A Further Details on Definitions of Entropy, Mutual Information and Kullback-Leibler Divergence

Definition 6 *For a pair of discrete random variables (X, Y), the empirical mutual information is defined as:*

$$I(X,Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X,Y) \tag{44}$$

where $\mathcal{H}(X)$ and $\mathcal{H}(Y)$ are the emperical marginal entropies, and $\mathcal{H}(X,Y)$ is the joint entropy of X and Y defined as follows:

$$\mathcal{H}(X) = -\sum_{i=1}^{k} \frac{m_i^x}{S} \log_2(\frac{m_i^x}{S}) \tag{45}$$

$$\mathcal{H}(Y) = -\sum_{i=1}^{l} \frac{m_j^y}{S} \log_2(\frac{m_j^y}{S}) \tag{46}$$

$$\mathcal{H}(X,Y) = -\sum_{i=1}^{k}\sum_{j=1}^{l} \frac{m_{ij}}{S} \log_2(\frac{m_{ij}}{S}) \tag{47}$$

Definition 7 (Kullback-Leibler Divergence) *Given two probability distributions $p_{k\times l}$ and $q_{k\times l}$*

$$D_{KL}(p||q) = \sum_{i=1}^{k}\sum_{j=1}^{l} p_{ij} \log(\frac{p_{ij}}{q_{ij}}) = \mathcal{H}(p,q) - \mathcal{H}(p) \tag{48}$$

where $\mathcal{H}(p) = \sum_{i=1}^{k}\sum_{j=1}^{l} p_{ij} \log p_{ij}$ is the entropy of p, and $\mathcal{H}(p,q) = \sum_{i=1}^{k}\sum_{j=1}^{l} p_{ij} \log q_{ij}$ is the cross entropy of p and q.

B Dirichlet Distribution

Consider the Dirichlet distribution with parameters $\alpha_1, \cdots, \alpha_K > 0$. The probability density function is represented as

$$f(x_1, x_2, \cdots, x_K) = \frac{1}{\beta(\boldsymbol{\alpha})} \prod_{k=1}^{K} x_k^{\alpha_k - 1} \tag{49}$$

where $\sum_{k=1}^{K} x_k = 1$. Moreover, β is the Beta function:

$$\begin{aligned} & \beta(\boldsymbol{\alpha}) \\ = \; & \beta(\alpha_1, \alpha_2, ..., \alpha_K) && (50) \\ = \; & \frac{\prod_{k=1}^{K} \Gamma(\alpha_i)}{\Gamma(\sum_{k=1}^{K} \alpha_i)} && (51) \end{aligned}$$

and $\Gamma(z)$ is the Gamma function:

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx, \text{ if real part of } z > 0 \tag{52}$$

When z is a natural number, we have

$$\Gamma(i) = (i-1)! \quad i \in \mathbb{N} \tag{53}$$

C Proof of Lemma 1

From (13) and (15), we have

$$\begin{aligned} & |\log\left(\frac{A_{X,Y}}{B_{X,Y}}\right) - S.I(X;Y)| \\ = \; & |\log\left(\frac{\beta(\mathbf{m}+\mathbf{1})\beta(\mathbf{1}^x)\beta(\mathbf{1}^y)}{\beta(\mathbf{m}^x+\mathbf{1}^x)\beta(\mathbf{m}^y+\mathbf{1}^y)\beta(\mathbf{1})}\right) - S.I(X;Y)| && (54) \\ = \; & |\sum_{i=1}^{k}\sum_{j=1}^{l} m_{ij}\log\frac{m_{ij}}{e} - \sum_{i=1}^{k} m_i^x \log\frac{m_i^x}{e} \\ & - \sum_{j=1}^{l} m_j^y \log\frac{m_j^y}{e} + S\log\frac{S}{e} - S.I(X;Y) \\ & + \log\left(\frac{t\sqrt{S}S^{k+l-1-kl}\prod_{i=1}^{k}\prod_{j=1}^{l} t_{ij}\sqrt{m_{ij}}}{\prod_{i=1}^{k} t_i^x\sqrt{m_i^x}\prod_{j=1}^{l} t_j^y\sqrt{m_j^y}}\right) \\ & + \log\left(\frac{\beta(\mathbf{1}^x)\beta(\mathbf{1}^y)}{\beta(\mathbf{1})}\right)| && (55) \end{aligned}$$

Note that from Sterling's formula, for any natural number n we have $\sqrt{2\pi n}\frac{n}{e}^n \leq n! \leq e\sqrt{n}\frac{n}{e}^n$. We define $t_{ij} = \frac{m_{ij}!}{\sqrt{m_{ij}}\left(\frac{m_{ij}}{e}\right)^{m_{ij}}}$, $t_i^x = \frac{m_i^x!}{\sqrt{m_i^x}\left(\frac{m_i^x}{e}\right)^{m_i^x}}$ and $t_j^y = \frac{m_j^y!}{\sqrt{m_j^y}\left(\frac{m_j^y}{e}\right)^{m_j^y}}$ and therefore, from Sterling's formula we have

$$\sqrt{2\pi} \leq t_{ij}, t_i^x, t_j^y \leq e \tag{56}$$

Therefore, we have

$$\begin{aligned} & |\log\left(\frac{A_{X,Y}}{B_{X,Y}}\right) - S.I(X;Y)| \\ = \; & |\log\left(\frac{t\sqrt{S}S^{k+l-1-kl}\prod_{i=1}^{k}\prod_{j=1}^{l} t_{ij}\sqrt{m_{ij}}}{\prod_{i=1}^{k} t_i^x\sqrt{m_i^x}\prod_{j=1}^{l} t_j^y\sqrt{m_j^y}}\right) \\ & + \log\left(\frac{\beta(\mathbf{1}^x)\beta(\mathbf{1}^y)}{\beta(\mathbf{1})}\right)| && (57) \\ \leq \; & O(kl!) + |k+l-\frac{1}{2}-kl|\log S && (58) \end{aligned}$$

(57) is true from Definition 2 as $\sum_{i=1}^{k}\sum_{j=1}^{l} m_{ij}\log\frac{m_{ij}}{S} - \sum_{i=1}^{k} m_i^x\log\frac{m_i^x}{S} - \sum_{j=1}^{l} m_j^y\log\frac{m_j^y}{S} - S.I(X;Y) = 0$ and as $\sum_{i=1}^{k}\sum_{j=1}^{l} m_{ij}$, $\sum_{i=1}^{k} m_i^x$ and $\sum_{j=1}^{l} m_j^y$ are equal to S. Finally, (58) follows from $\sqrt{2\pi} \leq t_{ij}, t_i^x, t_j^y \leq e$.

D Proof of Lemma 2

We first prove the statement about the true positive rate, and then we prove the statement about the complexity. Under Dirichlet assumption, the chance of

dependent pairs falling in the same bucket is:

$$
\begin{aligned}
&TP_1(G)\\
=&P\Big(\text{Collision}(X,Y) \mid (X,Y)\sim\mathbb{P},\mathbb{P}\sim Dir(\boldsymbol{\alpha})\Big) && (59)\\
=&\sum_{v\in V_{buckets}(G)} P\Big(\text{Collision}(X,Y,v) \mid (X,Y)\sim\mathbb{P},\\
&\mathbb{P}\sim Dir(\boldsymbol{\alpha})\Big) && (60)\\
=&\sum_{v\in V_{buckets}(G)} P\Big(\text{Prefix}(S_x(v),\pi(X))\\
&,\text{Prefix}(S_y(v),\pi(Y))\\
&\mid (X,Y)\sim\mathbb{P},\mathbb{P}\sim Dir(\boldsymbol{\alpha})\Big) && (61)\\
=&\sum_{v\in V_{buckets}(G)} A(v) && (62)\\
=&\mathbb{A}(G) && (63)
\end{aligned}
$$

where $\text{Collision}(X,Y)$ is the event that X and Y fall in the same bucket, $\text{Collision}(X,Y,v)$ is the event X and Y fall in buckets v, and $\text{Prefix}(S_1,S_2)$ is the event string S_1 is a prefix of string S_2.

Equation (62) holds, as:

$$
\begin{aligned}
&P\Big(\text{Prefix}(S_x(v),\pi(X)),\ \text{Prefix}(S_y(v),\pi(Y)) \mid\\
&(X,Y)\sim\mathbb{P},\mathbb{P}\sim Dir(\boldsymbol{\alpha})\Big) && (64)\\
=&P\Big(X_{\pi^{-1}(1)}=S_x(v,1),Y_{\pi^{-1}(1)}=S_y(v,1),\\
&X_{\pi^{-1}(2)}=S_x(v,2),Y_{\pi^{-1}(2)}=S_y(v,2),\\
&\dots \mid\\
&(X,Y)\sim\mathbb{P},\mathbb{P}\sim Dir(\boldsymbol{\alpha})\Big) && (65)\\
=&\prod_{s=1}^{\text{depth}(v)} P\Big(X_{\pi^{-1}(s)}=S_x(v,s)\\
&,Y_{\pi^{-1}(s)}=S_y(v,s)\Big) && (66)\\
=&\int_{\mathbb{P}}\prod_{i=1}^{k}\prod_{j=1}^{l} p_{ij}^{v.n_{ij}}.\frac{1}{\beta(\mathbf{1})}\\
&\times\prod_{i=1}^{k}\prod_{j=1}^{l} p_{ij}^{\alpha_{ij}}\,dp_{11}\dots dp_{kl} && (67)\\
=&\frac{\beta(v.\mathbf{n}+\mathbf{1})}{\beta(\mathbf{1})}=A(v) && (68)
\end{aligned}
$$

Therefore (25) holds. Now, we show that (24) holds. Building prefix trees on S_x (or S_y) has runtime of:

$$c_{\text{prefix}}.|V_{buckets}(G)|.depth(G) \tag{69}$$

Moreover, to search each data points against all the buckets using the prefix tree takes at most $c_{\text{search}}.depth(G)$ time, and the overall complexity of searching all the points in b band is:

$$c_{\text{search}}.b.(M+N).depth(G) \tag{70}$$

The expected value of the number of positive calls in each band under Dirichlet model is:

$$
\begin{aligned}
&E\Big(\text{ Number of positive calls in a single band}\Big)\\
=&N.M.P\Big(\text{Collision}(X,Y) \mid X\sim\mathbb{P}_x,Y\sim\mathbb{P}_y,\\
&\mathbb{P}_x\sim Dir(\boldsymbol{\alpha}_y),\mathbb{P}_y\sim Dir(\boldsymbol{\alpha}_x)\Big) && (71)\\
=&N.M.\sum_{v\in V_{buckets}(G)} P\Big(\text{Collision}(X,Y,v) \mid X\sim\mathbb{P}_x,\\
&Y\sim\mathbb{P}_y,\mathbb{P}_x\sim Dir(\boldsymbol{\alpha}_x),\mathbb{P}_y\sim Dir(\boldsymbol{\alpha}_y)\Big) && (72)\\
=&N.M.\sum_{v\in V_{buckets}(G)} P\Big(\text{Prefix}(S_x(v),\pi(X)),\\
&\text{Prefix}(S_y(v),\pi(Y)) \mid X\sim\mathbb{P}_x,Y\sim\mathbb{P}_y,\\
&\mathbb{P}_x\sim Dir(\boldsymbol{\alpha}_x),\mathbb{P}_y\sim Dir(\boldsymbol{\alpha}_y)\Big) && (73)\\
=&N.M.\sum_{v\in V_{buckets}(G)} B_x(v)B_y(v) && (74)\\
=&N.M.\sum_{v\in V_{buckets}(G)} B(v)=N.M.\mathbb{B}(G) && (75)
\end{aligned}
$$

Where in (74), we used:

$$
\begin{aligned}
&P(\text{Prefix}(S_x(v),\pi(X)) \mid X\sim\mathbb{P}_x,\mathbb{P}_x\sim Dir(\boldsymbol{\alpha}_x))\\
=&B_x(v) && (76)\\
&P(\text{Prefix}(S_y(v),\pi(Y)) \mid Y\sim\mathbb{P}_y,\mathbb{P}_y\sim Dir(\boldsymbol{\alpha}_y))\\
=&B_y(v) && (77)
\end{aligned}
$$

and proofs of (76) and (77) are similar to $((64)-(68))$. Then, to insert data points into matching buckets, the complexity is c_{insert} times the expected number of insertions needed. Using Dirichlet prior, we have

$$
\begin{aligned}
&E\Big(\text{Number of insertions}\Big) && (78)\\
=\ &E\Big(\text{Number of } X\_\text{ insertions}\Big)\\
&+E\Big(\text{Number of } Y\_\text{ insertions}\Big) && (79)\\
=\ &N\sum_{v\in V_{buckets}(G)} P\Big(\text{Prefix}(S_x(v),\pi(X)) \mid X\sim\mathbb{P}_x\\
&,\mathbb{P}_x\sim Dir(\boldsymbol{\alpha}_x)\Big) && (80)\\
+\ &M\sum_{v\in V_{buckets}(G)} P\Big(\text{Prefix}(S_y(v),\pi(Y)) \mid Y\sim\mathbb{P}_y,\\
&\mathbb{P}_y\sim Dir(\boldsymbol{\alpha}_y)\Big) && (81)\\
=\ &N\sum_{v\in V_{buckets}(G)} B_x(v)+M\sum_{v\in V_{buckets}(G)} B_y(v) && (82)\\
=\ &N.\mathbb{B}_x(G)+M.\mathbb{B}_y(G) && (83)
\end{aligned}
$$

Therefore, (24) holds. This completes the proof of Lemma 2.

E Proof of Lemma 3

Similar to Proof of Lemma 2, we derive $P(\text{Prefix}(S_x(v), \pi(X)) \mid X \sim \mathbb{P}_x, \mathbb{P}_x \sim F_x)$ for bounded density distribution F_x as follows

$$P(\text{Prefix}(S_x(v), \pi(X)) \mid X \sim \mathbb{P}_x, \mathbb{P}_x \sim F_x) \tag{84}$$

$$= P\Big(X_{\pi^{-1}(1)} = S_x(v,1), \quad X_{\pi^{-1}(2)} = S_x(v,2), \ldots, X_{\pi^{-1}(depth(v))} = S_x(v, depth(v)) \mid X \sim \mathbb{P}_x, \mathbb{P}_x \sim F_x\Big) \tag{85}$$

$$= \prod_{s=1}^{\text{depth}(v)} P\Big(X_{\pi^{-1}(s)} = S_x(v,s) \mid X \sim \mathbb{P}_x, \mathbb{P}_x \sim F_x\Big) \tag{86}$$

$$\leq \int_{\mathbb{P}} \prod_{i=1}^{k} \prod_{j=1}^{l} p_{ij}^{v.n_{ij}} . f_{\max} dp_{11} \ldots dp_{kl} \tag{87}$$

$$= f_{\max} \beta(v.\mathbf{n} + \mathbf{1}) = f_{\max} \beta(\mathbf{1}) A(v) \tag{88}$$

$P(\text{Prefix}(S_y(v), \pi(Y)) \mid Y \sim \mathbb{P}_y, \mathbb{P}_y \sim F_y)$ and $P\Big(\text{Collision}(X,Y) \mid X \sim \mathbb{P}_x, Y \sim \mathbb{P}_y, \mathbb{P}_x \sim F_x, \mathbb{P}_y \sim F_y\Big)$ are bounded similar to (88). This completes the proof of Lemma 3.

F Intuitions Behind the Selection of Decision Trees

Now, we consider the problem of maximizing the true positive rate while minimizing the complexity of a decision tree. In the case of a single band ($b = 1$), the expected true positive can be computed using (24). However, it is difficult to compute it in the case of multiple bands ($b > 1$):

$$TP_b(G) = P\Big(\text{Collision}(X,Y) \text{ in at least one of } b \text{ bands} \mid (X,Y) \sim P, P \sim Dir(\boldsymbol{\alpha})\Big) \tag{89}$$

as this involves b'th order integration. Therefore, instead of solving the optimization problem

$$\arg\max \quad TP_b(G) \tag{90}$$

$$\text{s.t.} \quad E(\text{Complexity}_b(G)) \leq C \tag{91}$$

we solve the easier optimization problem:

$$\arg\min_{\text{G}} \frac{E(\text{Complexity}_1(G))}{TP_1(G)} \tag{92}$$

The intuition behind optimizing this heuristic criteria is that we need roughly $O(\frac{1}{TP_1(G)})$ bands to push up the true positive rate from $TP_1(G)$ to nearly one. Therefore, the complexity is roughly equal to

$$\frac{E(\text{Complexity}_1(G))}{TP_1(G)} \tag{93}$$

Note that, while this heuristic criteria based on Dirichlet prior results in sub-optimal decision trees, in Appendix E, we proved that these decision trees provide sub-quadratic complexity for arbitrary priors with bounded density distribution.

Intuitively, to minimize (92) one needs to accept nodes v as buckets if $\frac{A(v)}{B(v)}$, $\frac{A(v)}{B_x(v)}$, and $\frac{A(v)}{B_y(v)}$ are large. In Algorithm 2, we presented an approach for designing decision trees based on the following constraints:

$$\begin{cases} \frac{A(v)}{B(v)} \geq (\log N)^{k+l-2kl} N^{1+\delta-\eta} : & \text{accept bucket} \\ \frac{A(v)}{B_x(v)} \leq (\log N)^{k-kl} N^{1-\eta} : & \text{prune} \\ \frac{A(v)}{B_y(v)} \leq (\log N)^{l-kl} N^{\delta-\eta} : & \text{prune} \\ depth(v) > \theta \log(N) : & \text{prune} \\ otherwise : & \text{branch into the } kl \text{ children} \end{cases} \tag{94}$$

The algorithm recursively constructs functions $A_v : V \to \mathbb{R}$, $B_v^x : V \to \mathbb{R}$, $B_v^y : V \to \mathbb{R}$, and $B_v : V \to \mathbb{R}$ representing the conditional probabilities defined in (29) - (33). We designate a node v as a bucket if the ratio of conditional probabilities $\frac{A(v)}{B(v)}$ surpass some threshold $(\log N)^{k+l-2kl} N^{1+\delta-\eta}$, and we prune a node v when $\frac{A(v)}{B_x(v)}$ and $\frac{A(v)}{B_y(v)}$ are less than $(\log N)^{k-kl} N^{1-\eta}$ and $(\log N)^{l-kl} N^{\delta-\eta}$, respectively. In Theorem 1 we show that these trees are highly sensitive, and have low complexity.

Remark 6 *C, C^x, and C^y are constants defined in* (146), (140), *and* (141).

Remark 7 *Each node v in the tree has a $k \times l$ matrix $v.\mathbf{n}$ which stores frequencies of occurrences of each pair of characters, and a sequence $(S_x, S_y) \in \mathcal{A}^{depth(v)} \times \mathcal{B}^{depth(v)}$ which stores the path from root to each node. We start constructing the tree by calling a DFS function on the root. The time complexity of this algorithm is equal to $O(size(tree))$. In lemma 7, we show that the size of the tree and the complexity of tree construction is $O(N^{\eta} \log(N))$. We also, show that mapping data points to the tree and checking positives are also efficient.*

G Proof of Lemma 4

To prove $\lambda(X,Y) \leq 1+\delta$, note that $\epsilon(X,Y,1+\delta) = 0$ by setting $n_{ij} = n.\frac{m_{ij}}{S}$ and $n = \min\big(\theta\log(N), \frac{\delta\log(N)}{\mathcal{H}(\frac{\mathbf{m}}{S})-\mathcal{H}(\frac{\mathbf{m}^x}{S})}, \frac{\log(N)}{\mathcal{H}(\frac{\mathbf{m}}{S})-\mathcal{H}(\frac{\mathbf{m}^y}{S})}\big)$. To prove $\max(1,\delta) \leq \lambda(X,Y)$, note that $\max\big(\mathcal{H}(\frac{n_i}{n}), \mathcal{H}(\frac{n_j}{n})\big) \leq \mathcal{H}(\frac{n_{ij}}{n})$. Therefore, using (40), $\epsilon(X,Y,\eta) < \infty$ only when $\max(1,\delta) \leq \eta$. ■

H Proof of Theorem 1

H.1 Proof of Theorem 1

Define n_{ij}'s as the minimizers of $\epsilon(X,Y,\eta)$ in (40), and

$$\begin{aligned} & V_{n_{11},\ldots,n_{kl}}(G^\eta) \\ = \; & \Big\{v \in V_{buckets}(G^\eta) \mid v.n_{ij} = n_{ij}\Big\} \end{aligned} \tag{95}$$

where $v.n_{ij}$ is defined in (36). Then we have:

$$|V_{n_{11},\ldots,n_{kl}}(G^\eta)| = \binom{n}{n_{11},\ldots,n_{kl}} \tag{96}$$

Lemma 5 *Consider node v in $V_{n_{11},\ldots,n_{kl}}$, where $\{n_{11},\ldots,n_{kl}\}$ satisfy the conditions in Definition 8. Then assuming $\alpha_{ij} = 1$, node v does not get pruned, and node v (or one of its ancestors) get accepted in Algorithm 2.*

Proof of Lemma 5 is relegated to Appendix L.1. Using Lemma 5, we know that for any node $v \in V_{n_{11},\ldots,n_{kl}}(G^\eta)$, v has an ancestor in $V_{buckets}(G^\eta)$. Therefore we can derive a lower-bound on true positive rate as:

$$\begin{aligned} & TP_1^{X,Y}(G^\eta) \\ = \; & \sum_{v\in V_{buckets}(G^\eta)} P_\pi\Big(\text{Prefix}(v.S_x,\pi(X)), \\ & \quad \text{Prefix}(v.S_y,\pi(Y))\Big) \end{aligned} \tag{97}$$

$$\geq \sum_{v\in V_{n_{11},\cdots,n_{kl}}(G^\eta)} \frac{\prod_{i=1}^k\prod_{j=1}^l \frac{m_{ij}!}{(m_{ij}-n_{ij})!}}{\frac{S!}{(S-n)!}} \tag{98}$$

$$= \sum_{v\in V_{n_{11},\cdots,n_{kl}}(G^\eta)} \frac{(S-n)!\prod_{i=1}^k\prod_{j=1}^l m_{ij}!}{S!\prod_{i=1}^k\prod_{j=1}^l (m_{ij}-n_{ij})!} \tag{99}$$

$$\geq \binom{n}{n_{11},\cdots,n_{kl}} \frac{(S-n)!\prod_{i=1}^k\prod_{j=1}^l m_{ij}!}{S!\prod_{i=1}^k\prod_{j=1}^l (m_{ij}-n_{ij})!} \tag{100}$$

$$= \frac{(S-n)!n!\prod_{i=1}^k\prod_{j=1}^l m_{ij}!}{S!\prod_{i=1}^k\prod_{j=1}^l (m_{ij}-n_{ij})!n_{ij}!} \tag{101}$$

$$\geq \frac{(\frac{S-n}{e})^{S-n}\sqrt{2\pi(S-n)}(\frac{n}{e})^n\sqrt{2\pi n}}{(\frac{S}{e})^S e\sqrt{S}\prod_{i=1}^k\prod_{j=1}^l(\frac{m_{ij}-n_{ij}}{e})^{m_{ij}-n_{ij}}} \tag{102}$$

$$\geq \frac{(\sqrt{2\pi})^{kl+2}}{e^{2kl+1}}.e^{n.\sum(\frac{n_{ij}}{n})\log(\frac{n_{ij}}{n})}.e^{-n\sum(\frac{n_{ij}}{n}).\log(\frac{m_{ij}}{s})} e^{(S-n)\sum\frac{m_{ij}-n_{ij}}{S-n}\log(\frac{m_{ij}-n_{ij}}{S-n})} e^{-(S-n)\sum\frac{m_{ij}}{S}\log(\frac{m_{ij}-n_{ij}}{S-n})}\sqrt{1-\frac{n}{S}}.n^{\frac{kl-1}{2}} \tag{103}$$

$$\geq N^{-\epsilon(X,Y,\eta)}.(\frac{\log(N).\log(kl)}{\eta-1})^{\frac{kl-1}{2}}.\sqrt{\frac{1}{2}} \tag{104}$$

$$= C.N^{-\epsilon(X,Y,\eta)}.\log(N)^{\frac{kl-1}{2}} \tag{105}$$

where we have used Stirling's formula (129) for factorials above, and we also assumed $\theta\log(N) \leq \frac{S}{2}$.

So far we have proved the statement about sensitivity, and next we are going to prove the complexity statement. From Lemma 6 presented below, we have $\sum_{v\in V_l(G^\eta)} A(v) = 1$.

Lemma 6 *For any decision tree G, $\sum_{v\in V_{leaf nodes}(G)} A(v) = 1$.*

For proof of Lemma 6, see Appendix L.2. Therefore, we conclude that

$$\begin{aligned} \mathbb{A}(G^\eta) &= \sum_{v\in V_{buckets}(G^\eta)} A(v) && (106)\\ &\leq \sum_{v\in V_l(G^\eta)} A(v) && (107)\\ &= 1 && (108)\end{aligned}$$

Using the prune/accept rules in Algorithm 2 we have:

$$\frac{\mathbb{A}(G^\eta)}{\mathbb{B}(G^\eta)} = \frac{\sum_{v\in V_{Buckets}(G^\eta)} A(v)}{\sum_{v\in V_{Buckets}(G^\eta)} B(v)} \geq (\log N)^{k+l-2kl} N^{1+\delta-\eta} \tag{109}$$

$$\frac{\mathbb{A}(G^\eta)}{\mathbb{B}_x(G^\eta)} = \frac{\sum_{v\in V_{Buckets}(G^\eta)} A(v)}{\sum_{v\in V_{Buckets}(G^\eta)} B_x(v)} \geq (\log N)^{k-kl} N^{1-\eta} \tag{110}$$

$$\frac{\mathbb{A}(G^\eta)}{\mathbb{B}_y(G^\eta)} = \frac{\sum_{v\in V_{Buckets}(G^\eta)} A(v)}{\sum_{v\in V_{Buckets}(G^\eta)} B_y(v)} \geq (\log N)^{l-kl} N^{\delta-\eta} \tag{111}$$

Therefore, from (109)-(111) and Lemma 7 presented

below, we conclude that

$$\begin{aligned} & E(\text{Complexity}_1(G^\eta)) \\ = \; & c_{\text{prefix}}.|V_{buckets}(G^\eta)|\text{depth}(G^\eta) \\ & +c_{\text{search}}.(N+M)\text{depth}(G^\eta) \\ & +c_{\text{insert}}.\big(N\mathbb{B}_x(G^\eta)+M\mathbb{B}_y(G^\eta)\big) \\ & +c_{\text{check}}.NM.B(G^\eta) \qquad (112) \\ = \; & O(N^\eta(\log N)^2) \\ & +N.O(N^{\eta-1}\log(N)^{k-kl}) \\ & +N^\delta.O(N^{\eta-\delta}\log(N)^{l-kl}) \\ & +N^{1+\delta}.O(N^{\eta-1-\delta}\log(N)^{k+l-kl}) \\ = \; & O(N^\eta(\log N)^2) \qquad (113) \end{aligned}$$

where using Lemma 7 below, $V_{buckets}(G^\eta)$ has size $O(N^\eta \log(N))$. This completes the proof of Theorem 1. ■

Lemma 7 *Number of nodes in the tree constructed in Algorithm 2 is at most* $O(N^\eta \log(N))$.

For proof of Lemma 7, see Appendix L.3.

I Further Theoretical Guarantees

Here, we present Algorithm 3, which is based on Algorithm 1 and 2, and given a value $\lambda > 0$ it finds (nearly) all λ-associated pairs with complexity $O(N^\lambda \log(N)^2)$. The algorithm is based on constructing decision trees with parameter η for various values of η, $0 < \eta < \lambda$. Then multiple bands from each tree is recruited, in a way that all λ-associated pairs are discovered, while the complexity remains $O(N^\lambda \log(N)^2)$

Definition 8 *For* $\mathbf{H} = \{\eta_0, ..., \eta_z\}$, (X, Y) *is called* $(\lambda, \mathbf{H})$*-associated if*

$$\min_{\eta \in \mathbf{H}} \epsilon(X, Y, \eta) + \eta \leq \lambda \qquad (114)$$

Algorithm 3 Decision Trees with Various Parameters

Input: λ, $\mathbf{H} = \{\eta_0, ..., \eta_z\}$, Data points $\mathcal{X} = \{X_1, X_2, ..., X_N\}$, Queries $\mathcal{Y} = \{Y_1, Y_2, ..., Y_M\}$
Output: $(\lambda, \mathbf{H})$-Associated Pairs

for $\eta \in \mathbf{H}$ **do**
  $b_\eta \leftarrow \frac{5}{C_0}(\log N)^{\frac{kl-1}{2}} N^{\lambda-\eta}$ ▷ See Theorem 8
  $G^\eta \leftarrow$ MAKE_BUCKETS(η, N, M) ▷ Using Algorithm 2
  MATCH_TO_BUCKETS$(G^\eta, b_\eta, \mathcal{X}, \mathcal{Y})$ ▷ Using Algorithm 1

Obviously, any λ-associated pair is also $(\lambda, \mathbf{H})$-associated for any set $\mathbf{H}$.

Lemma 8 *Consider parameter* $\lambda \geq 0$ *and* $\mathbf{H} = \{\eta_0, ..., \eta_z\}$ *where* $0 \leq \eta_0 < ... < \eta_z \leq 1$. *Then Algorithm 3 with parameters* λ *and* $\mathbf{H}$ *finds nearly all* $(\lambda, \mathbf{H})$*-associated pairs. Moreover, the complexity of Algorithm 3 under Dirichlet prior is* $O(N^\lambda \log(N)^2 z)$.

Lemma 8 proves that it is possible to find nearly all λ-associated pairs using Algorithm 3 with the complexity $O(N^\lambda \log(N)^2)$.

I.1 Proof of Lemma 8

Assume a $(\lambda, \mathbf{H})$-associated data pair (X, Y). The true positive rate over all the bands is :

$$\begin{aligned} & TP^{X,Y} \\ = \; & 1 - \prod_{\eta \in \mathbf{H}} \left(1 - TP_1^{X,Y}(G^\eta)\right)^{b_\eta} \qquad (115) \\ \geq \; & 1 - \left(1 - 5(\log N)^{\frac{1-kl}{2}} N^{-\epsilon(X,Y,\eta_{min})}\right)^{b_\eta} \qquad (116) \\ \geq \; & 1 - e^{-5} \approx 99.4\% \qquad (117) \end{aligned}$$

where in (115) we used Theorem 1, and in (116) we used:

$$\eta_{min} + \epsilon(X, Y, \eta_{min}) = \lambda_{\mathbf{H}}(X, Y) \leq \lambda \qquad (118)$$

where

$$\eta_{min} = \arg\min_{\eta \in \mathbf{H}} \epsilon(X, Y, \eta) + \eta \qquad (119)$$

Note that, (117) follows from the fact that $(1 - \frac{a}{x})^x < \frac{1}{e^a}$ for any positive a and x where $x \geq a$. Therefore, for b bands, complexity is:

$$\begin{aligned} & E(\text{Complexity}_b(G)) \\ = \; & c_{\text{prefix}}|V_{buckets}|\text{depth}(G) \\ & +c_{\text{search}}b(N+M)\text{depth}(G) \\ & +c_{\text{insert}}.\big(b.N\mathbb{B}_x(G)+b.M\mathbb{B}_y(G)\big) \\ & +c_{\text{check}}.b.NM.\mathbb{B}(G) \qquad (120) \\ = \; & O(N^\eta(\log N)^2) \\ & +O(N^{\lambda-\eta}).N.O(N^{\eta-1}\log(N)^{k-kl}) \\ & +O(N^{\lambda-\eta}).N^\delta.O(N^{\eta-\delta}\log(N)^{l-kl}) \\ & +O(N^{\lambda-\eta}).N^{1+\delta}.O(N^{\eta-1-\delta}\log(N)^{k+l-2kl}) \\ = \; & O(N^\eta(\log N)^2) \qquad (121) \end{aligned}$$

where in (121) we used $b = O(N^{\lambda-\eta})$. ■

J Finding Pairs with High Mutual Information

Lemma 9 below clarify the relationship between the mutual information and λ-associativity, and Theorem 2 provides guarantees on the performance of Algorithm 3 based on mutual information. First we need to define $\bar{\lambda}(X, Y)$.

Definition 9 *For any pairs of data points $X \in \mathcal{A}^S$ and $Y \in \mathcal{B}^S$, define $\bar{\lambda}(X,Y)$ as follows*[3]*:*

$$\bar{\lambda}(X,Y) = 1 + \delta - I(X,Y)\min(\frac{1}{\mathcal{H}(Y)}, \frac{\delta}{\mathcal{H}(X)}, \theta) \quad (122)$$

Lemma 9 $\lambda(X,Y) \leq \bar{\lambda}(X,Y)$.

J.1 Proof of Lemma 9

It is sufficient to show that

$$\epsilon(X,Y,\bar{\lambda}(X,Y)) = 0 \quad (123)$$

It can be shown that (40) holds by selecting $n = \frac{1+\delta-\bar{\lambda}(X,Y)}{I(X,Y)}.\log(N)$ and $n_{ij} = \lfloor \frac{n.m_{ij}}{S} \rfloor$. Also note that $n \leq \theta.\log(N)$. ■

J.2 Proof of Theorem 2

Using Lemma 9, $\lambda(X,Y) \leq \bar{\lambda}(X,Y)$. This makes Theorem 2 a special case of Lemma 8.

K Further Experiments

K.1 Further Discussion on Experiment 1

Similar to Experiment 1, G^η is computed using Algorithm 2, and true positives are calculated using algorithm 1 where $b = 1$.

K.2 Experiment 3

In this experiment, we run Algorithm 3 for $\lambda = 1.1, 1.3, 1.5$ and $\mathbf{H} = \{\eta_1, \eta_2, ..., \eta_z\}$, where $\eta_i = \frac{i\lambda}{z}$, $1 \leq i \leq z$, and $z = 100$. We sampled data points X and Y similar to the previous experiment and computed :

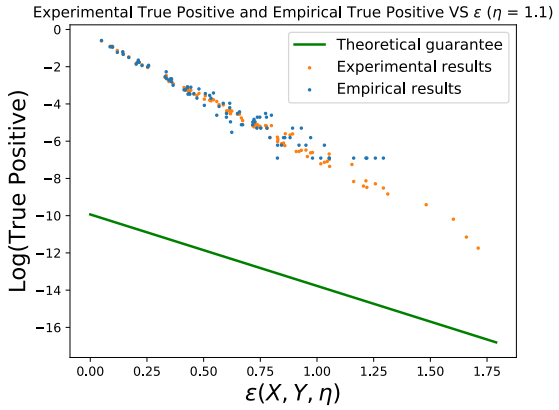
$$\lambda_{\mathbf{H}}(X,Y) = \min_{\eta \in \mathbf{H}} \eta + \epsilon(X,Y,\eta) \quad (124)$$

To compute the expected true positive rates we inserted the true positive rates computed on single bands from (43) into (115). Figure 8 shows expected and theoretical lower-bounds on true positive rates versus $\lambda(X,Y)$ for $\lambda = 1.3$. The theoretical lower-bounds are smaller than the expected results, as the bounds used for deriving theoretical guarantees are not tight.

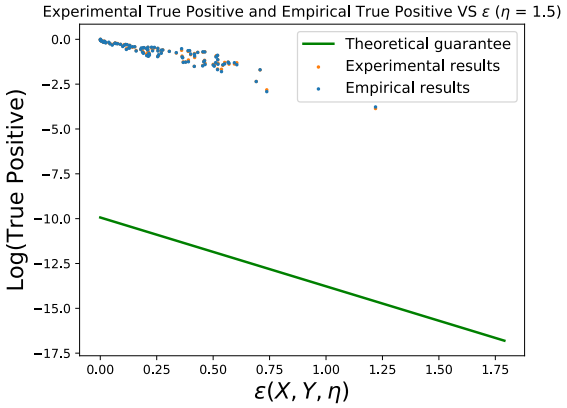
Remark 8 *When computing the true positive rates, Algorithm 3 suggest*

$$b_\eta = \frac{5}{C}(\log(N))^{\frac{kl-1}{2}} N^{\lambda-\eta} \quad (125)$$

[3]Note that ϵ and λ also depend on N, δ, and θ. However, for the sake of simplicity, we assume that N, δ, and θ are constants.



(a)



(b)

Figure 7: Experimental and empirical and theoretical lower-bounds on true positives rates $TP_1^{X,Y}(G^\eta)$ for various data points $(X,Y) \sim P$, $P \sim Dir(\boldsymbol{\alpha})$ for $\eta \in \{1.1, 1.5\}$.

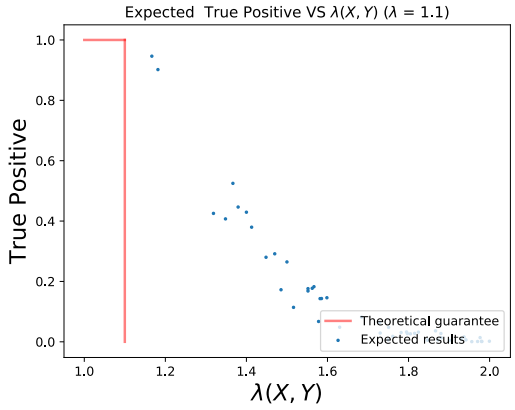
where $\frac{5}{C} \approx 5000$. Instead of this value, in experiment 2 we used:

$$b_\eta = \frac{1}{20}(\log(N))^{\frac{kl-1}{2}} N^{\lambda-\eta} \quad (126)$$

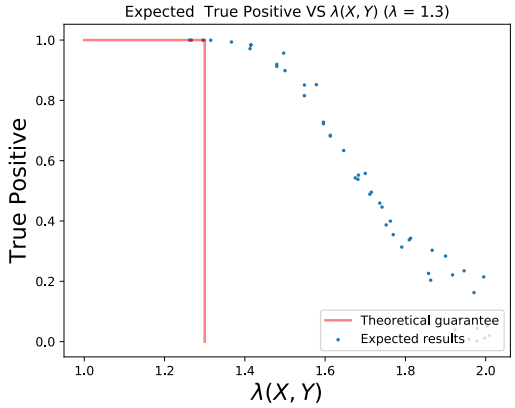
Generally, the lower-bounds by Theorem 1 and 8, and the parameter settings suggested are pessimistic , and more practical parameters can be tuned in the theorems.

K.3 Experiment 4

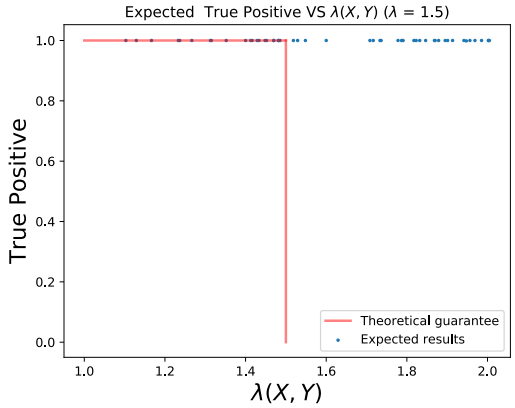
In this experiment, we verify Lemma 9 by plotting $\lambda(X,Y)$ versus $\bar{\lambda}(X,Y)$ for 100 randomly sampled data points (X,Y) from Dirichlet prior. We also plot $\lambda(X,Y)$ versus $I(X,Y)$. The results show that $\lambda(X,Y) \leq \bar{\lambda}(X,Y)$ as expected by Lemma 9, and $\lambda(X,Y) \leq 2 - I(X,Y)$ for the binary data as required by Theorem 2.



(a)

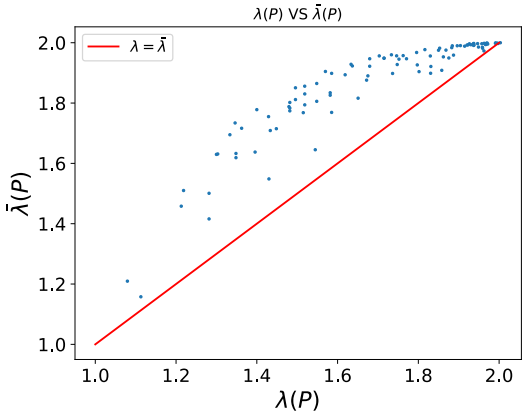


(b)

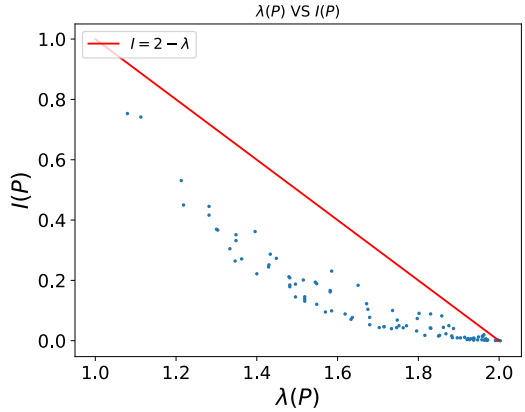


(c)

Figure 8: Expected and theoretical true positives $TP_1^{X,Y}(G^\eta)$ for various data points $(X, Y) \sim P$, $P \sim Dir(\boldsymbol{\alpha})$, and $\lambda \in \{1.1, 1.3, 1.5\}$ using Algorithm 3.



(a)



(b)

Figure 9: In this figure, $\lambda(\mathbb{P})$ versus $\bar{\lambda}(\mathbb{P})$ and $I(\mathbb{P})$ for distributions $\mathbb{P}$ sampled from Dirichlet are sketched. This figure confirms that $\lambda(X, Y) \leq \bar{\lambda}(X, Y)$ and $\lambda(X, Y) \leq 2 - I(X, Y)$.

L Proof of Lemmas 5, 6 and 7

L.1 Proof of Lemma 5

First we show that node v (or any of its ancestors) do not get pruned by algorithm 2:

$$\begin{aligned}
&\frac{A_v}{B_v^x} \\
= \;& \frac{\beta(v.\mathbf{n}+\boldsymbol{\alpha})\beta(\boldsymbol{\alpha}^x)}{\beta(\boldsymbol{\alpha})\beta(v.\mathbf{n}^x+\boldsymbol{\alpha}^x)} && (127)\\
= \;& \frac{\prod_{i=1}^k\prod_{j=1}^l(n_{ij}+\alpha_{ij}-1)!\prod_{i=1}^k(\alpha_i^x-1)!}{\prod_{i=1}^k\prod_{j=1}^l(\alpha_{ij}-1)!\prod_{i=1}^k(n_i^x+\alpha_i^x-1)!} \\
\geq \;& \frac{\prod_{i=1}^k(\alpha_i^x-1)!}{\prod_{i=1}^k\prod_{j=1}^l(\alpha_{ij}-1)!} \\
& \times\frac{\prod_{i=1}^k\prod_{j=1}^l(\frac{n_{ij}+\alpha_{ij}-1}{e})^{n_{ij}+\alpha_{ij}-1}}{\prod_{i=1}^k(\frac{n_i+\alpha_i^x-1}{e})^{n_i^x+\alpha_i^x-1}} \\
& \frac{\cdot\sqrt{2\pi(n_{ij})}}{e.\sqrt{(n_i^x)}} && (128)
\end{aligned}$$

where we used (29) and (31) in (127), and Sterling's inequality in (128):

$$\sqrt{2\pi n}(\frac{n}{e})^n \leq n! \leq e\sqrt{n}(\frac{n}{e})^n \quad (129)$$

Therefore, we have

$$\begin{aligned}
&\log(\frac{A_v}{B_v^x}) \\
\geq \;& C_\alpha^1+\sum_{i=1}^k\sum_{j=1}^l(n_{ij}+\alpha_{ij}-1)\log(n_{ij}+\alpha_{ij}-\frac{1}{2}) \\
& -\sum_{i=1}^k(n_i+\boldsymbol{\alpha}_i^x-1)\log(n_i+\alpha_i^x-\frac{1}{2}) && (130)\\
= \;& C_\alpha^1+\sum_{i=1}^k\sum_{j=1}^l n_{ij}\log(n_{ij})+n_{ij}\log(1+\frac{\alpha_{ij}-\frac{1}{2}}{n_{ij}}) \\
& +(\alpha_{ij}-1)\log(n_{ij})+(\alpha_{ij}-1)\log(1+\frac{\alpha_{ij}-\frac{1}{2}}{n_{ij}}) \\
& -\sum_{i=1}^k n_i\log(n_i)+n_i\log(1+\frac{\alpha_i^x-\frac{1}{2}}{n_i}) && (131)\\
& -(\alpha_i^x-1)\log(n_i)-(\alpha_i^x-1)\log(1+\frac{\alpha_i^x-\frac{1}{2}}{n_i}) \\
= \;& C_\alpha^2+n(\mathcal{H}(\frac{n_i}{n})-\mathcal{H}(\frac{n_{ij}}{n})) \\
& +\sum_{i=1}^k\sum_{j=1}^l n_{ij}\log(\frac{n_in_{ij}+n_i(\alpha_{ij}-\frac{1}{2})}{n_in_{ij}+n_{ij}(\alpha_i-\frac{1}{2})}) \\
& +(\alpha_{ij}-1)\log(n_{ij})-\sum_{i=1}^k(\alpha_i^x-1)\log(n_i) && (132)\\
= \;& C_\alpha^2+n(\mathcal{H}(\frac{n_i}{n})-\mathcal{H}(\frac{n_{ij}}{n})) \\
& +\sum_{i=1}^k\sum_{j=1}^l-\frac{n_{ij}^2(\alpha_i-\frac{1}{2})-n_in_{ij}(\alpha_{ij}-\frac{1}{2})}{n_in_{ij}+n_i(\alpha_{ij}-\frac{1}{2})} \\
& +(\alpha_{ij}-1)\log(n_{ij})-\sum_{i=1}^k(\alpha_i^x-1)\log(n_i) && (133)\\
= \;& C_\alpha^2+n(\mathcal{H}(\frac{n_i}{n})-\mathcal{H}(\frac{n_{ij}}{n})) \\
& +\sum_{i=1}^k\sum_{j=1}^l\frac{(\alpha_{ij}-\frac{1}{2})}{1+\frac{(\alpha_{ij}-\frac{1}{2})}{n_{ij}}}-\frac{(\alpha_i-\frac{1}{2})}{\frac{n_i}{n_{ij}}+\frac{n_i(\alpha_{ij}-\frac{1}{2})}{n_{ij}^2}} \\
& +(\alpha_{ij}-1)\log(n_{ij})-\sum_{i=1}^k(\alpha_i^x-1)\log(n_i) && (134)\\
\geq \;& C_\alpha^2+n(\mathcal{H}(\frac{n_i}{n})-\mathcal{H}(\frac{n_{ij}}{n})) \\
& +\sum_{i=1}^k\sum_{j=1}^l\frac{(\alpha_{ij}-\frac{1}{2})}{1+(\alpha_{ij}-\frac{1}{2})}-\frac{(\alpha_i-\frac{1}{2})}{\frac{n_i}{n_{ij}}+\frac{n_i(\alpha_{ij}-\frac{1}{2})}{n_{ij}^2}} \\
& +(\alpha_{ij}-1)\log(n_{ij})-\sum_{i=1}^k(\alpha_i^x-1)\log(n_i) && (135)\\
\geq \;& C_\alpha^3+n(\mathcal{H}(\frac{n_i}{n})-\mathcal{H}(\frac{n_{ij}}{n})) \\
& +\sum_{i=1}^k\sum_{j=1}^l\frac{(\alpha_{ij}-\frac{1}{2})}{1+(\alpha_{ij}-\frac{1}{2})}-\frac{(\alpha_i-\frac{1}{2})}{1+\frac{(\alpha_{ij}-\frac{1}{2})}{n_{ij}}} \\
& +(\alpha_{ij}-1)\log(n_{ij})-\sum_{i=1}^k(\alpha_i^x-1)\log(n_i) && (136)\\
\geq \;& C_\alpha^3+n(\mathcal{H}(\frac{n_i}{n})-\mathcal{H}(\frac{n_{ij}}{n})) \\
& +\sum_{i=1}^k\sum_{j=1}^l\frac{(\alpha_{ij}-\frac{1}{2})}{1+(\alpha_{ij}-\frac{1}{2})}-(\alpha_i-\frac{1}{2}) \\
& +(\alpha_{ij}-1)\log(n_{ij})-\sum_{i=1}^k(\alpha_i^x-1)\log(n_i) && (137)\\
= \;& C_\alpha^3+n(\mathcal{H}(\frac{n_i}{n})-\mathcal{H}(\frac{n_{ij}}{n})) \\
& +\sum_{i=1}^k\sum_{j=1}^l(\alpha_{ij}-1)\log(n_{ij}) \\
& -\sum_{i=1}^k(\alpha_i^x-1)\log(n_i) && (138)
\end{aligned}$$

where C_α^1, C_α^2, and C_α^3 are constants depending only on $\boldsymbol{\alpha}$ and δ.

Now assuming $\alpha_{ij}=1$ and using (40),

$$-\sum_{i=1}^k(\alpha_i^x-1)\log(n_i)\geq-(kl-k)\log(\theta\log(N)) \quad (139)$$

Therefore:

$$\frac{A_v}{B_v^x} \;\geq\; C^xN^{1-\eta}\log(N)^{-(kl-k)} \quad (140)$$

where C^x is a constant depending only on k, l, δ, and θ.

Similarly,

$$\frac{A_v}{B_v^y} \geq C^y N^{\delta-\eta}\log(N)^{-(kl-l)} \quad (141)$$

Therefore, we conclude that for any $v \in V_{n_{11},\ldots,n_{kl}}(G^\eta)$, the node or its ancestors do not get pruned by the first two pruning conditions. They do not get pruned by the third condition either, as $\text{depth}(v) = n \leq \theta\log(N)$ from (40). Now we show that node v or one of its ancestors get accepted:

$$\frac{A_v}{B_v} = \frac{\beta(v.\mathbf{n}+\boldsymbol{\alpha})\beta(\boldsymbol{\alpha}^x)\beta(\boldsymbol{\alpha}^y)}{\beta(\boldsymbol{\alpha})\beta(v.\mathbf{n}^x+\boldsymbol{\alpha}^x)\beta(v.\mathbf{n}^y+\boldsymbol{\alpha}^y)} \quad (142)$$

$$= \frac{\prod_{i=1}^k\prod_{j=1}^l(n_{ij}+\alpha_{ij}-1)!\prod_{i=1}^k(\boldsymbol{\alpha}_i^x-1)!}{\prod_{i=1}^k\prod_{j=1}^l(\alpha_{ij}-1)!\prod_{i=1}^k(n_{ij}+\boldsymbol{\alpha}_i^x-1)!} \frac{\prod_{j=1}^l(\boldsymbol{\alpha}_j^y-1)!(n+\sum_{i=1}^k\sum_{j=1}^l\alpha_{ij}-1)!}{\prod_{j=1}^l(n_{ij}+\boldsymbol{\alpha}_j^y-1)!(\sum_{i+1}^l\sum_{j=1}^l\alpha_{ij}-1)!} \quad (143)$$

$$\geq \frac{\prod_{i=1}^k(\alpha_i^x-1)!\prod_{j=1}^l(\alpha_j^y-1)!}{\prod_{i=1}^k\prod_{j=1}^l(\alpha_{ij}-1)!(\sum_{i=1}^k\sum_{j=1}^l\alpha_{ij}-1)!} \frac{\prod_{i=1}^k\prod_{j=1}^l(\frac{n_{ij}+\alpha_{ij}-1}{e})^{n_{ij}+\alpha_{ij}-1}}{\prod_{i=1}^k(\frac{n_i^x+\alpha_i^x-1}{e})^{n_i^x+\alpha_i^x-1}.e.\sqrt{(n_i^x+\alpha_i^x-1)}} \frac{\sqrt{2\pi(n_{ij}+\alpha_{ij}-1)}}{\prod_{j=1}^l(\frac{n_j^y+\alpha_j^y-1}{e})^{n_j^y+\alpha_j^y-1}.e.\sqrt{(n_j^y+\boldsymbol{\alpha}_j^y-1)}} \quad (144)$$

Therefore, we have

$$\begin{aligned}\log(\frac{A_v}{B_v}) &\geq C_\alpha + n(\mathcal{H}(\frac{n_i}{n})+\mathcal{H}(\frac{n_j}{n})-\mathcal{H}(\frac{n_{ij}}{n})) \\ &+\sum_{i=1}^k\sum_{j=1}^l(\alpha_{ij}-1)\log(n_{ij}) - \sum_{i=1}^k(\alpha_i^x-1)\log(n_i) \\ &-\sum_{j=1}^l(\alpha_j^y-1)\log(n_j) \quad (145)\end{aligned}$$

Setting $\alpha_{ij}=1$, and using (40):

$$\frac{A_v}{B_v} \geq CN^{1+\delta-\eta}\log(N)^{(2kl-k-l)} \quad (146)$$

where C is a constant depending only on k, l, δ, and θ. Therefore, v (or one of its ancestors) get accepted in Algorithm 2. This completes the proof of Lemma 5.

L.2 Proof of Lemma 6

The proof is done by induction on the number of vertices in the tree. For a tree with only one node, the statement is trivial as

$$A(root) = 1 \quad (147)$$

Assume $\sum_{v\in V_l(G)}A(v)=1$ holds for all the trees with the number of vertices below Z. Consider a tree G with Z nodes and assume w_0 is the node in G with maximum depth. Assume w is the parent of w_0, and w_{ij}, $1\leq i\leq k$ and $1\leq j\leq l$ are the children of w ($w_0\in\{w_{ij}\}$). First note that

$$\sum_{ij}A(w_{ij}) = \sum_{ij}A(w)\frac{w.n_{ij}+\mathbf{1}_{ij}}{\sum_{i,j}(w.n_{ij}+\mathbf{1}_{ij})} = A(w) \quad (148)$$

Now, define a tree G' as the tree obtained after removing $\{w_{ij}\}$ from G. Therefore, G' has $Z-kl$ nodes, and induction hypothesis holds for G'. Therefore:

$$\begin{aligned}&\sum_{v\in V_l(G)}A(v) \\ &= \sum_{v\in V_l(G')}A(v) - A(w) + \sum_{i,j}A(w_{ij}) \quad (149)\\ &= \sum_{v\in V_l(G')}A(v) = 1 \quad (150)\end{aligned}$$

This completes the proof of Lemma 6. ■

L.3 Proof of Lemma 7

For all the intermediate nodes, i.e., the nodes that are neither pruned nor accepted as a bucket, we have

$$\frac{A(v)}{B(v)} \leq (\log N)^{k+l-2kl}N^{1+\delta-\eta} \quad (151)$$

$$\frac{A(v)}{B_x(v)} \geq (\log N)^{k-kl}N^{1-\eta} \quad (152)$$

$$\frac{A(v)}{B_y(v)} \geq (\log N)^{l-kl}N^{\delta-\eta} \quad (153)$$

Therefore, for any intermediate node v, we have $A(v)\geq N^{-\eta}$ by multiplying (152) and (153) and dividing by (151). For each leaf node v (either pruned or accepted), its parent w is an intermediate node. Therefore, $A(w)\geq N^{-\eta}$, and we have

$$A(v)\geq A(w).\min_{v.n_{ij}\neq 0}\frac{v.n_{ij}+1}{n+kl} \geq N^{-\eta}\theta^{-1}\log(N)^{-1}. \quad (154)$$

Therefore, using Lemma 6:

$$\begin{aligned} 1 &= \sum_{v \in V_l(G^\eta)} A(v) && (155) \\ &\geq \sum_{v \in V_l(G^\eta)} N^{-\eta}\theta^{-1}(\log N)^{-1} && (156) \\ &= |V_l(G^\eta)|N^{-\eta}\theta^{-1}(\log N)^{-1} && (157) \end{aligned}$$

Therefore, $|V_l(G^\eta)| \leq \theta N^\eta \log(N)$. On the other hand, we know that in any homogeneous tree for which $kl > 1$, $|V(G)| \leq 2|V_l(G)|$ holds. Therefore, we have $|V(G^\eta)| \leq 2\theta N^\eta$.

This completes the proof of Lemma 7.■