A Technical Proofs

Proof of Theorem 2 Let $i' := \text{argmax}_i |Z_i|$. The attribute configuration $\lambda_{i'}$ corresponding to node i' appears at least B times in the set $\{\lambda_1, \ldots, \lambda_n\}$. By the pigeon-hole principle any partition of the set $\{1, \ldots, n\}$ which contains less than B sets must have a set which contains two nodes i and j with attribute configuration $\lambda_i = \lambda_j = \lambda_{i'}$. Therefore, the number of sets in the partition must be at least B. Our partitioning scheme produces exactly B sets, and is therefore optimal.

Proof of Theorem By definition,

$$A_{i,j} = \sum_{1 \leq k,l \leq B} A_{i,j}^{(k,l)}. \tag{14}$$

To prove the theorem, we first show that $A_{i,j} = A'^{|Z_i|,|Z_j|}_{\lambda_i,\lambda_j}$. This is straightforward from definition (9), since $D_1, \ldots, D_B$ is a partition of nodes and thus $i \in D_k$, $j \in D_l$ for only $k = |Z_i|$ and $l = |Z_j|$. This also implies

$$\begin{aligned}\mathbb{P}\left(A_{ij} = 1 \mid \tilde{\Theta}, \lambda_1, \ldots, \lambda_n\right) &= A'^{|Z_i|,|Z_j|}_{\lambda_i,\lambda_j} \\ &= P_{\lambda_i,\lambda_j} = Q_{i,j},\end{aligned}$$

using (8).

To prove independence, we show that if $(i,j) \neq (i',j')$, then $(\lambda_i, \lambda_j) \neq (\lambda_{i'}, \lambda_{j'})$ or $(|Z_i|, |Z_j|) \neq (|Z'_i|, |Z'_j|)$. Since we already showed $A_{i,j} = A'^{|Z_i|,|Z_j|}_{\lambda_i,\lambda_j}$, this implies independence of $A_{i,j}$ to other entries in A.

Now, suppose $(\lambda_i, \lambda_j) = (\lambda_{i'}, \lambda_{j'})$, since if it does not hold there is nothing to prove. Because $(i,j) \neq (i',j')$ by assumption, at least one of $i \neq i'$ or $j \neq j'$ is true. Without loss of generality, suppose $i \neq i'$. Then, since $\lambda_i = \lambda_{i'}$, $|Z_i| \neq |Z_{i'}|$ from definition of Z_i.

B Chernoff Bound of Poisson Distribution

Theorem 5 *Let X be the random variable which is distributed as Poisson distribution of parameter λ. Then,*

$$P(X \geq x) \leq \frac{e^{-\lambda}(e\lambda)^x}{x^x}. \tag{15}$$

C Upper bound of size of the partition when $n > 2^d$

As a binomial distribution with finite mean in limit, Y_c is approximately distributed as a Poisson distribution of parameter $\frac{n}{2^d}$. To use Chernoff bound (15) with $\lambda = \frac{n}{2^d}$, $x = 2^{t+1}\log_2(n)$, $t = d'' - d$, we first bound each term:

$$e^{-\lambda} = e^{-\frac{n}{2^d}} \leq e^{-2^t}, \tag{16}$$

$$e^x = e^{2^{t+1}\log_2(n)} = n^{2^{t+1}}, \tag{17}$$

$$\lambda^x = \left(\frac{n}{2^d}\right)^x \leq (2^{t+1})^{2^{t+1}\log_2(n)}, \tag{18}$$

$$x^x = (2^{t+1}\log_2(n))^{2^{t+1}\log_2(n)}. \tag{19}$$

Then, plugging these into (15),

$$\mathbb{P}\left(B = \max Y_c > 2^{t+1}\log_2(n)\right) \leq \sum_{c=1}^{n} \mathbb{P}\left(Y_c > 2^{t+1}\log_2(n)\right) \tag{20}$$

$$\leq n \cdot \frac{e^{-2^t} \cdot n^{2^{t+1}} \cdot (2^{t+1})^{2^{t+1}\log_2(n)}}{(2^{t+1}\log_2(n))^{2^{t+1}\log_2(n)}}. \tag{21}$$

By taking log,

$$\begin{aligned}\log \mathbb{P}\left(B = \max Y_c > 2^{t+1}\log_2(n)\right) &\leq -2^t + 2^{t+1}\log_2(n) \\ &+ (2^{t+1}\log_2(n))\log_2 2^{t+1} - 2^{t+1}\log_2(n)\log_2(2^{t+1}\log_2(n)) \\ &= -2^t + 2^{t+1}\log_2(n)(1 + \log_2 2^{t+1} - \log_2(2^{t+1}\log n)),\end{aligned} \tag{22}$$

and this goes to $-\infty$ as $n \to \infty$. Therefore, $\mathbb{P}[B = \max Y_c > 2^{t+1}\log_2(n)] \to 0$.