# Parallel Approximate Undirected Shortest Paths Via Low Hop Emulators

#### Alexandr Andoni, Clifford Stein, Peilin Zhong

Columbia University

April 20, 2020

Assess the efficiency of parallel algorithms by analyzing:

- Depth,  $D_P(n)$ , is length of the longest series of operations
- Work,  $W_P(n)$ , is runtime in a sequential setting

We seek algorithms that are:

- Polylog(arithmic) depth ,  $D_P(n) = \mathcal{O}(\operatorname{poly}(\log n))$
- Nearly work-efficient,  $W_P(n) = \mathcal{O}(T_S(n) \cdot \text{poly}(\log n))$

Given a (non-negative) weighted graph G = (V, E) shortest paths problem finds the shortest path from  $s \rightarrow t$  in G.

The single source shortest paths problem is the shortest path from  $s \in S$  to every vertex in V.

Let n = |V| and m = |E|. What algorithms are there?

Given a (non-negative) weighted graph G = (V, E) shortest paths problem finds the shortest path from  $s \rightarrow t$  in G.

The single source shortest paths problem is the shortest path from  $s \in S$  to every vertex in V.

Let n = |V| and m = |E|. What algorithms are there?

- Dijkstra's algorithm requires  $\mathcal{O}(m + n \cdot \log(n))$  work/depth
- Bellman Ford requires  $\mathcal{O}(nm)$  work and  $\mathcal{O}(n)$  depth
- Floyd–Warshall requires  $\mathcal{O}(n^3)$  work and  $\mathcal{O}(\log n)$  depth

Given a (non-negative) weighted graph G = (V, E) shortest paths problem finds the shortest path from  $s \rightarrow t$  in G.

The single source shortest paths problem is the shortest path from  $s \in S$  to every vertex in V.

Let n = |V| and m = |E|. What algorithms are there?

- Dijkstra's algorithm requires  $\mathcal{O}(m + n \cdot \log(n))$  work/depth
- Bellman Ford requires  $\mathcal{O}(nm)$  work and  $\mathcal{O}(n)$  depth
- Floyd–Warshall requires  $\mathcal{O}(n^3)$  work and  $\mathcal{O}(\log n)$  depth

Question: Does there exist a  $(1 + \varepsilon)$ -shortest paths algorithm with  $\mathcal{O}(m \cdot \operatorname{poly}(\log n))$  work and  $\mathcal{O}(\operatorname{poly}(\log n))$  depth?

### Previous and New Results

Previous Results

- Use hopset [Coh94] to solve  $(1 + \varepsilon)$ -shortest paths in polylog depth and  $\Omega(n^{2.1})$  work
- Any  $(1 + \varepsilon)$ -shortest paths algorithm with polylog depth using only hopsets requires  $\Omega(n^{1+\varepsilon'})$  work [ABP18]

## Previous and New Results

#### Previous Results

- Use hopset [Coh94] to solve  $(1 + \varepsilon)$ -shortest paths in polylog depth and  $\Omega(n^{2.1})$  work
- Any (1 + ε)-shortest paths algorithm with polylog depth using only hopsets requires Ω(n<sup>1+ε'</sup>) work [ABP18]

#### This Paper

- Construct a novel data structure, a low hop emulator, to solve approximate single source shortest paths
- Compute (1 + ε)-approximate shortest path in polylog depth and nearly linear work via [She17] and low hop emulators

## Previous and New Results

#### Previous Results

- Use hopset [Coh94] to solve  $(1 + \varepsilon)$ -shortest paths in polylog depth and  $\Omega(n^{2.1})$  work
- Any (1 + ε)-shortest paths algorithm with polylog depth using only hopsets requires Ω(n<sup>1+ε'</sup>) work [ABP18]

#### This Paper

- Construct a novel data structure, a low hop emulator, to solve approximate single source shortest paths
- Compute (1 + ε)-approximate shortest path in polylog depth and nearly linear work via [She17] and low hop emulators

Given a graph G = (V, E), a low hop emulator is a weighted graph H = (V, F) where any shortest (s - t)-path with  $\mathcal{O}(\log \log n)$  edge traversals and  $|F| = \mathcal{O}(m \cdot \operatorname{poly}(\log n))$ .

#### Constructing a Low Hop Emulator

- Constructing a subemulator
- Recursive subemulators
- Collapsing into a low-hop emulator
- Constructing a low-hop emulator (in parallel)

(1 + ε) shortest paths in polylog depth and nearly linear work
 Solving shortest paths via optimization and Sherman's framework

## Constructing a low hop emulator

To construct a low hop emulator,

- Construct a subemulator
- Recursive subemulators
- $\textcircled{O} \quad \text{Collapse all subemulators down to a single graph} \rightarrow \text{low-hop emulator}$

## Constructing a low hop emulator

To construct a low hop emulator,

- Construct a subemulator
- Recursive subemulators
- ${f 0}$  Collapse all subemulators down to a single graph ightarrow low-hop emulator

A subemulator is a graph H = (S, F') where  $S \subset V$  and F' is a weighted edge set that approximates distances well.

- Every vertex  $v \in V$  is close to a vertex in S
- ② Distances in H approximate distances in G

### Constructing a subemulator

A subemulator is a graph H = (S, F') where  $S \subset V$  and F' is a weighted edge set that approximates distances well.

Select vertices first by sampling

Our analysis depends on the ball  $B_{G,b}(v)$ , which is the closest *b* vertices (graph distance) to *v* in *G*,



### Constructing a subemulator

A subemulator is a graph H = (S, F') where  $S \subset V$  and F' is a weighted edge set that approximates distances well.

- Select vertices first by sampling
- 2 Add more vertices to ensure vertex in V is close to a vertex in S

Our analysis depends on the ball  $B_{G,b}(v)$ , which is the closest *b* vertices (graph distance) to *v* in *G*,



### Constructing a subemulator

A subemulator is a graph H = (S, F') where  $S \subset V$  and F' is a weighted edge set that approximates distances well.

- Select vertices first by sampling
- 2 Add more vertices to ensure vertex in V is close to a vertex in S
- Add edges and assign weights so that local distances are well-approximated

Our analysis depends on the ball  $B_{G,b}(v)$ , which is the closest *b* vertices (graph distance) to *v* in *G*,



#### Fix a ball size b (to be defined later).

- Construct S by sampling every vertex with probability  $p = \min(50 \frac{\log n}{b}, \frac{1}{2})$
- **2** If  $v \in V$  is not near any vertex in S, add v to S
- $\textbf{Store the leader } q(v) \leftarrow \text{closest vertex } u \in S \text{ to } v \in V$

Output: A sparse vertex set S and mapping  $q: V \rightarrow S$ 

#### Adding edges

Adding edges:

• For all  $(u, v) \in E$ , add edge (q(u), q(v))

**2** Fix 
$$v \in V$$
.  $\forall u \in B(v)$ , add edge  $((q(u), q(v)))$ 

When adding an edge e = (q(u), q(v)) to F, update weight

$$w(e) = \min \begin{cases} w(e), \text{ (initialize to } \infty) \\ d_G(q(u), u) + d_G(u, v) + d_G(v, q(v)) \end{cases}$$



#### Properties of subemulators

Given a subemulator H = (S, F'), Size

• 
$$\mathbb{E}[|S|] < n$$

• 
$$|F'| \leq m + nb$$

Distance approximation

• For any  $u, v \in S$ ,

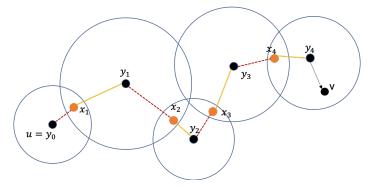
$$d_G(u, v) \leq d_H(u, v) \leq 8 \cdot d_G(u, v).$$

• For any  $u, v \in V$ ,

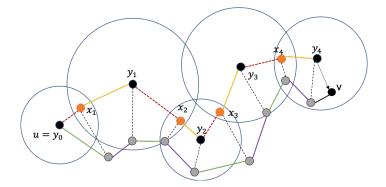
$$d_H(q(u),q(v)) \leq d_G(u,q(u)) + 22 \cdot d_G(u,v) + d_G(v,q(v)).$$

## 8-approximation for shortest (u, v)-path

Consider arbitrary  $u, v \in S$  and its shortest path in G.



## 8-approximation for shortest (u, v)-path



$$d_{H}(u, v) \leq \sum_{i=1}^{t} w_{H}(q(y_{i-1}), q(x_{i})) + w_{H}(q(x_{i}), q(y_{i})) + w_{H}(y_{t}, v)$$
  
$$\leq 8 \cdot \left(\sum_{i=1}^{t} w_{G}(y_{i-1}, x_{i}) + w_{G}(x_{i}, y_{i}) + w_{G}(y_{t}, v)\right)$$
  
$$= 8 \cdot d_{G}(u, v).$$

High level approach:

- Construct a subemulator
- 2 Recursive subemulators
- $\textcircled{O} Collapse all subemulators down to a single graph \rightarrow \mathsf{low-hop} \ \mathsf{emulator}$

Set  $b_0 \ll n$ . Let the first subemulator  $H_0$  be the original graph G. While  $|V(H_i)| > b_i$ : •  $\forall v \in V(H_i)$ , save  $B_{H_i, h_i}(v)$ 2  $H_{i+1} \leftarrow \text{SUBEMULATOR}(H_i, b_i)$ **③** Update  $b_{i+1} \leftarrow b_i^{1.25}$  and  $i \leftarrow i+1$ Output: Set of subemulators  $H_i$  and set of balls B(v'),  $\forall v' \in H_i$ 

#### Recursive subemulators

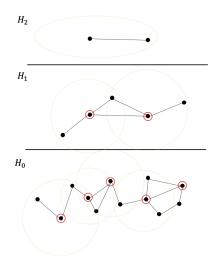


Figure: Recursive subemulators

Let u, v be arbitrary vertices whose distance we want to compute.

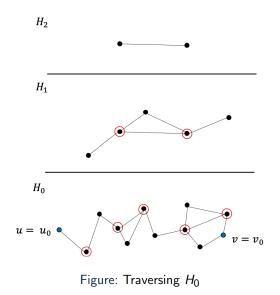
$$d = 0, i = 0$$
  

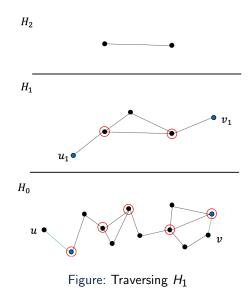
$$u_0 = u, v_0 = v$$
  
While  $u_i, v_i$  are not in the same ball:  

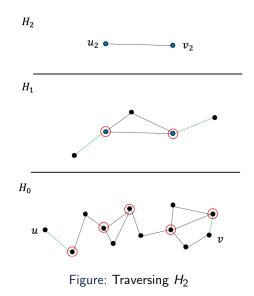
$$d \leftarrow d + d_{H_i}(q(u_i), u_i)) + d_{H_i}(v, q(v_i))$$
  

$$u_{i+1} \leftarrow q(u_i), v_{i+1} \leftarrow q(v_i)$$
  

$$i \leftarrow i + 1$$
  
Return  $d + d_{H_i}(u_i, v_i)$ .







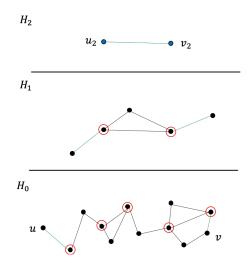
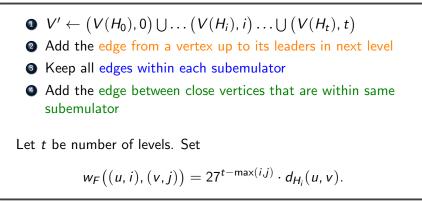


Figure: Near vertices exact distance

- Number of subemulators is  $\mathcal{O}(\log \log n)$
- **2** Traversing up-down subemulators to get (u, v)-shortest path is  $26^{\mathcal{O}(\log \log n)} = \operatorname{poly}(\log n)$ -approximation
- Store  $\mathcal{O}(\log \log n)$  subemulators and each vertex's ball

High level approach:

- Construct a subemulator
- ② Recursive subemulators
- $\textcircled{O} Collapse all subemulators down to a single graph \rightarrow \mathsf{low-hop} \ \mathsf{emulator}$



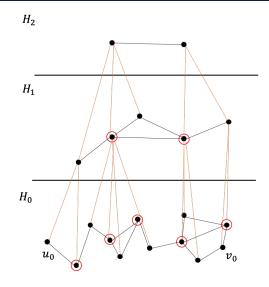


Figure: Edges between levels

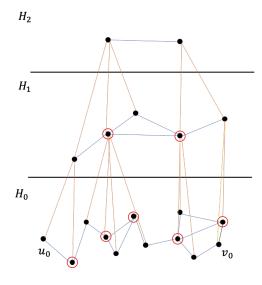


Figure: Edges within a level

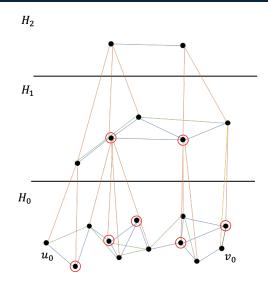


Figure: Edges between nearby vertices

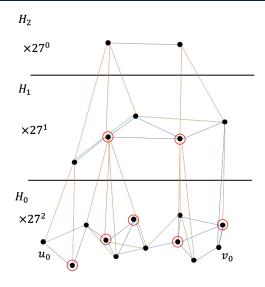


Figure: Scaling edges

Can remove copies of a vertex v, (v, 0), (v, 1), ..., and preserve edges by squishing all copies of vertex to bottom level

OK since have orange edges between (v, i) and (v, i+1) are weight zero.

There always exists a shortest (u - v)-path with minimial edge traversals without using blue edges

Close edges will be covered by green edges and far edges will never be used.

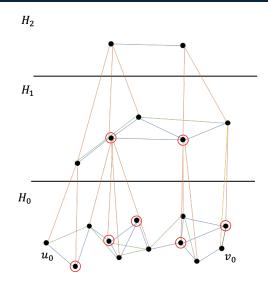


Figure: Almost low-hop emualtor

#### Collapsing towards a low-hop emulator

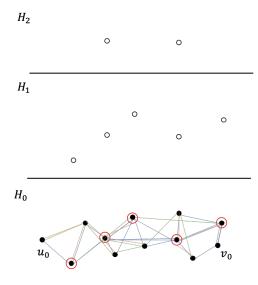


Figure: Removing redundant vertices

#### Collapsing towards a low-hop emulator

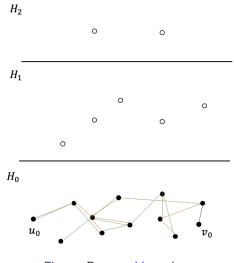


Figure: Remove blue edges

## Recap

High level approach:

- Construct a subemulator
- 2 Recursive subemulators
- $\textcircled{O} Collapse all subemulators down to a single graph \rightarrow \mathsf{low-hop} \ \mathsf{emulator}$

## Recap

High level approach:

- Construct a subemulator
- Recursive subemulators
- $\textcircled{O} \quad \text{Collapse all subemulators down to a single graph} \rightarrow \text{low-hop emulator}$

Our constructed graph H = (V', F):

- Shortest path requires  $\mathcal{O}(\log \log n)$  traversals
- **②** Distance is  $27^{\mathcal{O}(\log \log n)} = \operatorname{poly}(\log n) \operatorname{approximated}$

**3** 
$$V' = V$$

• 
$$\mathbb{E}[|F|] = \mathcal{O}(n \cdot \operatorname{poly}(\log n))$$

#### Corollary: H is a low-hop emulator.

Assume that we have computer that allows concurrent reads and concurrent writes (CRCW).

- Selecting vertices
- Adding edges
- Recursive subemulators
- Ollapsing towards a low-hop emulator

Uses path-doubling idea as done in Floyd-Warshall.

$$L_{0}(v) \leftarrow \text{closest } b \text{ neighbors of } v \text{ (including } v \text{ itself)}$$
  
For  $u \in L_{0}(v)$ , compute  $\text{dist}^{(0)}(v, u)$   
For  $i = 1, \dots, t = \lceil \log n \rceil$ :  
  $\bullet \forall v, u \in V$ , set  $\text{dist}^{(i)}(v, u) \leftarrow \infty$   
  $\bullet \text{ Fix } v \in V.$  Consider  $(v, x, u)$  such that  $x \in L_{i-1}(v)$  and  $u \in L_{i-1}(x)$ . Update  $\text{dist}^{(i)}(v, u)$  if  $\text{dist}^{(i-1)}(v, x) + \text{dist}^{(i-1)}(x, u)$  is smaller  
  $\bullet \text{ For every } v \in V$ , set  $L_{i}(v) \leftarrow \text{closest } b$  vertices  
  $\text{Output: } B(v) \leftarrow L_{t} \text{ and exact distance } d^{(t)}(v, \cdot)$ 

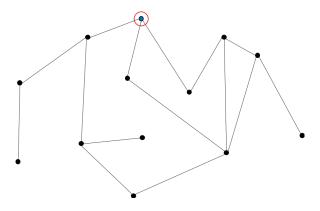


Figure: Expanding ball with b = 5. Reach of 0

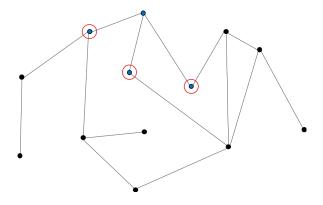


Figure: Expanding ball with b = 5. Reach of 1

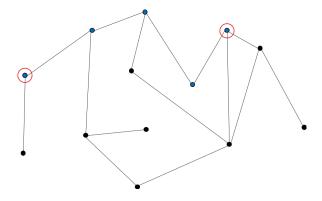


Figure: Expanding ball with b = 5. Reach of 2

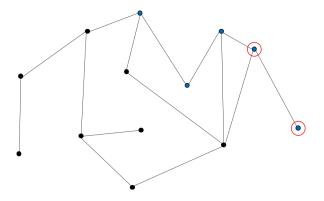


Figure: Expanding ball with b = 5. Reach of 4

 $L_0(v) \leftarrow \text{closest } b \text{ neighbors of } v.$  Compute  $\text{dist}^{(0)}(v, u)$ 

- Use polylog depth and nearly work-efficient parallel sorting algorithm  $\Rightarrow \mathcal{O}(\log^2 n)$  depth and  $\mathcal{O}(m \log^2 n)$  work
- $\bullet$  Initialize distance in  $\mathcal{O}(1)$  depth and  $\mathcal{O}(m)$  work

For 
$$i = 1, ..., t = \lceil \log n \rceil$$
:  
•  $\forall v, u \in V$ , set  $dist^{(i)}(v, u) \leftarrow \infty$   
• Fix  $v \in V$ . Consider  $(v, x, u)$  such that  $x \in L_{i-1}(v)$  and  $u \in L_{i-1}(x)$ . Update  $dist^{(i)}(v, u)$  if ...  
• For every  $v \in V$ , set  $L_i(v) \leftarrow$  closest b vertices

We perform ball computations on each vertex in G or a subemulator  $H_i$ . Repeat log n times:

- $\leq b^2$  vertex pairs per vertex  $v \in V$ .
- Use sorting algorithm

Fix a ball size b (to be defined later).

- Construct S by sampling every vertex
- 2 If  $v \in V$  is not near any vertex in S, add v to S
- **③** Store the leader  $q(v) \leftarrow$  closest vertex  $u \in S$  to  $v \in V$

Fix a ball size b (to be defined later).

- Construct S by sampling every vertex
- **2** If  $v \in V$  is not near any vertex in S, add v to S
- Store the leader  $q(v) \leftarrow$  closest vertex  $u \in S$  to  $v \in V$

- $\ \, {\cal O}(1) \ {\rm depth} \ {\rm and} \ {\cal O}(n) \ {\rm work} \ \,$
- Oitto

## Adding edges (in parallel)

Adding edges:

• For all  $(u, v) \in E$ , add edge (q(u), q(v))

**2** Fix  $v \in V$ .  $\forall u \in B(v)$ , add edge ((q(u), q(v)))

When adding an edge e = (q(u), q(v)) to F, update weight

$$w(e) = \min \begin{cases} w(e), \text{ (initialize to } \infty) \\ d_G(q(u), u) + d_G(u, v) + d_G(v, q(v)) \end{cases}$$

# Adding edges (in parallel)

Adding edges:

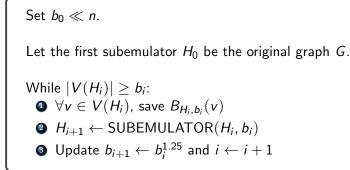
- For all  $(u, v) \in E$ , add edge (q(u), q(v))
- **2** Fix  $v \in V$ .  $\forall u \in B(v)$ , add edge ((q(u), q(v)))

When adding an edge e = (q(u), q(v)) to F, update weight

$$w(e) = \min \begin{cases} w(e), \text{ (initialize to } \infty) \\ d_G(q(u), u) + d_G(u, v) + d_G(v, q(v)) \end{cases}$$

- ${\small O } \ {\small Uses ball to compute exact distances. } {\small {\cal O}(1) \ {\small depth and } {\small {\cal O}(m) \ {\small work}}$
- 2 Ditto
- Oitto

## Recursive subemulators (in parallel)



## Recursive subemulators (in parallel)

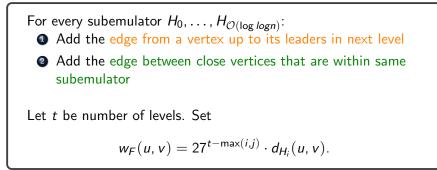
Set 
$$b_0 \ll n$$
.  
Let the first subemulator  $H_0$  be the original graph  $G$ .  
While  $|V(H_i)| \ge b_i$ :  
 $\forall v \in V(H_i)$ , save  $B_{H_i,b_i}(v)$   
 $H_{i+1} \leftarrow \text{SUBEMULATOR}(H_i, b_i)$   
 $\forall Update b_{i+1} \leftarrow b_i^{1.25} \text{ and } i \leftarrow i+1$ 

Repeat  $\mathcal{O}(\log \log n)$  times:

- Ball construction
- Proved is polylog depth and nearly linear
- 3 O(1)

## Collapsing towards a low-hop emulator (in parallel)

For every subemulator H<sub>0</sub>,..., H<sub>O(log logn)</sub>:
Add the edge from a vertex up to its leaders in next level
Add the edge between close vertices that are within same subemulator
Let t be number of levels. Set
w<sub>F</sub>(u, v) = 27<sup>t-max(ij)</sup> ⋅ d<sub>Hi</sub>(u, v).



Repeat  $\mathcal{O}(\log \log n)$  for each subemulator

- $\mathcal{O}(1)$  depth and  $\mathcal{O}(n \cdot \operatorname{poly}(\log n))$  work
- 2 Ditto
- Weights: Ditto

Recap. Assuming we have a concurrent read, concurrent write computational model, we can solve the following problems in polylog depth and near linear work:

- Selecting vertices
- Adding edges
- Recursive subemulators
- Sollapsing towards a low-hop emulator

Constructing a low-hop emulator can be done in same depth and work.

Compute the following problems with polylog approx, polylog depth, and nearly linear work:

- SSSP via Bellman-Ford
- Ø Bourgain's embedding
- O Low diameter decomposition
- Metric tree embedding

## Uncapacitated min-cost flow (transshipment) problem

Let  $W \in \mathbb{R}^{m \times m}$  be a diagonal matrix of weights. Let  $A \in \mathbb{R}^{n \times m}$  be the incidence matrix,

$$A_{iu} = \begin{cases} 1 : \exists edge \ u = (i, j) \\ -1 : \exists edge \ u = (j, i) \\ 0 : otherwise \end{cases}$$

Find a vector  $f \in \mathbb{R}^m$  such that  $\min_{f \in \mathbb{R}^m} ||Wf||_1$ s.t. Af = b, where  $b \in \mathbb{R}^n$  is the demand vector, where we require  $\sum_i b_i = 0$ .

If 
$$b(s) = 1$$
,  $b(t) = -1$ , then solves  $(s, t)$ -shortest path length

Andoni, Stein, Zhong (Columbia)

Low Hop Emulators and Applications

### Uncapacitated min-cost flow problem

An equivalent problem:

Let x = Wf. Find the optimal  $x^*$  such that  $x^* = \min_{x \mathbb{R}^m} ||x||_1$ s.t.  $AW^{-1}x = b$ .

## Uncapacitated min-cost flow problem

An equivalent problem:

Let x = Wf. Find the optimal  $x^*$  such that  $x^* = \min_{x \mathbb{R}^m} ||x||_1$ s.t.  $AW^{-1}x = b$ .

Lemma: There exists  $(1 + \varepsilon)$ -approximation algorithm to the optimization problem above in that runs in polylog depth if there exists a matrix P such that

 $||x^*||_1 \le ||Pb||_1 \le \mathcal{O}(\text{poly}\log n) \cdot ||x^*||_1$ .

The Earth Mover's Distance (EMD) problem is  

$$\min_{\pi: V \times V \to \mathbb{R}_{\geq 0}} \sum_{\substack{(u, v) \in V \times V}} \pi(u, v) \cdot \|\phi(u) - \phi(v)\|_{1}$$
s.t.  $\forall u \in V, \sum_{v \in V} \pi(u, v) - \sum_{v \in V} \pi(v, u) = b_{u}.$ 

Find a vector  $f \in \mathbb{R}^m$  such that

$$\min_{f \in \mathbb{R}^m} \|Wf\|_1$$
s.t.  $Af = b$ ,

#### Theorem (Bourgain's Embedding)

Every metric space  $(V, d_V)$  can be embedded in  $\ell_p$  with distortion  $\mathcal{O}(\log n)$ .

Given a graph G = (V, E) and distance  $d: V \times V \to \mathbb{R}^+$ , there exists a mapping  $\phi: V \to [\Delta]^{\mathcal{O}(\log^2 n)}$  such that

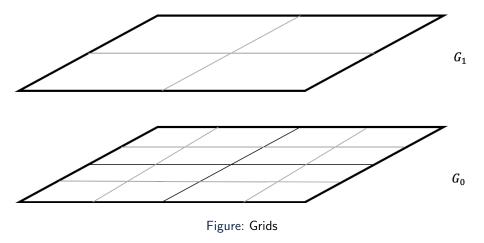
$$d(u, v) \leq \|\phi(u) - \phi(v)\|_1 \leq \mathcal{O}(\log n)d(u, v),$$

where  $\Delta \leq \sum_{e \in E} w_e$ .

Using Bourgain's embedding via low–hop emulators, we can find a mapping  $\phi: V \to [\Delta]^\eta$  such that

$$\mathsf{OPT}_{\mathit{EMD}}(b) \leq \mathcal{O}(\mathsf{poly}(\log n))\mathsf{OPT}_{\mathit{transshipment}}(b).$$

## Creating a preconditioner using grids



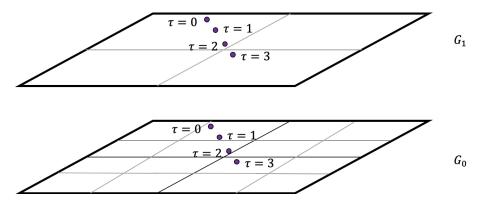


Figure: Grids with points

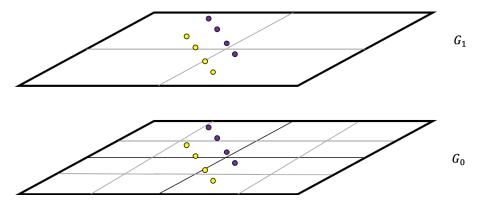
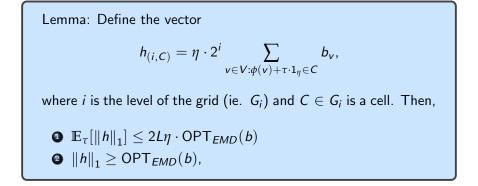


Figure: Points in same cell vs. random shift  $\boldsymbol{\tau}$ 

Construct a series of  $L = 1 + \log \Delta$  grids  $G_i$  as above.

Let  $\tau \in [\Delta]$  be a uniform random variable.



where  $\eta = \mathcal{O}(\log^2 n)$ .

Construct a vector h' such that  $\mathbb{E}[\|h\|_1] = \|h'\|_1$ ,

$$h'_{(i,C,\tau)} = \eta \cdot \sum_{v \in V: \phi(v) + \tau \cdot \mathbf{1}_{\eta} \in C} b_v.$$

Construct a vector h' such that  $\mathbb{E}[\|h\|_1] = \|h'\|_1$ ,

$$h'_{(i,C,\tau)} = \eta \cdot \sum_{v \in V: \phi(v) + \tau \cdot \mathbf{1}_{\eta} \in C} b_v.$$

Prescribes a matrix P' where

$$\mathcal{P}'_{(i,\mathcal{C}, au),\mathbf{v}} = egin{cases} \eta &: \ \phi(\mathbf{v}) + au \cdot \mathbf{1}_\eta \in \mathcal{C} \ 0 &: \ ext{otherwise} \end{cases}$$

Then h' = P'b, and

$$\left\|\boldsymbol{x}^*\right\|_1 \leq \left\|\boldsymbol{h}'\right\|_1 = \left\|\boldsymbol{P}'\boldsymbol{b}\right\|_1 \leq \left(\operatorname{poly}(\log n)\right) \left\|\boldsymbol{x}^*\right\|_1$$

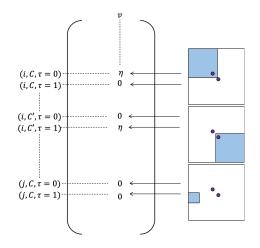


Figure: Preconditioner vs. Grids

## Uncapacitated min-cost flow problem

We want to solve the following problem:

Let 
$$x = Wf$$
. Find the optimal  $x^*$  such that  
 $x^* = \min_{x \in \mathbb{R}^m} ||x||_1$   
s.t.  $AW^{-1}x = b$ .

### Uncapacitated min-cost flow problem

We want to solve the following problem:

Let 
$$x = Wf$$
. Find the optimal  $x^*$  such that  
 $x^* = \min_{x \in \mathbb{R}^m} ||x||_1$   
s.t.  $AW^{-1}x = b$ .

Lemma: There exists  $(1+\varepsilon)-$ approximation algorithm to the optimization problem above in that runs in polylog depth using a matrix P' where

$$\|x^*\|_1 \le \|P'b\|_1 \le \mathcal{O}(\operatorname{poly}\log n) \cdot \|x^*\|_1$$
 .

#### Constructing a Low Hop Emulator

- Constructing a subemulator
- Recursive subemulators
- Collapsing into a low-hop emulator
- Constructing a low-hop emulator (in parallel)

(1 + ε) shortest paths in polylog depth and nearly linear work
 Solving shortest paths via optimization and Sherman's framework

# Thanks. Questions?