\Phi_{u,v_j}(0,0) - \Phi_{u,v_j}(0,1), z)\end{aligned} \tag{29}$$

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

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

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

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

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

Note that for any e, $\mathbb{E}|Y_e| \leq K_\Phi$ (see Assumption 2). Equation (11) can be rewritten as

$$B(r) = \mu_u(\mathbf{B}(r-1)) + \Phi_u(1) - \Phi_u(0) \tag{30}$$
$$B'(r) = \mu_u(\mathbf{B}'(r-1)) + \Phi_u(1) - \Phi_u(0) \tag{31}$$

Similarly, we have

$$B_j(r-1) = \mu_{v_j}(\mathbf{B_j}(r-2)) + \phi_{v_j}(1) - \phi_{v_j}(0) \tag{32}$$
$$B'_j(r-1) = \mu_{v_j}(\mathbf{B'_j}(r-2)) + \phi_{v_j}(1) - \phi_{v_j}(0) \tag{33}$$

Finally, equation (29) can be rewritten

$$\mu_{u\leftarrow v_j}(z) = \Phi^3_{u\leftarrow v} + \max(\Phi^1_{u\leftarrow v} z) - \max(\Phi^2_{u\leftarrow v}, z) \tag{34}$$

We call Y_e the *interaction coupling*; Y_e represents how strongly the interaction function $\Phi_{u,v}(x_u, x_v)$ is "coupling" the variables x_u and x_v. In particular, if Y_e is zero, the interaction function $\Phi_{u,v}(x_u, x_v)$ can be decomposed into a sum of two potential functions $\phi_u(x_u) + \phi_v(x_v)$, that is, the edge between

u and v is then be superfluous and can be removed. To see why this is the case, take $\phi_u(0) = 0$, $\phi_u(1) = \Phi_{u,v}(1,0) - \Phi_{u,v}(0,0)$, $\phi_v(0) = \Phi_{u,v}(0,0)$ and $\phi_v(1) = \Phi_{u,v}(0,1)$, which is also equal to $\Phi_{u,v}(1,1) - \Phi_{u,v}(1,0) + \Phi_{u,v}(0,0)$, since $Y_e = 0$.

D.2 Distance-dependent coupling and correlation decay

Definition 3. *A network $\mathcal{G}$ is said to exhibit (a,b)-coupling with parameters (a,b) if for every edge $e = (u,v)$, and every two real values x, x':*

$$\mathbb{P}\Big(\mu_{u\leftarrow v}(x + \Phi_v(1) - \Phi_v(0)) = \mu_{u\leftarrow v}(x' + \Phi_v(1) - \Phi_v(0))\Big) \geq (1-a) - b|x - x'| \tag{35}$$

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

$$\mathbb{P}\Big(\mu_{u\leftarrow v}(x) = \mu_{u\leftarrow v}(x')\Big) \geq (1-a) - b|x - x'| \tag{36}$$

then $\mathcal{G}$ exhibits (a,b) coupling, but in general the tightest coupling values found for equation (36) are much weaker than the ones we would find by analyzing condition (35). In a similar spirit, the following lemma guarantees that the regularization introduced in section 5.2 can only improve coupling:

Lemma 2. *If $\mathcal{G}$ exhibits (a,b) coupling, then for any $\delta \geq 0$, $\tilde{\mathcal{G}}$ also exhibits (a,b) coupling.*

Proof. Let $Z = Z_{v,1} - Z_{v,0}$. Then $\tilde{\Phi}_v(1) - \tilde{\Phi}_v(0) = \Phi_v(1) - \Phi_v(0) + Z$. For any x, x'

$$\mathbb{P}\Big(\mu_{u\leftarrow v}(x + \tilde{\Phi}_v(1) - \tilde{\Phi}_v(0)) = \mu_{u\leftarrow v}(x' + \tilde{\Phi}_v(1) - \tilde{\Phi}_v(0))\Big) =$$
$$\int_z dP_Z(z)\mathbb{P}\Big(\mu_{u\leftarrow v}(x + z + \Phi_v(1) - \Phi_v(0)) = \mu_{u\leftarrow v}(x' + z + \Phi_v(1) - \Phi_v(0))\Big)$$

Applying definition (35) to $x + z$ and $x' + z$, we obtain

$$\begin{aligned}\mathbb{P}\Big(\mu_{u\leftarrow v}(x + \tilde{\Phi}_v(1) - \tilde{\Phi}_v(0)) = \mu_{u\leftarrow v}(x' + \tilde{\Phi}_v(1) - \tilde{\Phi}_v(0))\Big) &\geq \int_z dP_Z(z)\big((1-a) - b|x - x'|\big) \\ &\geq (1-a) - b|x - x'|\end{aligned}$$

$\square$

This form of distance dependent coupling is a useful tool in proving that correlation decay occurs, as illustrated by the following theorem:

Theorem 9. *Suppose $\mathcal{G}$ exhibits (a,b)-coupling. If*

$$a(\Delta - 1) + \sqrt{bK_\Phi}(\Delta - 1)^{3/2} < 1 \tag{37}$$

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

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

$$a(\Delta-1) + bK_Y(\Delta-1)^2 < 1 \tag{38}$$

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

Suppose $\mathcal{G}$ exhibits (a,b)-coupling, that the network is locally tree-like ($\mathcal{B}(u,r)$ is a tree for every $u \in U$ and depth r used for the cavity recursion) and that for all edges $e = (u,v) \in E$, the random variables $(\Phi_e(x,y))_{x,y\in\{0,1\}^2}$ are i.i.d. If

$$(\Delta-1)(a + \sqrt{bK_\Phi}) < 1 \tag{39}$$

then the exponential correlation decay property holds with $\alpha = (\Delta-1)(a + \sqrt{bK_\Phi})$.

D.2.1 Proof of theorem 9

We begin by proving several useful lemmas. The most important, which will use frequently in the rest of the paper, states that in the computation tree of the cavity recursion, the cost function of an edge cost is statistically independent from the subtree below that edge.

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

Note however that Φ_{u,v_j} and $\mathcal{G}[u,k,x]$ are generally dependent when $i \neq k$

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 (5). □

Lemma 4. *For every (u,v), and every two real values x, x'*

$$|\mu_{u\leftarrow v}(x) - \mu_{u\leftarrow v}(x')| \leq |x - x'|. \tag{40}$$

Proof. From (29) we obtain

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

Using twice the relation $\max_x f(x) - \max_x g(x) \leq \max_x(f(x) - g(x))$, 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 5. *For every $u, v \in V$ and every two real values x, x'*

$$|\mu_{u\leftarrow v}(x) - \mu_{u\leftarrow v}(x')| \leq |Y_{u,v}| \tag{41}$$

Proof. Using (29) and (31), we have

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

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

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

Similarly

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

Adding up

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

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

Lemma 6. *Suppose (a,b)-coupling holds. Then,*

$$\mathbb{E}|B(r) - B'(r)| \leq a \sum_{1\leq j\leq d} \mathbb{E}|B_j(r-1) - B'_j(r-1)| + b \sum_{1\leq j\leq d} \mathbb{E}\big[|B_j(r-1) - B'_j(r-1)|^2\big]. \quad (42)$$

Proof. Using (11), we obtain:

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

By Lemma 4, we have $|\mu_{u\leftarrow v_j}(B_j(r-1)) - \mu_{u\leftarrow v_j}(B'_j(r-1))| \leq |B_j(r-1) - B'_j(r-1)|$. Also note from that from equation (32) and (33), $|B_j(r-1) - B'_j(r-1)| = |\mu_{v_j}(\mathbf{B_j}(r-2)) - \mu_{v_j}(\mathbf{B'_j}(r-2))|$; 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,

$$\begin{aligned}&\mathbb{E}\Big[\big|\mu_{u\leftarrow v_j}(B_j(r-1)) - \mu_{u\leftarrow v_j}(B'_j(r-1))\big| \;\Big|\mu_{v_j}(\mathbf{B_j}(r-2), \mu_{v_j}(\mathbf{B'_j}(r-2)\Big] \\ &\leq |B_j(r-1) - B'_j(r-1)|\, \mathbb{P}(\mu_{u\leftarrow v_j}(B_j(r-1)) \neq \mu_{u\leftarrow v_j}(B'_j(r-1)) \mid \mu_{v_j}(\mathbf{B_j}(r-2), \mu_{v_j}(\mathbf{B'_j}(r-2))\end{aligned} \quad (43)$$

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 lemma 3, these are independent from $\mu_{v_j}(\mathbf{B_j}(r-2)$ and $\mu_{v_j}(\mathbf{B'_j}(r-2))$. Thus, by the (a,b) coupling assumption, $\mathbb{P}(\mu_{u\leftarrow v_j}(B_j(r-1)) \neq \mu_{u\leftarrow v_j}(B'_j(r-1)) \mid \mu_{v_j}(\mathbf{B_j}(r-2), \mu_{v_j}(\mathbf{B'_j}(r-2)) \leq a + b|B_j(r-1) - B'_j(r-1)|$. The result then follows. $\square$

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 condition $\mathcal{C}, \mathcal{C}'$

$$\mathbb{E}|\mathrm{CE}(\mathcal{G}, u, r, \mathcal{C}) - \mathrm{CE}(\mathcal{G}, u, r, \mathcal{C}')| \leq K\alpha^r \tag{44}$$

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 (44), where the supremum is over all networks $\mathcal{G}'$ with degree at most Δ, such that the corresponding constant $K_{\Phi'} \leq K_\Phi$, over all nodes u in $\mathcal{G}$ with degree $|\mathcal{N}(u)| \leq \Delta - 1$ and all over all choices of boundary conditions $\mathcal{C}, \mathcal{C}'$. Each condition corresponds to a different recursive inequality for e_r.

Condition (37) Under (37), we claim that

$$e_r \leq a(\Delta - 1)e_{r-1} + b(\Delta - 1)^3 K_\Phi e_{r-2} \tag{45}$$

Applying (32) and (33), we have

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

Thus,

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

By Lemmas 4 and 5 we have $|\mu_{v_j \leftarrow v_{jk}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk}(r-2))| \leq |B_{jk}(r-2) - B'_{jk}(r-2)|$ and $|\mu_{v_j \leftarrow v_{jk}}(B_{jk}(r-2)) - \mu_{v_j \leftarrow v_{jk}}(B'_{jk}(r-2))| \leq |Y_{jk}|$. Also, $d_j \leq \Delta - 1$.Therefore,

$$|B_j(r-1) - B'_j(r-1)|^2 \leq (\Delta - 1) \sum_{1 \leq k \leq d_j} |B_{jk}(r-2) - B'_{jk}(r-2)| \,.\, |Y_{jk}| \tag{46}$$

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

$$\begin{aligned}\mathbb{E}|B_j(r-1) - B'_j(r-1)|^2 \leq &(\Delta - 1) \sum_{1 \leq k \leq d_j} \mathbb{E}|B_{jk}(r-2) - B'_{jk}(r-2)| \,.\, \mathbb{E}|Y_{jk}| \\ \leq &(\Delta - 1)K_\Phi \Big(\sum_{1 \leq k \leq d_j} \mathbb{E}|B_{jk}(r-2) - B'_{jk}(r-2)| \Big) \\ \leq &(\Delta - 1)^2 K_\Phi e_{r-2}\end{aligned} \tag{47}$$

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

6 and the definition of e_r, we obtain

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

This implies (45).

From (45) we obtain that $e_r \le K\alpha^r$ for $K=\Delta K_\Phi$ and α given as the largest in absolute value root of the quadratic equation $\alpha^2 = a(\Delta-1)\alpha + b(\Delta-1)^3K_\Phi$. We find this root to be

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

where the last inequality follows from assumption (37). 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 (30) and (31) we have

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

Applying again Lemma 4, the right-hand side is at most

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

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

Condition (38) Recall from lemma 6 that for all r, we have:

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

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, $|\mu_{v_j\leftarrow v_{jk}}(B_{jk})-\mu_{v_j\leftarrow v_{jk}}(B'_{jk})| \le |Y_{jk}| \le K_Y$ (the second inequality follows from Lemma 5, the third by assumption). As a result,

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

We obtain:

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

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

Condition (39) Since the network is locally tree-like, all Y_e encountered in the cavity recursion $\mathrm{CE}(\mathcal{G},u,r)$ are independent. Another important observation is the following: since $(\Phi_e(x,y))_{x,y\in\{0,1\}^2})$ are i.i.d., the random variables Y_e are symmetric (Y_e and $-Y_e$ have the same distribution). The next step of the proof is to observe that (34) can be rewritten as follows:

$$\begin{aligned}
\mu_{u\leftarrow v_j}(z) &= \Phi^3_{u\leftarrow v_j} + \max(\frac{X_{u\leftarrow v_j} - Y_{u,v_j}}{2}, z) - \max(\frac{X_{u\leftarrow v_j} + Y_{u,v_j}}{2}, z) \\
&= \Phi^3_{u\leftarrow v_j} + \mathrm{sign}(Y)\Big(\max(\frac{X_{u\leftarrow v_j} - |Y_{u,v_j}|}{2}, z) - \max(m\frac{X_{u\leftarrow v_j} + |Y_{u,v_j}|}{2}, z)\Big) \\
&= \Phi^3_{u\leftarrow v_j} + \mathrm{sign}(Y_{u,v_j}) h(X_{u\leftarrow v_j}, |Y_{u,v_j}|, z)
\end{aligned}$$

where $h(x,y,b) \stackrel{\Delta}{=} \max(\frac{1}{2}(x-y),b) - \max(\frac{1}{2}(x+y),b)$ is a nondecreasing function of b for $y \geq 0$. It follows that for two real numbers B_j, B'_j,

$$\mu_{u\leftarrow v_j}(B_j) - \mu_{u\leftarrow v_j}(B'_j) = \mathrm{sign}(Y_{u,v})\big(h(X_{u\leftarrow v_j}, |Y_{u,v}|, B_j) - h(X_{u\leftarrow v_j}, |Y_{u,v}|, B'_j)\big)$$

and so the sign of $\mu_{u\leftarrow v_j}(B_j) - \mu_{u\leftarrow v_j}(B'_j)$ is the product of the sign of $Y_{u,v}$ and the sign of $B_j - B'_j$. For a symmetric random variable Y, $\epsilon = \mathrm{sign}(Y)$ and $|Y|$ are independent, and $P(\epsilon = +1) = P(\epsilon = -1) = \frac{1}{2}$. For all $j = 1, \ldots d$ and $k = 1, \ldots n_j$, let

$$h_{jk} = (h(X_{v_j\leftarrow v_{jk}}, |Y_{v_j,v_{jk}}|, B_{jk}) - h(X_{v_j\leftarrow v_{jk}}, |Y_{v_j,v_{jk}}|, B'_{jk})$$

and

$$\epsilon_{jk} = \mathrm{sign}(Y_{v_j,v_{jk}})$$

For any (j,k),

$$\mu_{v_j\leftarrow v_{jk}}(B_{jk}(r-2)) - \mu_{v_j\leftarrow v_{jk}}(B'_{jk}(r-2)) = \epsilon_{jk} h_{jk}$$

Let $\mathcal{F}$ the σ-field generated by the set of random variables $\{(B_{jk}(r-1)), (B'_{jk}(r-1)), (|Y_{v_j,v_{jk}}|), (X_{v_j\leftarrow v_{jk}})\}$. Note that h_{jk} is measurable with respect to $\mathcal{F}$. We obtain:

$$\mathbb{E}|B_j(r-1) - B'_j(r-1)|^2 = \mathbb{E}\left[\mathbb{E}\Big[|\sum_{1\leq k\leq d_j} \epsilon_{jk} h_{jk})|^2 \Big| \mathcal{F}\Big]\right]$$

Conditional on $\mathcal{F}$, the inner expectation (which is taken only w.r.t the ϵ_{jk}) is simply the variance of the random variable $\sum_k \epsilon_k c_k$, where ϵ_k are independent variables, and $c_k = h_{jk}$ are fixed constants. The variance is then $\sum_k c_k^2 = \sum_k (h_{jk})^2$.

Therefore:

$$\mathbb{E}|B_j(r-1) - B'_j(r-1)|^2 \leq \mathbb{E}\big[\sum_{1\leq k\leq d_j} (h_{jk})^2\big]$$

Using inequalities $|h_{jk}| \leq |Y_{v_j,v_{jk}}|$, $|h_{jk}| \leq |B_{jk} - B'_{jk}|$, and $\mathbb{E}|Y_{v_j,v_{jk}}| \leq K_\Phi$, we obtain:

$$\mathbb{E}|B_j(r-1) - B'_j(r-1)|^2 \leq K_\Phi \sum_k \mathbb{E}|B_{jk}(r-2) - B'_{jk}(r-2)|$$

Therefore, using the same notations as previously, this implies:

$$e_r \leq a(\Delta - 1)e_{r-1} + bK_\Phi(\Delta - 1)^2 e_{r-2}$$

which, given $(\Delta - 1)(a + \sqrt{K_\Phi b}) < 1$, implies correlation decay at the desired rate.

D.3 Establishing coupling bounds

D.3.1 Coupling Lemma

Theorem 9 details sufficient condition under which the distance-dependent coupling induces correlation decay (and thus efficient decentralized algorithms, vis-à-vis Proposition 3 and Theorem 8). It remains to show how can we prove coupling bounds. The following simple observation can be used to achieve this goal.
For any edge $(u, v) \in \mathcal{G}$, and any two real numbers x, x', consider the following events

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

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

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

Lemma 7. *If $E_{u\leftarrow v}(x, x')$ occurs, then $\mu_{u\leftarrow v}(x + \Phi_v(1) - \Phi_v(0)) = \mu_{u\leftarrow v}(x' + \Phi_v(1) - \Phi_v(0))$. Therefore*

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

Proof. The result is obtained directly from representation (29). □

Note that Lemma 7 implies that the probability of coupling not occuring $P(\mu_{u\leftarrow v}(x + \Phi_v(1) - \Phi_v(0)) \neq \mu_{u\leftarrow v}(x' + \Phi_v(1) - \Phi_v(0)))$ is upper bounded by the probability of $(E_{u\leftarrow v}(x, x'))^c$. When obvious from context, we drop the subscript $u \leftarrow v$. We will often use the following description of $(E_(x, x'))^c$: for two real values $x \geq x'$,

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

D.3.2 Uniform Distribution: proof of Theorem 1

For a given family of distribution, all remains to do in order to prove a correlation decay theorem is to compute the coupling parameters a, b for this distribution and apply one form of Theorem 9. In this section we compute the coupling parameters of the uniform distribution; together with the second condition of Theorem 9, this proves Theorem **1**.

Lemma 8. *The network with uniformly distributed rewards described in section 3.1.1 exhibits (a, b) coupling with*

$$a = \frac{I_2}{2I_1} \qquad \textit{and} \qquad b = \frac{1}{2I_1}$$

Proof. Since the distribution of each $\Phi_e(x,y)$ is distributed as a uniform random variable over $[-I_2, I_2]$, it follows that for any fixed edge $(u,v) \in \mathcal{G}$, $\Phi^1_{u\leftarrow v}$ and $\Phi^2_{u\leftarrow v}$ are i.i.d. random variables with a triangular distribution with support $[-2I_2, 2I_2]$ and mode 0. Because $\Phi^1_{u\leftarrow v}$ and $\Phi^2_{u\leftarrow v}$ are i.i.d., by symmetry we obtain:

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

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

$$P(E(x,x')^c) \leq \frac{x-x'}{2I_1} + \frac{1}{I_1}\int_{-2I_2}^{2I_2} dP_{\Phi^1}(a^1)\int_{a^1}^{2I_2} dP_{\Phi^2}(a^2)(a^2-a^1)$$

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

$$\begin{aligned}
\int_{a^1}^{2I_2} dP_{\Phi^2}(a^2)(a^2-a^1) &= \frac{1}{4I_2^2}\int_{a^1}^{2I_2}(2I_2-a^2)(a^2-a^1)\, d(a^2) \\
&= \frac{1}{4I_2^2}\Big(-\int_{a^1}^{2I_2}(2I_2-a^2)^2 d(a^2) + (2I_2-a^1)\int_{a^1}^{2I_2}(2I_2-a^2)d(a^2)\Big) \\
&= \frac{1}{4I_2^2}\Big(-\frac{1}{3}(2I_2-a^1)^3+\frac{1}{2}(2I_2-a_1)^3\Big) = \frac{1}{24I_2^2}(2I_2-a^1)^3
\end{aligned}$$

Similarly, for $a^1 \leq 0$,

$$\int_{a^1}^{2I_2} dP_{\Phi^2}(a^2)(a^2-a^1) = -a_1 + \frac{1}{24I_2^2}(a^1+2I_2)^3$$

The final integral is therefore equal to:

$$\begin{aligned}
&\int_{-2I_2}^{2I_2} dP_{\Phi^1}(a^1)\int_{a^1}^{2I_2} dP_{\Phi^2}(a^2)(a^2-a^1) \\
= &\frac{1}{4I_2^2}\Big(\int_{-2I_2}^{0}\big((a^1+2I_2)(-a^1+\frac{1}{24I_2^2}(a^1+2I_2)^3\big)d(a^1) + \int_0^{2I_2}\frac{1}{24I_2^2}(2I_2-a^1)^4 d(a^1)\Big) \\
= &\frac{1}{4I_2^2}\Big(\frac{24}{15}I_2^3+\frac{4}{15}I_2^3\Big) = \frac{7}{15}I_2
\end{aligned}$$

Finally,

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

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

Note that for all edges, $|Y_e| \leq 4I_2$, so that the condition $\beta(\Delta-1)^2 < 1$ implies $\frac{I_2}{2I_1}(\Delta-1) + \frac{4I_2}{2I_1}(\Delta-1)^2 < 1$.

D.3.3 Gaussian distribution: proof of Theorem 2

In this section, we compute the coupling parameters for Gaussian distributed reward functions. Rather than considering only the assumptions of Theorem **2**, we place ourselves in a more general framework; the proof will follow from the application of the first form of Theorem 9 and a special case of the computation detailed below (see corollary 2).
Assume that for any edge $e=(u,v)$ the value functions $(\Phi_{u,v}(0,0), \Phi_{u,v}(0,1), \Phi_{u,v}(1,0), \Phi_{u,v}(1,1))$ are independent, identically distributed four-dimensional Gaussian random variables, with mean $\mu = (\mu_i)_{i\in\{00,01,10,11\}}$, and covariance matrix $S = (S_{ij})_{i,j\in\{00,01,10,11\}}$. For every node $v \in V$, suppose $\Phi_v(1) = 0$ and that $\Phi_v(0)$ is a Gaussian random variable with mean μ_p and standard deviation σ_p. Moreover, suppose all the Φ_v and Φ_e are independent for $v \in V$, $e \in E$. Let

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

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

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

Lemma 9. *Assume $C < 1$. Then the network exhibits coupling with parameters (a, b) equal to:*

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

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

Proof. Under the conditions of corollary 2, 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 Lemma 9, the network exhibits coupling with parameters

$$a = \frac{1}{\pi}\arctan\left(\sqrt{\frac{\sigma_e^2}{\sigma_e^2+\sigma_p^2}}\right) \leq \frac{1}{\pi}\beta \leq \beta$$
$$b = \sqrt{\frac{1}{2\pi}}\frac{1}{\sqrt{\sigma_e^2+\sigma_p^2}} \text{ and so, } bK_\Phi \leq \sqrt{\frac{2}{\pi}}\beta \leq \beta$$

□

Remark that when $\sigma_e \to 0$, $\beta \to 0$ and correlation decay appears;

Proof of Lemma 9 . Fix an edge (u, v) in E; for simplicity, in the rest of this section denote $\bar{\Phi}^1 = \Phi^1_{u\leftarrow v} + \Phi_v(0) - \Phi_v(1)$ and $\bar{\Phi}^2 = \Phi^2_{u\leftarrow v} + \Phi_v(0) - \Phi_v(1)$. It follows that $(\overline{\Phi}^1, \overline{\Phi}^2)$ follows a bivariate

Gaussian distribution with mean (μ_1, μ_2):

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

and covariance matrix

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

with σ_1, σ_2, ρ defined as previously. With a slight abuse of notation, let $X = \overline{\Phi}^1 + \overline{\Phi}^2$, $Y = \overline{\Phi}^2 - \overline{\Phi}^1$. Then, (X, Y) is a bivariate Gaussian vector with means $\mathbb{E}[X] = \mu_1 + \mu_2$ and $\mathbb{E}[Y] = \mu_2 - \mu_1$, standard deviations σ_X, σ_Y and correlation C as defined previously. Denote also $\overline{X} \triangleq X - E[X]$ and $\overline{Y} \triangleq Y - E[Y]$ the centered versions of X and Y. Consider two real numbers $x \geq x'$, and let (b, t) be the two real numbers such that $x = b + t/2$, $x' = b - t/2$. By the coupling lemma, as well as the definitions of $b, t, \overline{\Phi}^1$ and $\overline{\Phi}^2$, we have

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

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

Lemma 10.

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

Proof.

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

□

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

$$\begin{aligned}
\mathbb{P}((E)^c(x, x')) =& \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - C^2}} \int_{S(t)} e^{-\frac{1}{2(1-C^2)}(\frac{(x-\mu_1-\mu_2+2b)^2}{\sigma_x^2} + \frac{(y-\mu_2+\mu_1)^2}{\sigma_y^2} - 2C\frac{(x-\mu_1-\mu_2+2b)(y+\mu_2-\mu_1)}{\sigma_x\sigma_y})} dxdy \\
=& \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - C^2}} \int_y e^{-\frac{1}{2(1-C^2)} \frac{(y-\mu_2+\mu_1)^2}{\sigma_y^2}} g(y)dy \qquad (48)
\end{aligned}$$

where:

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

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:

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

Now, $\tilde{x_b} - C\tilde{y} = \frac{x-\mu_1-\mu_2+2b-\frac{C\sigma_x(y-\mu_2+\mu_1)}{\sigma_y}}{\sigma_x}$. Recall Anderson's inequality [**?**]: for an arbitrary set S and vector z we denote $S+z$ the set $\{(x+z) : x \in S\}$. 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

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

Injecting that bound in equation (48), we obtain:

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

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

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

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

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

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,

$$\begin{aligned}\mathbb{P}(S_{\text{out}}(t')) =&4\mathbb{P}(\{(X,Y) : X \geq t, Y \geq 0, Y \geq X-t\}) \\ =&\frac{2}{\pi\sigma_x\sigma_y\sqrt{1-C^2}}\int_{\{(X,Y):X\geq t,Y\geq 0,Y\geq X-t\}} e^{-\frac{1}{2(1-C^2)}(\frac{x^2}{\sigma_x^2}+(1-C^2)\frac{y^2}{\sigma_y^2})}\,dxdy\end{aligned}$$

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

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

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

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

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

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

$$\mathbb{P}((E)^c(x,x')) \leq \left(\frac{1}{\pi}\arctan(\frac{\sigma_y}{\sigma_x\sqrt{1-C^2}}) + \sqrt{\frac{2}{\pi(1-C^2)}}\frac{|\mu_2-\mu_1|}{\sigma_x}\right) + \sqrt{\frac{2}{\pi(1-C^2)}}\frac{t}{\sigma_x} \tag{49}$$

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

□

E Proofs of results in section 3.2

E.1 Concentration argument for Theorem 3

We now return to the proof of Theorem 3. Recall the description of the algorithm in the end of Section 7.1. Specifically, we run $CE(t,\epsilon)$ algorithm for some even t to be specified later. Let $\mathcal{I}$ be the set of nodes i in $\mathcal{G}_0$ such that $C^-_{\mathcal{G}_0}(i,t) > 0$. Since t is even, by Lemma 1, $\mathcal{I} \subset I^*_0$, where we recall that I^*_0 is the largest weighted independent set in $\mathcal{G}_0$. Thus we need to bound $|W(I^*) - W(I^*_0)|$ and $W(I^*_0 \setminus \mathcal{I})$ and show that both quantities are small. Let ΔV_0 be the set of nodes in $\mathcal{G}$ which are not in $\mathcal{G}_0$. Trivially, $|W(I^*) - W(I^*_0)| \leq W(\Delta V_0)$. We have $\mathbb{E}[\Delta V_0] = \delta n$, and since the nodes were deleted irrespectively of their weights, then $\mathbb{E}[W(\Delta V_0)] = \delta n$.

To analyze $W(I^*_0 \setminus \mathcal{I})$, observe that by (second part of) Proposition 7, for every node $i, \mathbb{P}(i \in I^*_0 \setminus \mathcal{I}) \leq 4(1-\delta)^t \equiv \delta_1$. Thus $\mathbb{E}|I^*_0 \setminus \mathcal{I}| \leq \delta_1 n$. In order to obtain a bound on $W(I^*_0 \setminus \mathcal{I})$ we obtain 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 $\mathcal{G}_0$ with weights greater than C. We have $\mathbb{E}[W(V_C)] \leq (C + E[W - C|W > C]) \exp(-C)n = (C+1)\exp(-C)n$, where W is a generic random variable with $\exp(1)$ distribution. Then remaining nodes have a weight at most C. Therefore,

$$\mathbb{E}[W(I_0^* \setminus \mathcal{I})] \leq C\delta_1 n + (C+1)\exp(-C)n.$$

We conclude

$$\mathbb{E}[|W(I^*) - W(\mathcal{I})|] \leq \delta n + C\delta_1 n + (C+1)\exp(-C)n. \tag{50}$$

Now we obtain a lower bound on $W(I^*)$. Consider the standard greedy algorithm for generating an independent set: take arbitrary node, remove neighbors, 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 $n/4$. By Chebyshev's inequality

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

We now summarize the results.

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

where we have used Markov's inequality in the last step and $\delta_1 = 4(1-\delta)^t$. Thus it suffices to arrange δ and C so that the first ratio is at most $\epsilon/2$ and assuming, without the loss of generality, that $n \geq 128/\epsilon$, we will obtain that the sum is at most ϵ. It is a simple exercise to show that by taking $\delta = O(\epsilon^2), t = O(\log(1/\epsilon)/\epsilon^2)$ and $C = O(\log(1/\epsilon))$, we obtain the required result. This completes the proof of Theorem 3. □

E.2 Proofs of Theorems 4,

In this section we present a proof of Theorems 4.

Proof of Theorem 4. The mixture of Δ exponential distributions with rates $\alpha_j, 1 \leq j \leq \Delta$ and equal weights $1/\Delta$ can be viewed as first randomly generating a rate α with the probability law $\mathbb{P}(\alpha = \alpha_j) = 1/\Delta$ and then randomly generating exponentially distributed random variable with rate α_j, conditional on the rate being α_j.

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

Lemma 11. *Fix any subgraph $\mathcal{H}$, node $i \in \mathcal{H}$ with $N_{\mathcal{H}}(i) = \{i_1, \ldots, i_r\}$. Then*

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

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

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

Thus,

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

The other equalities follow identically. $\square$

By taking differences, we obtain

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

We now use identity

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

which further implies

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

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

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

This implies

$$|M^{-,j}_{\mathcal{H}}(i,t) - M^{+,j}_{\mathcal{H}}(i,t)| \leq \frac{r}{\Delta} \sum_{1\leq k\leq m} \frac{\alpha_j}{\alpha_j + \alpha_k} \max_{1\leq l\leq r} |M^{-,k}_{\mathcal{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l, t-1) - M^{+,k}_{\mathcal{H}\setminus\{i,i_1,\ldots,i_{l-1}\}}(i_l, t-1)| \quad (51)$$

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

$$e_{t,j} = \sup_{\mathcal{H}\subset\mathcal{G}, i\in\mathcal{H}} |M^{-,j}_{\mathcal{H}}(i,t) - M^{+,j}_{\mathcal{H}}(i,t)| \quad (52)$$

By taking maximum on the right and left hand side successively, inequality (51) implies

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

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

$$\mathbf{e_t} \leq M\mathbf{e_{t-1}}.$$

Therefore, if M^t converges to zero exponentially fast in each coordinate, then also $\mathbf{e_t}$ 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 by CE$(t, 0)$ is exponentially small in t . Note that $\frac{r}{\Delta} \leq 1$. Recall that $\alpha_j = \rho^j$. Therefore, for each j, k, we have $M_{j,k} \leq \frac{\rho^j}{\rho^j+\rho^k}$. It then suffices to show that M^t_Δ converges to zero exponentially fast, where where M_Δ is 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 $1 \leq j, k \leq \Delta$.

Proof of theorem 4 will thus be completed with the proof of the following lemma:

Lemma 12. *Under the condition $\rho > 17$, there exists $\delta = \delta(\rho) < 1$ such that the absolute value of every entry of M^t_Δ is at most $\delta^t(\rho)$.*

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

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

It is easy to verify that when $\rho > 17$, that is $\epsilon < 1/17$, $(1/2 + 2\sqrt{\frac{\epsilon}{1-\epsilon}}) < 1$, and the proof would be complete. Fix $1 \leq j \leq \Delta$. Then,

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

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 and of the theorem. □

□

E.3 Hardness Result (joint work with David Goldberg)

We turn to our third and last result - the hardness of approximating $W(I^*)$ when the weights are exponentially distributed and the degree of the graph is large. We need to keep in mind that since we dealing with instances which are random (in terms of weights) and worst-case (in terms of the underlying graph) at the same time, we need to be careful as to the notion of hardness we use. In fact we will prove a result using the standard (non-average case) notions of complexity theory.

Theorem 10. *There exist Δ_0 and c_1^*, c_2^* such that for all $\Delta \geq \Delta_0$ the problem of computing $\mathbb{E}[W(I^*)]$ to within a multiplicative factor $\rho = \Delta/(c_1^* \log \Delta 2^{c_2^* \sqrt{\log \Delta}})$ for graphs with degree at most Δ is NP-complete.*

Remark : One can in principle compute a concrete Δ_0 such that for all $\Delta \geq \Delta_0$ the claim of the theorem holds. But computing such Δ_0 explicitly does not seem to offer much insight. We note that in the related work by Trevisan [**?**], no attempt is made to compute a similar bound either.

Proof of Theorem 10. We only provide a sketch of the proof. Given a graph $\mathcal{G}$ with degree bounded by Δ, let $\mathcal{I}$ denote (any) maximum cardinality independent set, and let $\mathcal{I}^*$ denote the unique maximum weight independent set corresponding to i.i.d. weights with $\mathbb{E}(1)$ distribution. We make use of the following result due to Trevisan [**?**].

Theorem 11. *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.

Proposition 9. *For every graph $\mathcal{G}$ with n large enough,*

$$\frac{1}{3\log\Delta} \le \frac{\mathbb{E}[|\mathcal{I}^*|]}{|\mathcal{I}|} \le 1$$

This in combination with Theorem 11 leads to the desired result.

Proof sketch. Let $W(1) < W(2) < \cdots < W(n)$ be the ordered weights associated with our graph $\mathcal{G}$. Fix δ, which we later will choose to be $1/(3\log\Delta)$. let $m = \lceil \delta|\mathcal{I}|\rceil$. Observe that the event $|\mathcal{I}^*| < \delta|\mathcal{I}|$ implies that $\sum_{n-m+1\le j\le n} W(j) \ge \sum_{i\in\mathcal{I}^*} W_i \ge \sum_{i\in\mathcal{I}} W_i$. The exponential distribution implies $\mathbb{E}[W(j)] = H(n) - H(n-j)$, where $H(k)$ is the harmonic sum $1 + 1/2 + \ldots + 1/k = \log(k) + O(1)$. Thus

$$\begin{aligned}
\sum_{n-m+1\le j\le n} \mathbb{E}[W_j] &= \sum_{n-m+1\le j\le n} (H(n) - H(n-j)) \\
&= mH(n) - \sum_{j\le m-1} H(j) \\
&\le mH(n) - m\log(m) + O(m) \\
&= \delta|\mathcal{I}|\log\frac{n}{\delta|\mathcal{I}|} + O(m) \\
&\le \delta|\mathcal{I}|(\log\frac{\Delta+1}{\delta} + O(1)),
\end{aligned}$$

where a straightforward bound $|\mathcal{I}| \ge n/(\Delta+1)$ is used. It is easy to check that for $\delta = 1/(C\log\Delta)$ with sufficiently large universal constant C, we have $\delta(O(1) + \log\frac{\Delta+1}{\delta}) < 1$ implying

$$\sum_{n-m+1\le j\le n} \mathbb{E}[W_j] - \mathbb{E}[W(\mathcal{I})] \le |\mathcal{I}|(\delta(O(1) + \log\frac{\Delta+1}{\delta}) - 1) < -\delta_2|I|,$$

for some constant $\delta_2 = \delta_2(\Delta)$. Our next step is to use standard methods to show that $\sum_{n-m\le j\le n} W_j$ is concentrated around its mean. This together with the previous bound implies that the probability of the event $\sum_{n-m\le j\le n} W_j > W(\mathcal{I})$ converges to zero at the rate at least $O(1/n)$. This is used in the final step to argue that $\mathbb{E}[W(\mathcal{I}^*)]/|\mathcal{I}| \ge \frac{1}{O(\log(\Delta))}$. □

□