Vector-Space Markov Random Fields via Exponential Families Appendix

A. Proof of Theorem 1

The proof follows the same lines as the proof in Yang et al. (2014). Let us denote $Q(X)$ as $\log(P(X)/P(0))$. Note that $X = (X_1, X_2, \cdots X_p)$ and each X_r belongs to a vector space. Given any X, let us denote $\bar{X}_s$ as $\bar{X}_s = (X_1, \cdots, X_{s-1}, 0, X_{s+1}, \cdots, X_p)$. Consider the following expansion for $Q(X)$:

$$Q(X) = \sum_{t\in\{1,\cdots,p\}} \mathcal{I}[X_t \neq 0] G_t(X_t) + \cdots + \sum_{\substack{t_1,\cdots t_k \in \\ \{1,\cdots,p\}}} \mathcal{I}[X_{t_1} \neq 0, \dots X_{t_k} \neq 0] G_{t_1\dots t_k}(X_{t_1} \dots X_{t_k}) \quad \text{(A.1)}$$

where $\mathcal{I}$ is the indicator function which takes value 1 if its argument evaluates to true and 0 otherwise.

Using some simple algebra and the definition $Q(X) = \log(P(X)/P(0))$ we can show that

$$\exp(Q(X) - Q(\bar{X}_s)) = \frac{P(X_s|X_1,\cdots,X_{s-1},X_{s+1},\cdots,X_p)}{P(0|X_1,\cdots,X_{s-1},X_{s+1},\cdots,X_p)} \quad \text{(A.2)}$$

From (A.1) we have the following:

$$(Q(X) - Q(\bar{X}_s)) = \mathcal{I}[X_s \neq 0] \Bigg(G_s(X_s) + \sum_{t\in\{1,\cdots,p\}\backslash s} \mathcal{I}[X_t \neq 0] G_{s,t}(X_s, X_t) + \sum_{\substack{t_2,\cdots t_k \in \\ \{1,\cdots,p\}\backslash s}} \mathcal{I}[X_{t_2} \neq 0, \dots X_{t_k} \neq 0] G_{s,t_2\dots t_k}(X_s, \dots X_{t_k}) \Bigg) \quad \text{(A.3)}$$

Since the node conditional distribution follows the exponential family distribution defined in (6) we can show that:

$$\log \frac{P(X_s|X_1,\cdots,X_{s-1},X_{s+1},\cdots,X_p)}{P(0|X_1,\cdots,X_{s-1},X_{s+1},\cdots,X_p)} = \langle E_s(X_{-s}), B_s(X_s) - B_s(0)\rangle + (C_s(X_s) - C_s(0)) \quad \text{(A.4)}$$

Using (A.3) and (A.4) for left and right hand sides of (A.2) and setting $X_t = 0$ for all $t \neq s$ we obtain:

$$\mathcal{I}[X_s \neq 0] G_s(X_s) = \langle E_s(0), B_s(X_s) - B_s(0)\rangle + (C_s(X_s) - C_s(0))$$

Similarly setting $X_r = 0$ for all $r \notin \{s,t\}$ we obtain:

$$\mathcal{I}[X_s \neq 0] G_s(X_s) + \mathcal{I}[X_s \neq 0, X_t \neq 0] G_{s,t}(X_s, X_t) = \langle E_s(0 \cdots X_t \cdots, 0), B_s(X_s) - B_s(0)\rangle + (C_s(X_s) - C_s(0))$$

Similarly, replacing X_s with X_t in (A.2) and setting $X_r = 0$ for all $r \notin \{s,t\}$ we obtain:

$$\mathcal{I}[X_t \neq 0] G_t(X_t) + \mathcal{I}[X_s \neq 0, X_t \neq 0] G_{s,t}(X_s, X_t) = \langle E_t(0 \cdots X_s \cdots, 0), B_t(X_t) - B_t(0)\rangle + (C_t(X_t) - C_t(0))$$

From the above three equations we arrive at the following equality:

$$\langle E_s(0 \cdots X_t \cdots, 0) - E_s(0), B_s(X_s) - B_s(0)\rangle = \langle E_t(0 \cdots X_s \cdots, 0) - E_t(0), B_t(X_t) - B_t(0)\rangle \quad \text{(A.5)}$$

The above equality should hold for the node conditional distributions to be consistent with the joint MRF distribution over X with respect to graph G. So we need to find the form of $E_r()$ that satisfies the above equation. Omitting zero vectors for clarity from (A.5), we get the following:

$$\begin{aligned} \langle E_t(X_s), B_t(X_t)\rangle &= \langle E_s(X_t), B_s(X_s)\rangle \\ \sum_j E_{tj}(X_s) B_{tj}(X_t) &= \sum_l E_{sl}(X_t) B_{sl}(X_s) \end{aligned} \quad \text{(A.6)}$$

We rewrite the natural parameter functions as

$$\begin{aligned} E_{tj}(X_s) &= \sum_l \theta_{sl;tj} B_{sl}(X_s) + \overline{B}_{tj}(X_s) \\ E_{sl}(X_t) &= \sum_j \overline{\theta}_{sl;tj} B_{tj}(X_t) + \overline{B}_{sl}(X_t) \end{aligned} \quad \text{(A.7)}$$

where $\forall j\ \overline{B}_{tj}(X_s)$ are functions in the Hilbert space $\mathcal{H}_s$ orthogonal to the span of functions $B_s(X_s)$, and $\forall j\ \overline{B}_{sl}(X_t)$ are functions in the Hilbert space $\mathcal{H}_t$ orthogonal to the span of functions $B_t(X_t)$; and $\theta_{sl;tj}, \overline{\theta}_{sl;tj}$ are scalars. Combining (A.6) and (A.7), we get

$$\begin{aligned} &\sum_j \sum_l \theta_{sl;tj} B_{sl}(X_s) B_{tj}(X_t) + \sum_j \overline{B}_{tj}(X_s) B_{tj}(X_t) \\ &= \sum_l \sum_j \overline{\theta}_{sl;tj} B_{sl}(X_s) B_{tj}(X_t) + \sum_l \overline{B}_{sl}(X_t) B_{sl}(X_s) \end{aligned} \quad \text{(A.8)}$$

Rearranging terms in the above equation gives us the following equation:

$$\sum_j \left(\sum_l (\theta_{sl;tj} - \overline{\theta}_{sl;tj}) B_{sl}(X_s) + \overline{B}_{tj}(X_s) \right) B_{tj}(X_t) = \sum_l B_{sl}(X_s)\overline{B}_{sl}(X_t) \tag{A.9}$$

However, since $\forall l\ \overline{B}_{sl}(X_t)$ is orthogonal to $B_t(X_t)$, the left and right hand sides of the above equation are equal to 0, which leads us to the following equations.

$$\sum_l B_{sl}(X_s)\overline{B}_{sl}(X_t) = 0$$

$$\sum_j \left(\sum_l (\theta_{sl;tj} - \overline{\theta}_{sl;tj}) B_{sl}(X_s) + \overline{B}_{tj}(X_s) \right) B_{tj}(X_t) = 0 \tag{A.10}$$

However since we assumed that the sufficient statistics are minimal we get $\forall l\ \overline{B}_{sl}(X_t) = 0$ from the first equality and $\forall j, l\ \theta_{sl;tj} = \overline{\theta}_{sl;tj}, \overline{B}_{tj}(X_s) = 0$ from the second equality.

Hence from (A.7), we obtain $E_s(X_t) = \theta_{st}(B_t(X_t) - B_t(0))$ and $E_t(X_s) = \theta_{st}^T(B_s(X_s) - B_s(0))$ where θ_{st} is a matrix formed by the scalars $\theta_{sl;tj}$ such that $(\theta_{st})_{lj} = \theta_{sl;tj}$ and:

$$\begin{aligned} &\mathcal{I}[X_s \neq 0, X_t \neq 0] G_{s,t}(X_s, X_t) = \\ &(B_t(X_t) - B_t(0))^T \theta_{st}^T (B_s(X_s) - B_s(0)) \end{aligned} \tag{A.11}$$

By extending this argument to higher order factors we can show that the natural parameters are required to be in the form specified by (7).

B. Proof of Sparsistency

Before proving the sparsistency result, we will show that the sufficient statistics $B_r(X_r)$ are well behaved. Recall that $B_{ri}(X_r)$ indicates i^{th} component of the vector $B_r(X_r)$. We set the convention that whenever a variable has the subscript $\backslash r$ attached we will be referring to the set of indexes $\{(t, j, k) : \theta_{rj;tk} \in \theta_{r\cdot}, t \neq r\}$.

Proposition 1. *Let $\{X^{(j)}\}_{j=1}^n$ have joint distribution as in* (10)*, then,*

$$P\left(\frac{1}{n} \sum_{j=1}^n \left(B_{ri}\left(X_r^{(j)} \right) \right)^2 \geq \delta \right) \leq \exp\left(-n \frac{\delta^2}{4k_h^2} \right) \tag{B.12}$$

for $\delta \leq \min\left\{ 2\frac{k_v}{3}, k_h + k_v \right\}$.

Proof. It is clear from Taylor Series expansion and assumption 4 that

$$\begin{aligned} &\log E\left[\exp\left(t B_{ri}\left(X_r \right)^2 \right) \right] = \\ &\log \int_{\otimes_{s \in [p]} \mathcal{X}_s} \exp\left\{ t B_{ri}\left(X_r \right)^2 + \right. \\ &\sum_{s \in V} \left\langle B_s\left(X_s \right), \theta_s^* + \sum_{t \in N(r)} \theta_{st}^* B_t\left(X_t \right) \right\rangle + \\ &\left. \sum_{s \in V} C_s\left(X_s \right) - A(\theta^*) \right\} v\left(dx \right) \\ &= \bar{A}_{r,i}\left(\eta; \theta \right)\left(t; \theta^* \right) - \bar{A}_{r,i}\left(\eta; \theta \right)\left(0; \theta^* \right) \\ &\leq t \tfrac{\partial \bar{A}_{r,i}(\eta;\theta)}{\partial \eta}(0) + \tfrac{t^2}{2} \tfrac{\partial^2 \bar{A}_{r,i}(\eta;\theta)}{\partial \eta^2}(ut) \\ &\leq t\, k_v + \tfrac{t^2}{2}\, k_h \end{aligned} \tag{B.13}$$

where $u \in [0, 1]$

Therefore, by the standard Chernoff bounding technique, for $t \leq 1$, it follows that

$$\begin{aligned} &\mathbb{P}\left(\tfrac{1}{n} \textstyle\sum_{j=1}^n B_{ri}\left(X_r^{(j)} \right)^2 \geq \delta \right) \leq \\ &\exp\left(-n\delta t + n\, k_v t + \tfrac{t^2}{2} K_h n \right) \leq \\ &\exp\left(-n \tfrac{\delta^2}{4k_h^2} \right) \end{aligned} \tag{B.14}$$

for $\delta \leq \min\left\{ 2\frac{k_v}{3}, k_h + k_v \right\}$. □

Proposition 2. *Let X be a random vector with the distribution specified in* (10)*. Then, for any positive constant δ and some constant $c > 0$*

$$P\Big(|B_{ri}(X_r)| \geq \delta log(\eta) \Big) \leq c\eta^{-\delta} \tag{B.15}$$

Proof. Let $\bar{v}$ be a unit vector with the same dimensions as $\theta_{r\cdot}^*$ and exactly one non-zero entry, corresponding to the sufficient statistic $B_{ri}(X_r)$. Then we can write $\log\Big(E[\exp(B_{ri}(X_r))] \Big)$ as:

$$\log\Big(E[\exp(B_{ri}(X_r))] \Big) = A(\theta^* + \bar{v}) - A(\theta^*)$$

By Taylor series expansion, for some $u \in [0, 1]$, we can rewrite last equation as

$$\begin{aligned} A(\theta^* + \bar{v}) - A(\theta^*) =& \nabla A(\theta^*).\bar{v} + \frac{1}{2} \bar{v}^T \nabla^2 A(\theta^* + u\,\bar{v}) \bar{v} \\ =& E[B_{ri}(X_r)] \|\bar{v}\|_2 \\ &+ \frac{1}{2} \frac{\partial^2 A(\theta^* + u\,\bar{v})}{\partial \theta_{r_i}^2} \|\bar{v}\|_2^2 \end{aligned}$$

Using Assumption 4 we get the inequality :

$$A(\theta^* + \bar{v}) - A(\theta^*) \leq k_m + \frac{1}{2} k_h$$

Now, by using Chernoff bound, for any positive constant a, we get $P(B_{ri}(X_r) \geq a) \leq \exp(-a + k_m + \frac{1}{2}k_h)$. By setting $a = \delta log(\eta)$ it follows that

$$P(B_{ri}(X_r) \geq \delta log(\eta)) \leq \exp(-\delta log(\eta) + k_m + \frac{1}{2}k_h) \leq c\eta^{-\delta}$$

where $c = \exp(k_m + \frac{1}{2}k_h)$ □

The proof of Sparsistency is based on the primal dual witness proof technique. First note that the optimality condition of (14), can be written as:

$$\nabla \ell(\hat{\theta}_{r\cdot}; \mathcal{D}) + \lambda_1 \sum_{t:r\neq t} \sqrt{\nu_{rt}} \hat{Z}_{1,rt} + \lambda_2 \hat{Z}_2 = 0 \quad \text{(B.16)}$$

where $\hat{Z}_{1,rt} \in \partial \parallel \hat{\theta}_{rt} \parallel_2$, $\hat{Z}_2 \in \partial \parallel \hat{\theta}_{\setminus r} \parallel_1$ and we denote $\hat{Z} = \left(\hat{Z}_1, \hat{Z}_2\right)$, where $\hat{Z}_1 = \{\hat{Z}_{1,rt}\}_{t\in V\setminus r}$. And subgradients $\hat{Z}_1$, $\hat{Z}_2$ should satisfy the following conditions:

$$\begin{aligned} \forall i \;\; (\hat{Z}_2)_i &= sign\left((\hat{\theta}_{r\cdot})_i\right) \;\text{ if }\; (\hat{\theta}_{r\cdot})_i \neq 0 \\ & \quad |(\hat{Z}_2)_i| \leq 1 \;\text{ otherwise} \\ \forall t \;\; \hat{Z}_{1,rt} &= \frac{\hat{\theta}_{rt}}{\|\hat{\theta}_{rt}\|_2} \;\text{ if }\; \hat{\theta}_{rt} \neq 0 \\ & \quad \left\|\hat{Z}_{1,rt}\right\|_2 \leq 1 \;\text{ otherwise} \end{aligned} \quad \text{(B.17)}$$

Note that we can think of $\hat{Z}_1$ and $\hat{Z}_2$ as dual variables by appealing to Lagrangian theory. The next lemma shows that graph structure recovery is guaranteed if the dual is strictly feasible.

Lemma 1. *Suppose that there exists a primal-dual pair* $\left(\hat{\theta}_{r\cdot}, \hat{Z}\right)$ *for* (14) *such that* $\left\|\hat{Z}_{1,S^c}\right\|_{\infty,2} < 1$ *and* $\left\|\hat{Z}_{2,S^c}\right\|_{\infty} < 1$. *Then, any optimal solution* $\tilde{\theta}_{r\cdot}$ *must satisfy* $\left(\tilde{\theta}_{r\cdot}\right)_{S^c} = 0$. *Moreover, if the Hessian sub-matrix* $[\nabla^2 \ell(\hat{\theta}_{r\cdot})]_{SS}$ *is positive definite then* $\hat{\theta}_{\setminus r}$ *is the unique optimal solution.*

Proof. First, note that by Cauchy−Schwarz's and Holder's inequalities

$$\langle \hat{Z}_{1,rt}, \tilde{\theta}_{rt}\rangle \leq \parallel \tilde{\theta}_{rt} \parallel_2 \text{ and} \langle \hat{Z}_2, \tilde{\theta}_{\setminus r}\rangle \leq \parallel \tilde{\theta}_{\setminus r} \parallel_1 . \quad \text{(B.18)}$$

But from (B.16) and the primal optimality of $\hat{\theta}_{r\cdot}$ and $\tilde{\theta}_{r\cdot}$ for (14),

$$\begin{aligned} &\ell\left(\tilde{\theta}_{r\cdot}\right) + \textstyle\sum_{t\neq r} \lambda_1 \sqrt{\nu_{rt}} \langle \hat{Z}_{1,rt}, \tilde{\theta}_{rt}\rangle + \lambda_2 \langle \hat{Z}_2, \tilde{\theta}_{\setminus r}\rangle \\ &\geq \min_{\theta} \ell\left(\theta_{r\cdot}\right) + \textstyle\sum_{t\neq r} \lambda_1 \sqrt{\nu_{rt}} \langle \hat{Z}_{1,rt}, \theta_{rt}\rangle + \lambda_2 \langle \hat{Z}_2, \theta_{\setminus r}\rangle \\ &= \ell\left(\hat{\theta}_{r\cdot}\right) + \textstyle\sum_{t\neq r} \lambda_1 \sqrt{\nu_{rt}} \langle \hat{Z}_{1,rt}, \hat{\theta}_{rt}\rangle + \lambda_2 \langle \hat{Z}_2, \hat{\theta}_{\setminus r}\rangle \\ &= \ell\left(\tilde{\theta}_{r\cdot}\right) + \textstyle\sum_{t\neq r} \lambda_1 \sqrt{\nu_{rt}} \left\|\tilde{\theta}_{rt}\right\|_2 + \left\|\tilde{\theta}_{\setminus r}\right\|_1 \end{aligned} \quad \text{(B.19)}$$

hence, combining with (B.18) with (B.19) it follows that $\sum_{t\neq r} \lambda_1 \sqrt{\nu_{rt}} \left\|\tilde{\theta}_{rt}\right\|_2 + \left\|\tilde{\theta}_{\setminus r}\right\|_1 = \sum_{t\neq r} \lambda_1 \sqrt{\nu_{rt}} \left\|\hat{\theta}_{rt}\right\|_2 + \left\|\hat{\theta}_{\setminus r}\right\|_1 = \sum_{t\neq r} \lambda_1 \sqrt{\nu_{rt}} \langle \hat{Z}_{1,rt}, \hat{\theta}_{rt}\rangle + \lambda_2 \langle \hat{Z}_2, \hat{\theta}_{\setminus r}\rangle$. The result follows.

If the Hessian sub-matrix is positive definite for the restricted problem then the problem is strictly convex and has a unique solution. □

Based on the above lemma, we prove sparsistency theorem by constructing a primal-dual witness $(\hat{\theta}_{r\cdot}, \hat{Z})$ with the following steps:

1. Set $(\hat{\theta}_{r\cdot})_S = argmin_{((\theta_{r\cdot})_S, 0)} \ell\left((\theta_{r\cdot})_S; \mathcal{D}\right) + \lambda_1 \sum_{t\in S} \sqrt{\nu_{rt}} \|\theta_{rt}\|_2 + \lambda_2 \|(\theta_{r\cdot})_S\|_1$
2. For $t \in S$, we define $\hat{Z}_{1,rt} = \frac{\theta_{rt}}{\|\theta_{rt}\|_2}$ and then construct $\hat{Z}_{2,S}$ by the stationary condition.
3. Set $(\hat{\theta}_{r\cdot})_{S^c} = 0$
4. Set $\hat{Z}_{2,S^c}$ such that $\left\|\hat{Z}_{2,S^c}\right\|_\infty < 1$
5. Set $\hat{Z}_{1,S^c}$ such that condition (B.16) is satisfied.
6. The final step consists of showing, that the following conditions are satisfied:
   (a) *strict dual feasibility* : the condition in Lemma 1 holds with high probability
   (b) *correct neighbourhood recovery*: the primal-dual pair specifies the neighbourhood of r, with high probability

We begin by proving some key lemmas that are key to our main theorem. The sub-gradient optimality condition (B.16) can be rewritten as:

$$\nabla \ell(\hat{\theta}_{r\cdot}; \mathcal{D}) - \nabla \ell(\theta^*_{r\cdot}; \mathcal{D}) = W^n - \lambda_1 \sum_{t:r\neq t} \sqrt{\nu_{rt}} \hat{Z}_{1,rt} - \lambda_2 \hat{Z}_2 \quad \text{(B.20)}$$

where $W^n = -\nabla \ell(\theta^*_{r\cdot}; \mathcal{D})$ and $\theta^*_{r\cdot}$ is the true model parameter. By applying mean-value theorem coordinate wise to (B.20), we get:

$$\begin{aligned} \nabla^2 \ell(\theta^*_{r\cdot}; \mathcal{D})[\hat{\theta}_{r\cdot} - \theta^*_{r\cdot}] &= W^n - \lambda_1 \textstyle\sum_{t:r\neq t} \sqrt{\nu_{rt}} \hat{Z}_{1,rt} \\ & \quad - \lambda_2 \hat{Z}_2 + R^n \end{aligned} \quad \text{(B.21)}$$

where R^n is the remainder term after applying mean-value theorem:$R^n_j = [\nabla^2 \ell(\theta^*_{r\cdot}; \mathcal{D}) - \nabla^2 \ell(\bar{\theta}^j_{r\cdot}; \mathcal{D})]^T_j (\hat{\theta}_{r\cdot} - \theta^*_{r\cdot})$ for some $\bar{\theta}^j_{r\cdot}$ on the line between $\hat{\theta}_{r\cdot}$ and $\theta^*_{r\cdot}$, and with $[.]^T_j$ denoting the j-th row of matrix. The following lemma controls the score term W^n

Lemma 2. *Recall $\nu_{r\,max} = \max_t \nu_{rt}$, $\nu_{r\,min} = \min_t \nu_{rt}$, $p' = max(n, p)$. Assume that*

$$\begin{aligned} &\tfrac{8(2-\alpha)}{\alpha} \sqrt{k_1(n,p)\, k_4\, \tfrac{\nu_{r\,max} log(p\nu_{r\,max})}{n\nu_{r\,min}}} \leq \\ &\lambda_1 + \lambda_2 \leq \tfrac{4(2-\alpha)\sqrt{\nu_{r\,max}}}{\alpha\sqrt{\nu_{r\,min}}}\, k_1(n,p)\, k_2(n,p)\, k_4 \end{aligned} \tag{B.22}$$

for some constant $k_4 \leq min\left\{2\frac{k_v}{3}, k_h + k_v\right\}$ and suppose also that $n \geq \frac{8\,k_h^2}{k_4^2}\ log\left(\sum_t m_t\right)$ then,

$$\begin{aligned} &P\left(\|W^n_{\backslash r}\|_{\infty,2} > \tfrac{\alpha}{2-\alpha}\, \tfrac{\sqrt{\nu_{r\,min}}\,(\lambda_1+\lambda_2)}{4}\right) \leq \\ &1 - c_1 p'^{-3}\left(\textstyle\sum_t m_t\right) - \exp\left(-c_2 n\right) - \exp\left(-c_3 n\right) \end{aligned} \tag{B.23}$$

Proof. Define $W^n_t = -\nabla_{\theta_{rt}} \ell\left(\theta^*_{r\cdot}; \mathcal{D}\right)$. Let $W^n_{t,jk}$ be the element in W^n_t corresponding to parameter $\theta_{rj;tk}$. Note that $W^n_{t,jk} = \frac{1}{n}\sum_{t=1}^n V^i_{t,jk}$ where

$$\begin{aligned} V^i_{t,jk} \quad &= B_{rj}\left(X^{(i)}_r\right) B_{tk}\left(X^{(i)}_t\right) - \\ &\nabla_{\theta_{rj;tk}} A_r\left(\theta^*_r + \textstyle\sum_{s\in V\backslash r} \theta^*_{rs} B_s\left(X^{(i)}_s\right)\right) B_{tk}\left(X^{(i)}_t\right) \end{aligned}$$

so for $t' \in \mathbb{R}$

$$\begin{aligned} &E\left[\exp\left(t' V^i_{t,jk}\right) | X^{(i)}_{V\backslash r}\right] = \\ &\int_{X^{(i)}_r} \exp\Bigg\{ t'\Bigg[B_{rj}\left(X^{(i)}_r\right) B_{tk}\left(X^{(i)}_t\right) \\ &-\nabla_{\theta_{rj;tk}} A_r\left(\theta^*_r + \textstyle\sum_{s\in V\backslash r} \theta^*_{rs} B_s\left(X^{(i)}_s\right)\right) B_{tk}\left(X^{(i)}_t\right)\Bigg] \\ &+C\left(X^{(i)}_r\right) + \theta^*_r B_r\left(X^{(i)}_r\right) + \sum_{s\in V\backslash r} B_r\left(X^{(i)}_r\right) \theta^*_{rs} B_s\left(X^{(i)}_s\right) \\ &-A_r\left(\theta^*_r + \sum_{s\in V\backslash r} \theta^*_{rs} B_s\left(X^{(i)}_s\right)\right)\Bigg\} dX_r \\ &= \exp\Bigg\{ A_r\left(\theta^*_r + t' B_{tk}\left(X^{(i)}_t\right) + \sum_{s\in V\backslash r} \theta^*_{rs} B_s\left(X^{(i)}_s\right)\right) \\ &-A_r\left(\theta^*_r + \sum_{s\in V\backslash r} \theta^*_{rs} B_s\left(X^{(i)}_s\right)\right) \\ &-\nabla_{\theta_{rj;tk}} A_r\left(\theta^*_r + \sum_{s\in V\backslash r} \theta^*_{rs} B_s\left(X^{(i)}_s\right)\right) t' B_{tk}\left(X^{(i)}_t\right)\Bigg\} \\ &= \exp\left\{ \frac{\nabla^2_{\theta_{rj;tk},\theta_{rj;tk}} A_r(c)}{2} B_{tk}\left(X^{(i)}_t\right)^2 t'^2 \right\} \end{aligned}$$

where $c = \theta^*_r + \sum_{s\neq r} \theta^*_{rs} B_s\left(X^i_s\right) + v_1 t' B_{tk}\left(X^{(i)}_t\right)$ for some $v_1 \in [0,1]$. Therefore,

$$\begin{aligned} &\tfrac{1}{n}\sum_{i=1}^n \log E\left[\exp\left(t' V^i_{t,jk}\right) | X^{(i)}_{V\backslash r}\right] = \\ &\tfrac{1}{n}\sum_{i=1}^n \frac{1}{2} \nabla^2_{\theta_{rj;tk},\theta_{rj;tk}} A_r(c) B_{tk}\left(X^{(i)}_t\right)^2 t'^2 \end{aligned}$$

Next lets define event $\varepsilon_1 = \left\{ \max_{i,t} \|B_t\left(X^{(i)}_t\right)\|_\infty \leq 4\log p' \right\}$. Then, from Proposition 2 we get $P(\varepsilon_1^c) \leq c_1 n p'^{-4}\left(\sum_t m_t\right) \leq c_1 p'^{-3}\left(\sum_t m_t\right)$. If $t' \leq k_2(n,p)$, Assumption 5 implies that

$$\begin{aligned} &\tfrac{1}{n}\sum_{i=1}^n \log E\left[\exp\left(t' V^i_{t,jk}\right) | X^{(i)}_{V\backslash r}\right] \leq \\ &\tfrac{k_1(n,p)}{2}\, \tfrac{1}{n}\textstyle\sum_{i=1}^n B_{tk}\left(X^{(i)}_t\right)^2 t'^2 \end{aligned}$$

Now, lets define event $\varepsilon_2 = \left\{ \max_{t,j} \frac{1}{n}\sum_{i=1}^n \left(B_{tj}\left(X^{(i)}_t\right)\right)^2 \leq k_4 \right\}$ where $k_4 \leq \min\{2k_v/3, k_h + k_v\}$. Then, by proposition (1) we obtain that if $n \geq \frac{8\,k_h^2}{k_4^2}\log(\sum_{t\in V} m_t)$:

$$P\left(\varepsilon_2^c\right) \leq \exp\left(-n\frac{k_4^2}{4k_h^2} + \log\left(\sum_{t\in V} m_t\right)\right) \leq \exp\left(-n\, c_2\right) \tag{B.24}$$

Therefore, for $t' \leq k_2(n,p)$,

$$\frac{1}{n}\sum_{i=1}^n \log E\left[\exp\left(t' V^i_{t,jk}\right) | X^{(i)}_{V\backslash r}\right] \leq \frac{k_1(n,p) k_4 {t'}^2}{2} \tag{B.25}$$

Hence, by the standard Chernoff bound technique, for $t' \leq k_2(n,p)$

$$\begin{aligned} &P\left(\tfrac{1}{n}\sum_{i=1}^n |V^i_{t,jk}| > \delta \mid \varepsilon_1, \varepsilon_2\right) \leq \\ &2\exp\left(n\left(\tfrac{k_1(n,p) k_4 {t'}^2}{2} - t'\delta\right)\right) \end{aligned} \tag{B.26}$$

Setting $t' = \frac{\delta}{k_1(n,p) k_4}$, for $\delta \leq k_1(n,p)\, k_2(n,p)\, k_4$, we arrive to:

$$P\left(\frac{1}{n}\sum_{i=1}^n |V^i_{t,jk}| > \delta \mid \varepsilon_1, \varepsilon_2\right) \leq 2\exp\left(\frac{-n\delta^2}{2k_1(n,p) k_4}\right) \tag{B.27}$$

Supposing that $\frac{\alpha\sqrt{\nu_{r\,\min}}}{2-\alpha}\, \frac{\lambda_1+\lambda_2}{4\sqrt{m_r\, m_t}} \leq k_1(n,p)\, k_2(n,p)\, k_4$. It then follows that $\delta = \frac{\alpha\sqrt{\nu_{r\,\min}}}{2-\alpha}\, \frac{\lambda_1+\lambda_2}{4\sqrt{m_r\, m_t}}$ satisfies

$$\begin{aligned} &P\left(\tfrac{1}{n}\sum_{i=1}^n |V^i_{t,jk}| > \frac{\alpha\sqrt{\nu_{r\,\min}}}{2-\alpha}\, \frac{\lambda_1+\lambda_2}{4\sqrt{m_r\, m_t}} \mid \varepsilon_1, \varepsilon_2\right) \leq \\ &2\exp\left(\tfrac{-\alpha^2}{(2-\alpha)^2}\, \tfrac{\nu_{r\,\min}\, n\, (\lambda_1+\lambda_2)^2}{32\, k_1(n,p)\, k_4\, m_r\, m_t}\right) \end{aligned} \tag{B.28}$$

Form which, we obtain the following using union bound

$$
\begin{aligned}
&P\left(\|W_t^n\|_2 > \tfrac{\alpha}{2-\alpha}\tfrac{\sqrt{\nu_{r\,\min}}\,(\lambda_1+\lambda_2)}{4}\,|\varepsilon_1,\varepsilon_2\right) \leq \\
&P\left(\|W_t^n\|_\infty > \tfrac{\alpha}{2-\alpha}\tfrac{\sqrt{\nu_{r\,\min}}\,(\lambda_1+\lambda_2)}{4\sqrt{m_r\,m_t}}\,|\varepsilon_1,\varepsilon_2\right) \\
&\leq\; 2\exp\left(\tfrac{-\alpha^2}{(2-\alpha)^2}\tfrac{\nu_{r\,\min}\,n\,(\lambda_1+\lambda_2)^2}{32\,k_1(n,p)k_4\,m_r\,m_t}+\log\left(\nu_{rt}\right)\right)
\end{aligned} \tag{B.29}
$$

and hence,

$$
\begin{aligned}
&P\left(\|W^n\|_{\infty,2} > \tfrac{\alpha}{2-\alpha}\tfrac{\sqrt{\nu_{r\,\min}}\,(\lambda_1+\lambda_2)}{4}\,|\varepsilon_1,\varepsilon_2\right) \leq \\
&2\exp\left(\frac{-\alpha^2}{(2-\alpha)^2}\frac{\nu_{r\,\min}\,n\,\left(\lambda_1+\lambda_2\right)^2}{32\,k_1(n,p)k_4\,\nu_{r\,\max}}+\log\left(\nu_{r\,\max}\right)+\log p\right)
\end{aligned} \tag{B.30}
$$

Finally for $\lambda_1+\lambda_2 \geq \frac{8(2-\alpha)}{\alpha}\sqrt{k_1(n,p)\,k_4\,\frac{\nu_{r\,\max}\log(p\nu_{r\,\max})}{n\nu_{r\,\min}}}$, we obtain

$$
\begin{aligned}
&P\left(\|W^n\|_{\infty,2} > \tfrac{\alpha}{2-\alpha}\tfrac{\sqrt{\nu_{r\,\min}}\,(\lambda_1+\lambda_2)}{4}\right) \leq \\
&c_1 p'^{-3}\left(\textstyle\sum_t m_t\right)+\exp\left(-c_2 n\right)+\exp\left(-c_3 n\right)
\end{aligned} \tag{B.31}
$$

□

Lemma 3. *Suppose that* $\lambda_1+\lambda_2 \leq \frac{C_{min}^2}{40\,logp'\,D_{max}\,d_r\,k_3(n,p)\,\nu_{r\,max}^2}$ *and* $\|W_{\backslash r}^n\|_{\infty,2} \leq \frac{(\lambda_1+\lambda_2)\,\alpha\,\sqrt{\nu_{r\,min}}}{4\,(2-\alpha)}$*, then,*

$$
\begin{aligned}
&P\left(\|(\theta_{r\cdot}^*)_S-(\hat{\theta}_{r\cdot})_S\|_{\infty,2} \leq \tfrac{5\sqrt{\nu_{rmax}}}{C_{min}}\;(\lambda_1+\lambda_2)\right) \\
&\geq 1-cp'^{-3}\left(\textstyle\sum_t m_t\right)
\end{aligned} \tag{B.32}
$$

for some constant $c > 0$.

Proof. We define $F(u_S)$ as:

$$
\begin{aligned}
F(u_S) = \;& \ell\left(\left(\theta_{r\cdot}^*\right)_S+u_S;\mathcal{D}\right)-\ell\left(\left(\theta_{r\cdot}^*\right)_S;\mathcal{D}\right) \\
&+\lambda_1\sum_{t\in N(r)}\sqrt{\nu_{rt}}\left(\|\theta_{rt}^*+u_{rt}\|_2-\|\theta_{rt}^*\|_2\right) \\
&+\lambda_2\left(\|\left(\theta_{r\cdot}^*\right)_S+u_S\|_1-\|\left(\theta_{r\cdot}^*\right)_S\|_1\right)
\end{aligned} \tag{B.33}
$$

From the construction of $\hat{\theta}_{r\cdot}$ it is clear that $\hat{u}_s = (\hat{\theta}_{r\cdot})_S - (\theta_{r\cdot}^*)_S$ minimizes F. And since $F(0) = 0$, we have $F\left(\hat{u}_s\right) \leq 0$. We now show that for some $B > 0$ with $\|u_S\|_{\infty,2} = B$, we have $F(u_S) > 0$. Using this and the fact that F is convex we can then show that $\|\hat{u}_S\|_{\infty,2} \leq B$.

Let u_S an arbitrary vector with $\|u_S\|_{\infty,2} = \frac{5\sqrt{\nu_{rmax}}}{C_{\min}}\left(\lambda_1+\lambda_2\right)$. Then, from the Taylor Series expansion of log likelihood function in F, we have:

$$
\begin{aligned}
F(u_S) = \;& \nabla\ell\left(\left(\theta_{r\cdot}^*\right)_S;\mathcal{D}\right)^T u_S \\
&+u_S^T\nabla^2\ell\left(\left(\theta_{r\cdot}^*\right)_S+v\,u_S\right)\,u_S \\
&+\lambda_1\sum_{t\in N(r)}\sqrt{\nu_{rt}}\left(\|\theta_{rt}^*+u_{rt}\|_2-\|\theta_{rt}^*\|_2\right) \\
&+\lambda_2\left(\|\left(\theta_{r\cdot}^*\right)_S+u_S\|_1-\|\left(\theta_{r\cdot}^*\right)_S\|_1\right)
\end{aligned} \tag{B.34}
$$

for some $v \in [0,1]$

We now bound each of the terms in the right hand side of (B.34). From (B.29) and using Cauchy-Schwarz inequality we obtain:

$$
\begin{aligned}
&\left|\nabla\ell\left(\left(\theta_{r\cdot}^*\right)_S;\mathcal{D}\right)^T u_S\right| \\
&\leq \|\nabla\ell\left(\left(\theta_{r\cdot}^*\right)_S;\mathcal{D}\right)\|_\infty\,\|u_S\|_1 \\
&\leq \|\nabla\ell\left(\left(\theta_{r\cdot}^*\right)_S;\mathcal{D}\right)\|_\infty\,d_r\,\sqrt{\nu_{r\,\max}}\,\|u_S\|_{\infty,2} \\
&\leq \tfrac{\alpha}{2-\alpha}\,\tfrac{\lambda_1+\lambda_2}{4}\,d_r\,\nu_{r\,\max}\,\tfrac{5}{C_{\min}}\,\left(\lambda_1+\lambda_2\right) \\
&= \tfrac{5\,\nu_{r\,\max}}{4\,C_{\min}}\,d_r\,\left(\lambda_1+\lambda_2\right)^2
\end{aligned} \tag{B.35}
$$

where the last inequality holds because $\alpha \in (0,1]$. Moreover, from triangle inequality we have:

$$
\begin{aligned}
&\lambda_1\sum_{t\in S}\sqrt{\nu_{rt}}\left(\|\theta_{rt}^*+u_{rt}\|_2-\|\theta_{rt}^*\|_2\right) \\
&\geq -\lambda_1\sum_{t\in N(r)}\sqrt{\nu_{rt}}\|u_{rt}\|_2 \\
&\geq -\lambda_1\,d_r\,\sqrt{\nu_{r\,\max}}\,\|u_S\|_{\infty,2} \\
&= -\tfrac{5\,d_r\,\nu_{r\,\max}}{C_{\min}}\,\lambda_1\left(\lambda_1+\lambda_2\right)
\end{aligned} \tag{B.36}
$$

Also,

$$
\begin{aligned}
&\lambda_2\,\left(\|\left(\theta_{r\cdot}^*\right)_S+u_S\|_1-\|\left(\theta_{r\cdot}^*\right)_S\|_1\right) \\
&\geq -\lambda_2\,\|u_S\|_1 \\
&\geq -\lambda_2\,\|u_S\|_1 \\
&\geq -\lambda_2\,d_r\,\sqrt{\nu_{r\,\max}}\,\|u_S\|_{\infty,2} \\
&= -\tfrac{5\,\nu_{r\,\max}}{C_{\min}}\,d_r\,\lambda_2\left(\lambda_1+\lambda_2\right)
\end{aligned} \tag{B.37}
$$

On the other hand, by Taylor's approximation of $\nabla^2\ell$, there exists $\alpha_{jk} \in [0,1]$ and $\tilde{u}_{jk}^i$ between $\theta_{\backslash r}$ and $\theta_{\backslash r}+vu_S$ such that

$$
\begin{aligned}
&\Lambda_{\min}\left(\nabla^2\ell\left(\left(\theta_{r\cdot}^*\right)_S+v\,u_S\right)\right) \\
&\geq \min_{\beta\in[0,1]}\Lambda_{\min}\left(\nabla^2\ell\left(\left(\theta_{r\cdot}^*\right)_S+\beta\,u_S\right)\right) \\
&\geq \Lambda_{\min}\left(Q_{SS}^n\right)- \\
&\max_{v\in[0,1]}\max_{\|y\|\leq 1}\left\{\frac{1}{n}\sum_{i=1}^n\sum_{j,k,l,t,h,s,m,s\prime,m\prime}\alpha_{jk}\,v\,\left(\nabla^3 A_r\left(\tilde{u}_{jk}^i\right)\right)_{jkl}\right. \\
&\left. u_{rt;lh}\,B_{th}\left(X_t^i\right)\,y_{s,m,j}\,B_{sm}\left(X_s^i\right)\,B_{s'm'}\left(X_{s'}^i\right)\,y_{s',m',j'}\right\}
\end{aligned} \tag{B.38}
$$

Consider the event ε_1 as defined in the previous proof. We know that $P(\varepsilon_1) \geq 1 - c_1 p'^{-3} \left(\sum_t m_t\right)$. Conditioned on ε_1 and using Assumption 5 we arrive to the following:

$$\begin{aligned}&\Lambda_{\min}\left(\nabla^2 \ell\left((\theta^*_{r\cdot})_S + v\, u_S\right)\right)\\ &\geq C_{\min} - 4\log p' \|u_S\|_1\, D_{\max}\, \nu_{r\max} k_3\,(n,p)\\ &\geq C_{\min} - 4\log p' d_r \sqrt{\nu_{r\max}}\, \|u_S\|_{\infty,2}\, D_{\max}\, \nu_{r\max} k_3\,(n,p)\\ &\geq \tfrac{C_{\min}}{2}\end{aligned} \tag{B.39}$$

where the last inequality holds for $\lambda_1 + \lambda_2 \leq \frac{C^2_{\min}}{40\log p'\, D_{\max}\, d_r\, k_3(n,p)\, \nu^2_{r\max}}$.

Finally using the above bounds we arrive at the following:

$$F\left(u_S\right) \geq d_r\, \nu_{r\max} \frac{5}{C_{\min}}\, (\lambda_1 + \lambda_2)^2 \left(-1 - \frac{1}{4} + \frac{5}{2}\right) > 0 \tag{B.40}$$

Therefore

$$\|(\theta^*_{r\cdot})_S - (\hat{\theta}_{r\cdot})_S\|_{\infty,2} \leq \frac{5\sqrt{\nu_{r\max}}}{C_{\min}}\, (\lambda_1 + \lambda_2)$$

□

Lemma 4. *Suppose that* $\lambda_1 + \lambda_2 \leq \frac{\alpha}{2-\alpha} \frac{\sqrt{\nu_{r\,min}}\, C^2_{min}}{400\, \nu^{\frac{5}{2}}_{r\,max} logp'\, D_{max}\, k_3(n,p)\, d_r}$ *and* $\|W^n_{\setminus r}\|_{\infty,2} \leq \frac{\alpha\,(\lambda_1+\lambda_1)\sqrt{\nu_{r\,min}}}{4\,(2-\alpha)}$*, then,*

$$P\left(\frac{\|R^n\|_{\infty,2}}{\lambda_1 + \lambda_2} \leq \frac{\alpha\sqrt{\nu_{r\,min}}}{4\,(2-\alpha)}\right) \geq 1 - cp'^{-3}\left(\sum_t m_t\right) \tag{B.41}$$

for some constant $c > 0$.

Proof. Recall that $R^n_j = \left[\nabla^2\ell(\theta^*_{r\cdot}; \mathcal{D}) - \nabla^2\ell(\bar{\theta}^j_{r\cdot}; \mathcal{D})\right]^T_j \left(\hat{\theta}_{r\cdot} - \theta^*_{r\cdot}\right)$ where $\left[\ \cdot\ \right]^T_j$ denotes the j-th row of a matrix. Let us also refer to $R^n_{t;jk}$ to the cordinate of R^n corresponding to $\theta_{rj;tk}$. Then,

$$R^n_{t;jk} = \left[\frac{1}{n}\sum_{i=1}^n B_{tk}(X^i_t)\left[\nabla^2 A_r\left(\theta^*_r + \sum_{s\neq r}\theta^*_{rs}B_s(X^i_s)\right) - \nabla^2 A_r\left(\bar{\theta}^{t_{jk}}_r + \textstyle\sum_{s\neq r}\bar{\theta}^{t_{jk}}_{rs}B_s(X^i_s)\right)\right]_j \otimes B^i_{r\cdot}\right]^T \left(\hat{\theta}_{r\cdot} - \theta^*_{r\cdot}\right) \tag{B.42}$$

with $B^i_{r\cdot}$ the vector of sufficient statistics evluate at the i-th sample. Intoroducing the notation $\langle\theta_{r\cdot}, B^i_{r\cdot}\rangle =: \theta_r + \sum_{s\neq r}\theta_{rs}B_s\left(X^i_s\right)$, from the mean value theorem we obtain

$$\begin{aligned}&\nabla^2 A_{r;jl}\left(\langle\theta^*_{r\cdot}, B^i_{r\cdot}\rangle\right) - \nabla^2 A_{r;jl}\left(\langle\bar{\theta}^{t_{jk}}_{r\cdot}, B^i_{r\cdot}\rangle\right) =\\ &-v^i_{j,l}\left[\langle\hat{\theta}_{r\cdot} - \theta^*_{r\cdot}, B^i_{r\cdot}\rangle\right]\left(\nabla^3 A_r\right)_{jl:}\left(\langle\bar{\bar{\theta}}^{i;t_{jk}}_{r\cdot}, B^i_{r\cdot}\rangle\right)\end{aligned} \tag{B.43}$$

Therefore, combining (B.42) with (B.43) and using basic properties of krockner product we ontain that

$$\begin{aligned}\left|R^n_{t;jk}\right| &\leq \frac{1}{n}\sum_{i=1}^n\left|B_{tk}(X^i_t)\right|\nu_{r\max}k_3\,(n,p)\ D_{max}\,\|\hat{\theta}_{r\cdot} - \theta^*_{r\cdot}\|^2_2\\ &\leq\ 4\log p'\,\nu_{r\max}k_3\,(n,p)\ D_{max}\,\|\hat{\theta}_{r\cdot} - \theta^*_{r\cdot}\|^2_2\end{aligned} \tag{B.44}$$

which implies

$$\begin{aligned}&\|R^n_t\|_{\infty,2} \leq\\ &4\sqrt{\nu_{r\max}}\log p'\,\nu_{r\max}k_3\,(n,p)\ D_{max}\,\|\hat{\theta}_{r\cdot} - \theta^*_{r\cdot}\|^2_2 \leq\\ &4\sqrt{\nu_{r\max}}\log p'\,\nu_{r\max}k_3\,(n,p)\ D_{max}\,\tfrac{25\,d_r\,\nu_{r\max}}{C^2_{\min}}\,\lambda^2_1 \leq\\ &\tfrac{\lambda_1+\lambda_2}{4}\,\tfrac{\alpha\sqrt{\nu_{r\min}}}{2-\alpha}\end{aligned} \tag{B.45}$$

with probbility at least $1 - cp'^{-3}\left(\sum_t m_t\right)$. □

We now prove theorem 2 using lemmas 2-4. Recalling that $Q^n = \nabla^2\ell(\theta^*_{r\cdot}; \mathcal{D})$ and the fact that we have set $(\hat{\theta}_{r\cdot})_{S^c} = 0$ in our primal-dual construction, we can rewrite condition (B.21) as the following equations:

$$\begin{aligned}&Q^n_{S^cS}[(\hat{\theta}_{r\cdot})_S - (\theta^*_{r\cdot})_S] =\\ &W^n_{S^c} - \lambda_1\textstyle\sum_{t\notin N(r)}\sqrt{\nu_{rt}}\hat{Z}_{1,rt} - \lambda_2\hat{Z}_{2,S^c} + R^n_{S^c}\end{aligned} \tag{B.46}$$

$$\begin{aligned}&Q^n_{SS}[(\hat{\theta}_{r\cdot})_S - (\theta^*_{r\cdot})_S] =\\ &W^n_S - \lambda_1\textstyle\sum_{t\in N(r)}\sqrt{\nu_{rt}}\hat{Z}_{1,rt} - \lambda_2\hat{Z}_{2,S} + R^n_S\end{aligned} \tag{B.47}$$

Since the matrix Q^n_{SS} is invertible, the conditions (B.46) and (B.47) can be rewritten as :

$$\begin{aligned}&Q^n_{S^cS}(Q^n_{SS})^{-1}[W^n_S - \lambda_1\sum_{t\in N(r)}\sqrt{\nu_{rt}}\hat{Z}_{1,rt} - \lambda_2\hat{Z}_{2,S} + R^n_S] =\\ &W^n_{S^c} - \lambda_1\textstyle\sum_{t\notin N(r)}\sqrt{\nu_{rt}}\hat{Z}_{1,rt} - \lambda_2\hat{Z}_{2,S^c} + R^n_{S^c}\end{aligned} \tag{B.48}$$

Rearranging yields the following condition:

$$\begin{aligned}&\lambda_1\textstyle\sum_{t\notin N(r)}\sqrt{\nu_{rt}}\hat{Z}_{1,rt} =\\ &W^n_{S^c} + R^n_{S^c} - Q^n_{S^cS}(Q^n_{SS})^{-1}[W^n_S + R^n_S] -\\ &\lambda_2\hat{Z}_{2,S^c} + Q^n_{S^cS}(Q^n_{SS})^{-1}[\lambda_1\textstyle\sum_{t\in N(r)}\sqrt{\nu_{rt}}\hat{Z}_{1,rt} + \lambda_2\hat{Z}_{2,S}]\end{aligned} \tag{B.49}$$

Strict Dual Feasibility: we now show that for the dual sub-vector $\hat{Z}_{1,S^c}$, we have $\left\|\hat{Z}_{1,S^c}\right\|_{\infty,2} < 1$. We get the following equation from B.49, by applying triangle inequality:

$$\lambda_1 \sqrt{\nu_{r\,\min}} \left\| \hat{Z}_{1,S^c} \right\|_{\infty,2} \leq$$
$$\lambda_1 \left\| \sum_{t \notin N(r)} \sqrt{\nu_{rt}} \hat{Z}_{1,rt} \right\|_{\infty,2} \leq$$
$$\left[\|W^n\|_{\infty,2} + \|R^n\|_{\infty,2} \right] \left(1 + \left\| Q^n_{S^cS} (Q^n_{SS})^{-1} \right\|_{\infty,2} \sqrt{d_r} \right)$$
$$+ \lambda_2 \sqrt{\nu_{r\,\max}}$$
$$+ \left\| Q^n_{S^cS} (Q^n_{SS})^{-1} \right\|_{\infty,2} \left[(\lambda_1 + \lambda_2) \sqrt{d_r\, \nu_{r\,\max}} \right] \tag{B.50}$$

where $\nu_{r\,\min} = \min_t \nu_{rt}$, $\nu_{r\,\max} = \max_t \nu_{rt}$ and $d_r = |N(r)|$ Using mutual incoherence bound 2 on the above equation gives us:

$$\left\| \hat{Z}_{1,S^c} \right\|_{\infty,2} \leq$$
$$\frac{1}{\lambda_1 \sqrt{\nu_{r\,\min}}} \left[\|W^n\|_{\infty,2} + \|R^n\|_{\infty,2} \right] (2 - \alpha)$$
$$+ \frac{\lambda_2 \sqrt{\nu_{r\,\max}}}{\lambda_1 \sqrt{\nu_{r\,\min}}} \left[1 + \left\| Q^n_{S^cS} (Q^n_{SS})^{-1} \right\|_{\infty,2} \sqrt{d_r} \left(\frac{\lambda_1}{\lambda_2} + 1 \right) \right] \tag{B.51}$$

Using the previous lemmas we obtain the following:

$$\begin{aligned} &\left\| \hat{Z}_{1,S^c} \right\|_{\infty,2} \leq \frac{1}{2\lambda_1} \left[\alpha \left(\lambda_1 + \lambda_2 \right) \right] \\ &+ \frac{\lambda_2 \sqrt{\nu_{r\,\max}}}{\lambda_1 \sqrt{\nu_{r\,\min}}} \left[1 + \frac{m_{min}}{m_{max}} (1 - \alpha) \left(\frac{\lambda_1}{\lambda_2} + 1 \right) \right] \end{aligned} \tag{B.52}$$

If $\lambda_2 < \left(\frac{\alpha}{2 - \alpha + 2 \frac{\sqrt{\nu_{r\,\max}}}{\sqrt{\nu_{r\,\min}}}} \right) \lambda_1$, then,

$$\left\| \hat{Z}_{1,S^c} \right\|_{\infty,2} \quad < \quad 1 \tag{B.53}$$

We have shown that the dual is strictly feasible with high probability and also the solution is unique. And hence based on Lemma 1 the method correctly excludes all edges not in the set of edges.

Correct Neighbourhood Recovery: To show that all correct neighbours are recovered, it suffices to show that

$$\|(\theta^*_{r\cdot})_S - (\hat{\theta}_{r\cdot})_S\|_{\infty,2} \leq \frac{\theta_{min}}{2}$$

where $\theta_{min} = \min_{t \in V \setminus r} \|\theta_{rt}\|_2$.

Using Lemma 3 we can show the above inequality holds if $\theta_{min} \geq \frac{10\, \sqrt{\nu_{r\,max}}}{C_{\min}} (\lambda_1 + \lambda_2)$

C. Full MyFitnessPal Graph

Figure 1 shows a high-level view of the entire VS-MRF learned from the MyFitnessPal food database. The three macro-nutrients (fat, carbs, and protein) correspond to the three largest hubs with the remaining nine micro-nutrients representing smaller hubs.

References

Yang, Eunho, Baker, Yulia, Ravikumar, Pradeep, Allen, Genevera, and Liu, Zhandong. Mixed graphical models via exponential families. *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, pp. 1042–1050, 2014.

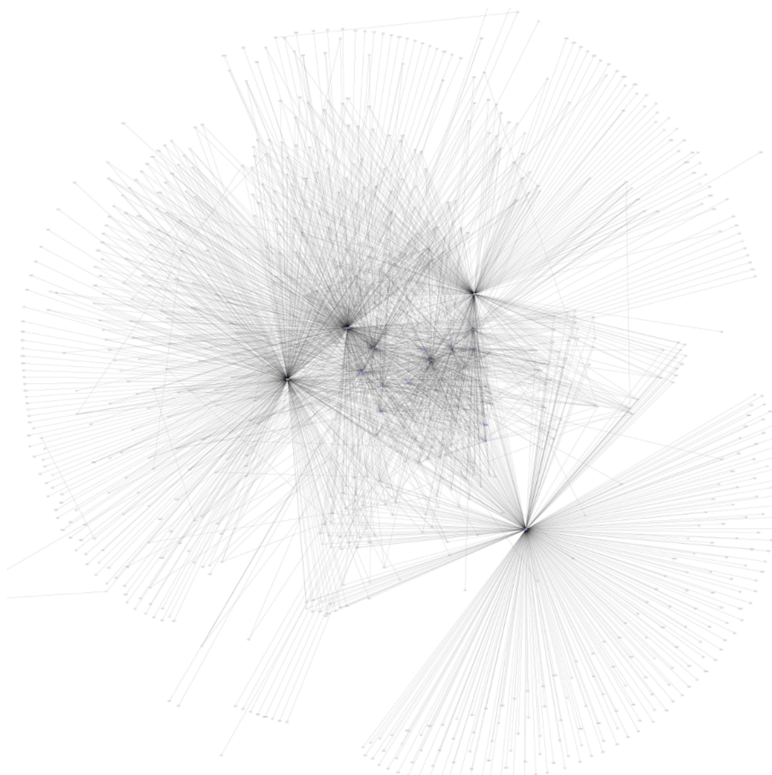

Figure 1. Full MRF learned from the MyFitnessPal food database. The hubs correspond to point-inflated gamma nutrient nodes, with the three largest hubs being the macro-nutrients (fat, carbs, and protein).