20240930 SERENA Algorithm

Instead of replacing the matching, merge two matchings.

For each alternating cycle, select a better side for the matching.

20241001 Dynamic Programming: Knapsack

① An algorithm that runs in $O(nB)$ $\begin{cases} B: \text{budget size} \\ n: \text{number of elements} \end{cases}$ s_i : size of item i; v_i : value of item i

$M(i, w)$ = largest possible value using jobs $\{0, \dots, i\}$ and total size $\le W$

$$M(0, w) = \begin{cases} v_0 & \text{if } w \ge s_0 \\ 0 & \text{o/w} \end{cases}$$

$$M(i, w) = \max\{M(i-1, w),\ v_i + M(i-1, w-s_i) \text{ if } w \ge s_i \text{, o/w } (-\infty)\}$$

Time: $O(nB)$; Space: $O(nB)$ } Answer: $M(n-1, B)$

② An algorithm that runs in $O(nV)$ { V: sum of v_i

$M(i, v)$ = min total knapsack size that suffices to get exactly value v from jobs $\{0, \dots, i\}$

$$M(i, v) = \min\{M(i-1, v),\ s_i + M(i-1, v-v_i) \text{ if } v \ge v_i \text{, o/w } (-\infty)\}$$

Time: $O(nv)$; Space: $O(nV)$ } Answer: max over feasible values

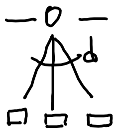
③ A fully polynomial time approximation scheme

$\forall \varepsilon$: runtime is poly (input size, ε^{-1})

$\tilde{v}_i = \lfloor \frac{v_i}{k} \rfloor$ where $k = \frac{\varepsilon v_{max}}{n}$ & run algorithm ② on $\tilde{v}$, multiply k to the answer.

Cummulative error is at most $\frac{\varepsilon v_{max}}{n} \cdot n$ & $v_{max} < V$

20241002 LT Process



For each ball, select degree d with prob. $P(d)$

i.i.d. uniform selection of d bins

k : # of input symbols

K : # of encoded symbols

$q(d, l)$: select an encoded symbol of degree d exactly d−2 of the input symbols is covered in the rest $k-l-1$ symbols & one symbol is decoded at step $k-l$.

20241002 SQL to CUDA

A GROUP BY P — When to materialize?

A JOIN B — Formalize indices bx, by, bz; wx, wy, wz

20241002 Traffic Patterns of Switches

- Uniform
- Quasi-diagonal
- Log-diagonal
- Diagonal

Bursty Arrivals:

- with ε probability use the same output port

20241002 Small Batch QPS & Sliding Window QPS

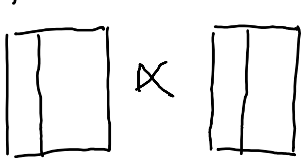
Multi-proposal Combine Proposals from Different Rounds

20241003 Rateless Bloom Filtering Semi-join

A large amount of data is indexed by LSM Trees.

Therefore they are ordered internally.

If I want to compute semi-join of two distributed tables,



Currently the database needs a scan before it starts join.

Co-filter?

20241003 Markov Decision Process (MDP)

Finite State Space : S

Finite Action Space : A

Start State $s_1 \in S$

Transition Probability : $p(s' | s, a)$

Reward $r(s,a) \in \mathbb{R}$

Goal : Given time region T, design a policy $\pi(a | s_{[1..t]})$ that maximize the expected total reward.

$$\max_{\pi} \mathbb{E}\left\{\sum_{t=1}^{T} r(s_t, a_t)\right\}$$

$$s_{t+1} \sim p(\cdot | s_t, a_t)$$

$$a_t \sim \pi(a_t | s_{[1...t]})$$

Observation 1 : The optimal policy is deterministic.

Observation 2 : Only the current state and the remaining time matters.

Let $v(s,t)$ be the optimal reward and time region t.

$$v(s,t) = \max_{a \in A} \left\{ r(s,a) + \sum_{s' \in S} p(s'|s,a) \cdot v(s', t-1) \right\}$$

20241004 Randomized Multi-cut

Given $G=(V,E)$ and a metric $d: V\times V \to \mathbb{R}_{\geq 0}$

The diameter of a graph G is $D(G)$.

Divide G's vertices & get their induced subgraph (G_i),

Definition of (Low Diameter Decomposition)

$D(G_i) \leq D/2$ for each i (Property 1)

$\forall\, u,v \in V$: $\mathbb{P}\{u,v \text{ are from different subgraphs}\} \leq \frac{d(u,v)}{D}\cdot\alpha$ (Property 2)

For $\alpha = |\log V|$ (or better $\alpha = \log\left|\frac{B(u,D/2)}{B(u,D/16)}\right|$), we can

$B(u,d)$: size of vertex nearer then d (ball)

construct a such a random partition.

Choose $r \sim \mathrm{Unif}(\frac{D}{8},\frac{D}{4})$

Choose a random permutation σ

For each vertex $\sigma(i)$, draw a ball of radius r.

Proof via Tree Embeddings

20241004 Tree Embeddings

Given a graph $G=(V,E)$ with metric d.

Find a random tree T

① $d(u,v) \leq d_T(u,v)$

② $\mathbb{E}\, d_T(u,v) \leq \alpha\cdot d(u,v)$ usually α is $O(\log n)$

20241004 Group Vertex Cover

Find a subset of edges $E' \subseteq E$, s.t. at least one vertex from each vertex set V_i is covers.

State formally how this LP gives a relaxation of group vertex cover

$$\min \sum_e c_e x_e$$

s.t. $x_e \geq 0$ $\quad \forall i \in [k]$

$$\sum_{e \in (S, V \setminus S)} x_e \geq 1 \quad \forall S : S \ni r \wedge |S \cap G_i| = 0$$

We can assume W.L.O.G. that the tree has height $O(\log |G|)$

1) Round all x_e^* to power of 2

2) Discard $x_e^* \leq \frac{1}{10n}$

What's the expected number of G connected to r ?

It is should be at least 1.

Define $X_i = \begin{cases} 1 & \text{if } i \in G \text{ is connected to } r \\ 0 & \text{o/w} \end{cases}$

$$\mathbb{P}\{\sum_i X_i = 0\} \leq 1 - (\log n)^{-1}$$

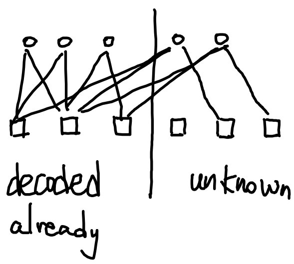
Therefore, we bound $\sum_{i,j} \mathbb{E}[X_i X_j] = \Delta = O(\log n)$.

Using Chebyshev $\mathbb{P}\{\sum_i X_i = 0\} \leq \frac{\mathrm{Var}(X)}{(\mathbb{E} X)^2} \lesssim \frac{\mu + \Delta}{\mu^2}$

Janson's Inequality $\mathbb{P}\{\sum_i X_i = 0] \leq \exp(-\frac{\mu^2}{2\Delta}) = \exp(-\frac{1}{\log n}) \approx 1 - \frac{1}{\log n}$

20241004 Luby Code & Good Degree Distribution

Observation: If two encoded packets are the same modulo already decoded packets, then removing one of them will not effect the decoding process.



Observation: If at some point, no encoded packet has remaining degree one, the process will also fail (or stuck)

20241004 Set Reconcilation for (?)

LSM Tree Compaction

When we merge two SSTs, the set of keys will merge.

But value updates cannot be handled.

IBLT

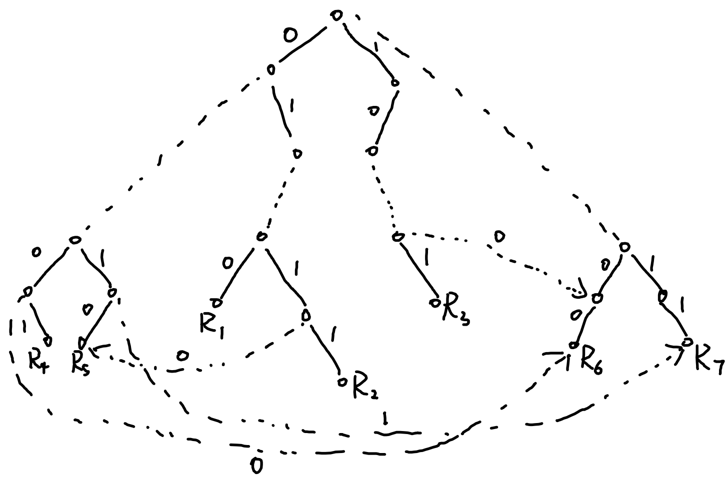
Set Reconcilation for CRDT operations?

Maintain a IBLT along a given CRDT.

20241005 Discussion: Soliton Distribution

Soliton distribution is derived to make each decoding step produce one expected input symbol.

20241007 Grid of Tries



R4 R5 R6 R7　　R1 R2 R4 R5　R3 R6 R7　　R6 R7

R6 R7

20241007 iSLIP