Proof of Proposition 3

We first see that there always exists a lower bound of α such that the distance term becomes smaller (or equivalent) after AM for an initial $\boldsymbol{y}_0$, *i.e.*,

$$d_h(\boldsymbol{x}^*, \boldsymbol{y}^*) \leq d_h(\boldsymbol{x}_1, \boldsymbol{y}_0)$$

where $\boldsymbol{x}_1 \in \operatorname{argmin}_{\boldsymbol{x}} \mathcal{E}_\alpha(\boldsymbol{x}, \boldsymbol{y_0})$ and pair $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is the output of AM with inputs $(\alpha, \boldsymbol{y_0})$. First, it is obvious that $d_h(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is a non-increasing function for α (while $E(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is non-decreasing). If $\alpha = 0$, then $\boldsymbol{x}^* \in \operatorname{argmin}_{\boldsymbol{x}} f(\mathcal{S}(\boldsymbol{x}))$ and $\boldsymbol{y}^* \in \operatorname{argmin}_{\boldsymbol{y}} \sum_{(i,j)\in\mathcal{E}} \psi_{ij}(y_i, y_j)$ and thus $d_h(\boldsymbol{x}^*, \boldsymbol{y}^*)$ takes some finite value. Meanwhile, if $\alpha \to \infty$, then $d_h(\boldsymbol{x}^*, \boldsymbol{y}^*)$ becomes 0. Therefore, the statement holds (we denote by $\tilde{\alpha}$ the lower bound). Since we can find a small value of E for an initial with a smaller distance term for a common α, the statement of the proposition follows from $d_h(\boldsymbol{x}_1^*, \boldsymbol{y}_1^*) < d_h(\boldsymbol{x}_1, \boldsymbol{y}_0)$ from the above statement.

Proof of Proposition 4

For the given $\boldsymbol{\delta}^*$ and $\boldsymbol{x}^*$, it is obvious that

$$\begin{aligned} L(\boldsymbol{\delta}^*) &= E(\boldsymbol{x}^*, \boldsymbol{x}^*) + \boldsymbol{\delta}^*(\boldsymbol{x}^* - \boldsymbol{x}^*) = E(\boldsymbol{x}^*, \boldsymbol{x}^*) \\ &\geq \min_{\boldsymbol{x},\boldsymbol{y}\in\{0,1\}^{\mathcal{V}}, x_i = y_i (i\in\mathcal{V})} E(\boldsymbol{x}, \boldsymbol{y}) = \min_{\boldsymbol{x}\in\{0,1\}^{\mathcal{V}}} E(\boldsymbol{x}). \end{aligned}$$

Meanwhile, since $L(\boldsymbol{\delta})$ is the Lagrangian relaxation of the original problem, we always have

$$L(\boldsymbol{\delta}^*) \leq \min_{\boldsymbol{x}\in\{0,1\}^{\mathcal{V}}} E(\boldsymbol{x}). \tag{12}$$

Thus, taking the above two equations together, we have the equality in Eq. (12), which shows the statement of the proposition.

Proof of Lemma 5

This result directly follows from the fact that the modular upper bound $m^f(X)$ is an approximation of f such that [15],

$$f(X) \leq m^f(X) \leq \frac{|X|}{1 + (|X|-1)(1-\kappa_f(X))} f(X) \tag{13}$$

where $\kappa_f(X)$ is the curvature of f [14]. In the worst case, this factor is $|X|$. Now let $\mathbf{x}^*$ be the optimal solution, and $\mathcal{S}(\mathbf{x}^*)$ be the corresponding set. Denote $\alpha(X) = \frac{|X|}{1+(|X|-1)(1-\kappa_f(X))}$, and let $\hat{\mathbf{x}}$ be the exact solution to the problem $\min_{\mathbf{x}} m^f(\mathcal{S}(\mathbf{x})) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(x_i, x_j)$. The following chain of inequalities hold:

$$\begin{aligned} & f(\mathcal{S}(\hat{\mathbf{x}})) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(\hat{x}_i, \hat{x}_j) \\ & \leq m^f(\mathcal{S}(\hat{\mathbf{x}})) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(\hat{x}_i, \hat{x}_j) \\ & \leq m^f(\mathcal{S}(\mathbf{x}^*)) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(x_i^*, x_j^*) \\ & \leq \alpha(S(\mathbf{x}^*)) f(S(\mathbf{x}^*)) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(x_i^*, x_j^*) \\ & \leq \alpha(S(\mathbf{x}^*)) [f(S(\mathbf{x}^*)) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(x_i^*, x_j^*)] \end{aligned}$$

Hence this provides a $\alpha(S(\mathbf{x}^*)) \leq |S(\mathbf{x}^*)|$ approximation.

Proof of Lemma 6

In this case, we assume we are given the maximization problem,

$$\max_{\mathbf{x}} f(S(\mathbf{x})) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(x_i, x_j) \tag{14}$$

where ψ is a submodular tree, and f is a supermodular function. Note that this is equivalent to the original problem, just changing the min to a max, and correspondingly interchanging the submodularity and supermodularities. This is different from the original problem in the sense that simple interchanging the max and min (which can be done my adding a minus sign), changes the signs of the submodular function. In order to ensure that the functions f and ψ are positive even after changing the sign, we would need to shift the functions.

Assuming this is done, we can provide an approximation guarantee for this setup. In this case, we use a simple surrogate for the submodular function ψ. Since we assume ψ is monotone submodular, it is easy to see that, $\psi_{ij}(x_i, x_j) \leq \psi_{ij}(x_i) + \psi_{ij}(x_j) \leq 2\psi_{ij}(x_i, x_j)$. The algorithm then just uses the function $\psi_{ij}(x_i) + \psi_{ij}(x_j)$ as a surrogate, and solves the problem $\max_{\mathbf{x}} f(S(\mathbf{x})) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(x_i) + \psi_{ij}(x_j)$. Since f is supermodular, this is submodular minimization, which can be performed exactly. Again, let $\hat{\mathbf{x}}$ be the solution using the surrogate function, and $\mathbf{x}^*$ be the optimal solution. Then, the following chain of inequalities hold:

$$
\begin{aligned}
& f(\mathcal{S}(\hat{\mathbf{x}})) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(\hat{x}_i, \hat{x}_j) \\
& \geq f(\mathcal{S}(\hat{\mathbf{x}})) + \sum_{i,j\in\mathcal{E}} \frac{1}{2}[\psi_{ij}(\hat{x}_i) + \psi_{ij}(\hat{x}_j)] \\
& \geq \frac{1}{2}\{f(\mathcal{S}(\hat{\mathbf{x}})) + \sum_{i,j\in\mathcal{E}} [\psi_{ij}(\hat{x}_i) + \psi_{ij}(\hat{x}_j)]\} \\
& \geq \frac{1}{2}\{f(\mathcal{S}(\mathbf{x}^*)) + \sum_{i,j\in\mathcal{E}} [\psi_{ij}(x_i^*) + \psi_{ij}(x_j^*)]\} \\
& \geq \frac{1}{2}\{f(\mathcal{S}(\mathbf{x}^*)) + \sum_{i,j\in\mathcal{E}} \psi_{ij}(x_i^*, x_j^*)\}
\end{aligned}
$$

Hence this provides a $1/2$ approximation.

Proof of Theorem 7

To prove Theorem 7, we need the following lemmas:

Lemma 8. *Let f be a submodular function. For any $\beta \in \mathbb{R}$ and $\mathbf{b} \in \mathbb{R}^{\mathcal{V}}_{>0}$, $\mathbf{t}^*$ is optimal for $\min_{t\in P(f)} \frac{w_i(t_i)}{b_i}$ if and only if $\mathbf{t}^* + \beta\mathbf{b}$ is optimal for $\min_{\mathbf{t}\in P(f+\beta\mathbf{b})} \frac{w_i(t_i)}{b_i}$, where $w_i : \mathbb{R} \to \mathbb{R}$.*

Lemma 9. *Let f be a submodular function with $f(\emptyset) = 0$. And, let $\mathbf{x} \in [0,1]^{\mathcal{V}}$ with unique values $u_1 > \cdots > u_l$, taken at sets $\mathcal{A}_1, \ldots, \mathcal{A}_l$. Then, for $\mathbf{c} \in \mathbb{R}^{\mathcal{V}}_{<0}$, $\mathbf{s}$ is optimal for $\max_{\mathbf{s}\in P(f^{\mathbf{c}})} \mathbf{x}^\top\mathbf{s}$ if and only if $\mathbf{s}(\mathcal{A}_1\cup\cdots\cup\mathcal{A}_i) = f(\mathcal{A}_1\cup\cdots\cup\mathcal{A}_i)$ for all $i = 1, \ldots, l-1$.*

Proof of Lemma 9. It is obvious that $(\mathbf{s}+\mathbf{c})(\mathcal{A}_1\cup\cdots\cup\mathcal{A}_i) = (f+\mathbf{c})(\mathcal{A}\cup\cdots\cup\mathcal{A}_i)$ if and only if $\mathbf{s}(\mathcal{A}_1\cup\cdots\cup\mathcal{A}_i) = f(\mathcal{A}\cup\cdots\cup\mathcal{A}_i)$ for all $i = 1, \ldots, l-1$. Therefore, the statement follows from Lemma 8 and, for example, Proposition 4.2 in [2] because $P((f+\mathbf{c})^{\mathbf{c}}) = P(f+\mathbf{c})$. □

Now, we have the proof of Theorem 7 as follows.

Proof of Theorem 7. Since $\mathbf{t}^* = 2\alpha(\mathbf{a} - \tilde{\mathbf{x}}^*)$, we know from Lemma 9 that the dual problem of Eq. (10) is

$$\max_{\mathbf{s}\in P(f^{2\lambda(\mathbf{a}-\mathbf{1}_{|\mathcal{V}|})})} -\sum_{i\in\mathcal{V}} s_i^2/(4\lambda) + s_i a_i. \tag{15}$$

Let $\psi_i(\tilde{x}_i) = \lambda(\tilde{x}_i - a_i)^2$ and $\psi_i^*(-t_i) = t_i^2/(4\lambda) - t_i a_i$. Also, for $i \in \mathcal{V}$, let s_i^* be a maximizer of $-\psi_i^*(-s_i)$ over $(-\infty, \max(t_i^*, 2\lambda(a_i - 1))]$. Then, the pair $(\mathbf{x}^*, \mathbf{s}^*)$ is optimal for Eq. (10) and Eq. (15) if and only if (a)

$$(\eta_i(x_i^*, s_i^*) :=)\ x_i^* s_i^* + \psi_i(x_i^*) + \psi_i^*(-s_i^*) = 0$$

and (b) $f(\tilde{\mathbf{x}}^*) = (\mathbf{s}^*)^\top\mathbf{x}^*$.

For i such that $\tilde{x}_i^* < 0$ (*i.e.*, $x_i^* = 0$), we have $t_i^* = 2\lambda(a_i - \tilde{x}_i)$ $(> 2\lambda(a_i - 1))$ and thus $s_i^* = 2\lambda a_i$. Hence, (a) is met because $\psi_i(0) = \lambda a_i^2$. For i such that $0 \leq \tilde{x}_i^* \leq 1$, (a) is met from the optimality of Eq. (11) because $x_i^* = \tilde{x}_i^*$ and t_i^* is still larger than $2\alpha(a_i - 1)$. And for i such that $\tilde{x}_i^* > 1$ (*i.e.*, $x_i^* = 1$), we have

$$\eta_i(x_i^*, s_i^*) = s_i^* + \lambda(1 - a_i)^2 + (s_i^*)^2/(4\lambda) - s_i^* a_i.$$

On the other hand, since $t_i^* < 2\lambda(a_i - 1)$, we have $s_i^* = 2\lambda(a_i - 1)$. Therefore, we have $\eta_i(1, s_i^*) = 0$. And, (b) follows from Lemma 9. □