▶ **Definition 4.6** (Parameter γ). *Given a Binary-BIDB circuit $\mathtt{C}$ define*

$$\gamma(\mathtt{C}) = \frac{\sum_{(\mathtt{v},\mathtt{c})\in \mathtt{E}(\mathtt{C})} |\mathtt{c}| \cdot \mathbb{1}_{\neg \textsc{isInd}(\mathtt{v_m})}}{|\mathtt{C}|(1,\ldots,1)}.$$

See comment on next pg to double check.

4.2 Our main result

We solve Problem 1.6 for any fixed $\epsilon > 0$ in what follows.

Algorithm Idea. Our approximation algorithm ($\textsc{Approximate}\widetilde{\Phi}$ pseudo code in Appendix D.1) is based on the following observation. Given a lineage polynomial $\Phi(\mathbf{X}) = \textsc{poly}(\mathtt{C})$ for circuit $\mathtt{C}$ over Binary-BIDB (recall that all c-TIDB can be reduced to Binary-BIDB by Proposition 2.4), we have:

$$\widetilde{\Phi}(p_1,\ldots,p_n) = \sum_{(\mathtt{v},\mathtt{c})\in \mathtt{E}(\mathtt{C})} \mathbb{1}_{\textsc{isInd}(\mathtt{v_m})} \cdot \mathtt{c} \cdot \prod_{X_i \in \mathtt{v}} p_i. \tag{2}$$

Given the above, the algorithm is a sampling based algorithm for the above sum: we sample (via $\textsc{SampleMonomial}$) $(\mathtt{v},\mathtt{c}) \in \mathtt{E}(\mathtt{C})$ with probability proportional to $|\mathtt{c}|$ and compute $\mathtt{Y} = \mathbb{1}_{\textsc{isInd}(\mathtt{v_m})} \cdot \prod_{X_i\in\mathtt{v}} p_i$. Repeating the sampling an appropriate number of times and computing the average of $\mathtt{Y}$ gives us our final estimate. $\textsc{OnePass}$ is used to compute the sampling probabilities needed in $\textsc{SampleMonomial}$ (details are in Appendix D).

Runtime analysis. We can argue the following runtime for the algorithm outlined above:

▶ **Theorem 4.7.** *Let $\mathtt{C}$ be an arbitrary* Binary-BIDB *circuit, define $\Phi(\mathbf{X}) = \textsc{poly}(\mathtt{C})$, let $k = \textsc{deg}(\mathtt{C})$, and let $\gamma = \gamma(\mathtt{C})$. Further let it be the case that $p_i \geq p_0$ for all $i \in [n]$. Then an estimate $\mathcal{E}$ of $\widetilde{\Phi}(p_1,\ldots,p_n)$ satisfying*

$$Pr\left(\left|\mathcal{E} - \widetilde{\Phi}(p_1,\ldots,p_n)\right| > \epsilon' \cdot \widetilde{\Phi}(p_1,\ldots,p_n)\right) \leq \delta \tag{3}$$

I think this should be $c^{-(k-1)}$

General comment (post-arxiv) Need a way to connect D to $\mathcal{D}$.

can be computed in time

$$O\left(\left(\textsc{size}(\mathtt{C}) + \frac{\log\frac{1}{\delta}\cdot k\cdot \log k \cdot \textsc{depth}(\mathtt{C})}{(\epsilon')^2\cdot(1-\gamma)^2\cdot p_0^{2k}}\right)\cdot \overline{\mathcal{M}}\left(\log\left(|\mathtt{C}|(1,\ldots,1)\right), \log\left(\textsc{size}(\mathtt{C})\right)\right)\right). \tag{4}$$

In particular, if $p_0 > 0$ and $\gamma < 1$ are absolute constants then the above runtime simplifies to $O_k\left(\left(\frac{1}{(\epsilon')^2}\cdot \textsc{size}(\mathtt{C})\cdot \log\frac{1}{\delta}\right)\cdot \overline{\mathcal{M}}\left(\log\left(|\mathtt{C}|(1,\ldots,1)\right), \log\left(\textsc{size}(\mathtt{C})\right)\right)\right)$.

The restriction on γ is satisfied by any 1-TIDB (where $\gamma = 0$ in the equivalent 1-BIDB of Proposition 2.4) as well as for all three queries of the PDBench BIDB benchmark (see Appendix D.10 for experimental results). Further, we can alo argue the following result:

▶ **Lemma 4.8.** *Given $\mathcal{RA}^+$ query Q and c-TIDB $\mathcal{D}$, let $\mathtt{C}$ be the circuit computed by $Q(D)$. Then, for the reduced Binary-BIDB $\mathcal{D}'$ there exists an equivalent circuit $\mathtt{C}'$ obtained from $Q(D')$, such that $\gamma(\mathtt{C}') \leq 1 - (c+1)^{-(k-1)}$ with $\textsc{size}(\mathtt{C}') \leq \textsc{size}(\mathtt{C}) + n\cdot\left(2^{(\lceil \log 2c\rceil)+1} - 1\right)$ and $\textsc{depth}(\mathtt{C}') = \textsc{depth}(\mathtt{C}) + \lceil \log 2c \rceil$.* $O(\log c)$

D

Write as $O(cn)$

Proof of Lemma 4.8. The circuit $\mathtt{C}$' is built from $\mathtt{C}$ in the following manner. For each input gate $\mathtt{g}_i$ with $\mathtt{g}_i.\mathtt{val} = X_t$, replace $\mathtt{g}_i$ with the circuit $\mathtt{S}$ encoding the sum $\sum_{j=1}^{c} j \cdot X_{t,j}$. We argue that $\mathtt{C}$' is a valid circuit by the following facts. Let $\mathcal{D} = \left(\{0,\ldots,c\}^D, D\right)$ be the original c-TIDB $\mathtt{C}$ was generated from. Then, by Proposition 2.4 there exists a ~~reduced~~

normal D

binary-BIDB

Expand on what you mean. E.g. explicitly say that X_{t_1,j_1} & X_{t_2,j_2} cannot cancel since they are in diff blocks

$\langle t, j\rangle$ where $D' = \{\langle t, j\rangle \mid t \in D, j \in [c]\}$

a (sub) circuit

$\mathcal{D}' = \left(\bigtimes_{t\in D'}\{0, c_t\}, \mathcal{P}'\right)$, from which the conversion from C to C' follows. Both POLY(C) and POLY(C') have the same expected multiplicity since (by Proposition 2.4) the distributions $\mathcal{P}$ and $\mathcal{P}'$ are equivalent and each $j \cdot \mathbf{W}'_{t,j} = \mathbf{W}_t$ for $\mathbf{W}' \in \{0,1\}^{cn}$ and $\mathbf{W} \in \{0,\ldots,c\}^{D}$. Finally, note that because there exists ~~$S' \in$ CSet(POLY(C))~~ encoding $\sum_{j=1}^{c} j \cdot X_{t,j}$ that is a *balanced* binary tree, the above conversion implies the claimed size and depth bounds of the lemma. Next, we argue the claim on $\gamma(C')$.

has this term been defined?

Consider the list of expanded monomials E(C) for c-TIDB circuit C. Let v be an arbitrary monomial such that the set of variables in v is $v_m = X_{t,1}, \ldots, X_{t,\ell}$ with ~~the number of~~ variables ~~$|v_m| = \ell$~~. Then v yields the set of monomials $E_v(C') = \left(j_1 \cdot X_{t,j}^{d_1}, \ldots, j_\ell \cdot X_{t,j_\ell}^{d_\ell}\right)_{j_1,\ldots,j_\ell \in [0,c]}$ in E(C'). Observe that cancellations can only occur for each $X_t^{d_t} \in v_m$. Consider the number of cancellations for $X_t^{d_t}$. Then $\gamma \le 1-(c+1)^{d_t-1}$, since for each element in $\left\{\bigtimes_{i\in[d_t], j_i\in[0,c]} X_{j_i}\right\}$ there are *exactly* $c+1$ surviving elements with $j_1 = \cdots = j_{d_t}$, i.e. $X_j^{d_t}$ for each $j \in [0,c]$. The rest of the $(c+1)^{d_t-1}$ cross terms cancel. Regarding the whole monomial v it is the case that the proportion of non-cancellations across each $X_t^{d_t} \in v_m$ multiply as non-cancelling terms for X_t can only be joined with non-cancelling terms of $X_{t'}^{d_{t'}}$. This then yields the ~~inequality~~ $1 - \prod_{i=1}^{\ell} (c+1)^{d_i - 1} \le \gamma \le 1 - (c+1)^{-(k-1)}$ where the inequalities take into account the fact that $\sum_{i=1}^{\ell} d_i \le k$.

Since this is true for arbitrary v, the bound follows for POLY(C'). ◀

I think you meant $t_1, \ldots, t_\ell$

not compress but product. i.e. distinct tuples in D'.

PROPAGATE

Update as per comment

fraction of cancelled monomials

Indexing seems wrong here.

We briefly connect the runtime in Eq. (4) to the algorithm outline earlier (where we ignore the dependence on $\overline{\mathcal{M}}(\cdot,\cdot)$, which is needed to handle the cost of arithmetic operations over integers). The SIZE(C) comes from the time taken to run OnePass once (OnePass essentially computes $|C|(1,\ldots,1)$ using the natural circuit evaluation algorithm on C). We make $\frac{\log\frac{1}{\delta}}{(\epsilon')^2\cdot(1-\gamma)^2\cdot p_0^{2k}}$ many calls to SampleMonomial (each of which essentially traces $O(k)$ random sink to source paths in C all of which by definition have length at most DEPTH(C)).

Finally, we address the $\overline{\mathcal{M}}(\log(|C|(1,\ldots,1)), \log(\text{SIZE}(C)))$ term in the runtime.

▶ **Lemma 4.9.** *For any* Binary-BIDB *circuit* C *with* DEG(C) $= k$, *we have* $|C|(1,\ldots,1) \le 2^{2^k\cdot \text{DEPTH}(C)}$. *Further, if* C *is a tree, then we have* $|C|(1,\ldots,1) \le \text{SIZE}(C)^{O(k)}$.

Also it would be better/useful to state the monomials in the expansion of $\left(\sum_{j=1}^{c} j \cdot X_{t,j}\right)^{d_t}$ are the monomial $X_{t,j}^{d_t}$

Note that the above implies that with the assumption $p_0 > 0$ and $\gamma < 1$ are absolute constants from Theorem 4.7, then the runtime there simplifies to $O_k\left(\frac{1}{(\epsilon')^2}\cdot \text{SIZE}(C)^2 \cdot \log\frac{1}{\delta}\right)$ for general circuits C. If C is a tree, then the runtime simplifies to $O_k\left(\frac{1}{(\epsilon')^2}\cdot \text{SIZE}(C) \cdot \log\frac{1}{\delta}\right)$, which then answers Problem 1.6 with yes for such circuits.

Finally, note that by Proposition E.1 and Lemma E.2 for any $\mathcal{RA}^+$ query Q, there exists a circuit C* for $\Phi[Q, D, t]$ such that DEPTH(C*) $\le O_{|Q|}(\log n)$ and SIZE(C) $\le O_k(T_{det}(Q, D, c))$. Using this along with Lemma 4.9, Theorem 4.7 and the fact that $n \le T_{det}(Q, D, c)$, we have the following corollary:

▶ **Corollary 4.10.** *Let Q be an $\mathcal{RA}^+$ query and $\mathcal{D}$ be a* Binary-BIDB *with $p_0 > 0$ and $\gamma < 1$ (where p_0, γ as in Theorem 4.7) are absolute constants. Let $\Phi(\mathbf{X}) = \Phi[Q, D, t]$ for any result tuple t with $\deg(\Phi) = k$. Then one can compute an approximation satisfying Eq. (3) in time $O_{k,|Q|,\epsilon',\delta}(T_{det}(OPT(Q), D, c))$ (given Q, D and p_i for each $i \in [n]$ that defines $\mathcal{P}$).*

Next, we note that the above result along with Lemma 4.8 answers Problem 1.5 in the affirmative as follows:

Assuming I did not cancel something. PROPAGATE.

▶ **Corollary 4.11.** *Let Q be an $\mathcal{RA}^+$ query and $\mathcal{D}$ be a c-TIDB with $p_0 > 0$ (where p_0 as in Theorem 4.7) is an absolute constant. Let $\Phi(\mathbf{X}) = \Phi[Q, D, t]$ for any result tuple*

Not correct. Argument is c out of c^{d_t} terms survive. I.e. frac of canceling terms $= 1 - \frac{c}{c^{d_t}} = 1 - c^{-(d_t-1)}$

this is why the final frac should be $1 - c^{-(k-1)}$