Sparsification Framework for Directed Densest Subgraph: MPC, semi-streaming, sublinear-time

Slobodan Mitrović
(UC Davis)

Theodore Pan
UC Davis

Densest subgraph

Undirected densest subgraph

G = (V,E), $S \subseteq V$

$$\boldsymbol{\rho(S) = \frac{|E(S)|}{|S|}}$$

DS(G) = $\max_{S} \rho(S)$

Densest subgraph

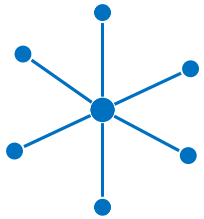
Undirected densest subgraph

G = (V,E),   S ⊆ V

$$\rho(S) = \frac{|E(S)|}{|S|}$$

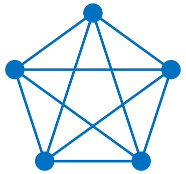
DS(G) = $\max_S \rho(S)$

k-star



$\rho = k/(k+1)$

clique



$\rho = (k-1)/2$

Densest subgraph

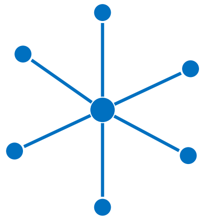
Undirected densest subgraph

G = (V,E),   S ⊆ V

$$\boldsymbol{\rho(S) = \frac{|E(S)|}{|S|}}$$

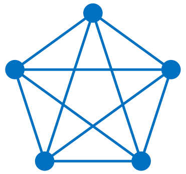
DS(G) = $\max_S \rho(S)$

k-star



$\rho \;=\; k/(k+1)$

clique



$\rho \;=\; (k-1)/2$

Directed densest subgraph

G = (V,E),   S,T ⊆ V

$E(S,T)$: arcs from S to T

$$\boldsymbol{\rho(S,T) = \frac{|E(S,T)|}{\sqrt{|S| \cdot |T|}}}$$

DDS(G) = $\max_{S,T} \rho(S,T)$

Densest subgraph

Undirected densest subgraph

G = (V,E),   S ⊆ V

$$\rho(S) = \frac{|E(S)|}{|S|}$$

DS(G) = $\max_{S} \rho(S)$

Directed densest subgraph

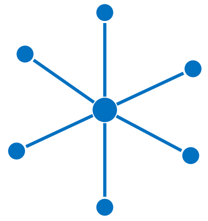
G = (V,E),   S,T ⊆ V

$E(S,T)$: arcs from S to T

$$\rho(S,T) = \frac{|E(S,T)|}{\sqrt{|S| \cdot |T|}}$$

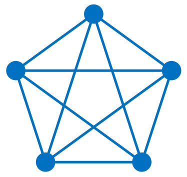
DDS(G) = $\max_{S,T} \rho(S,T)$

k-star



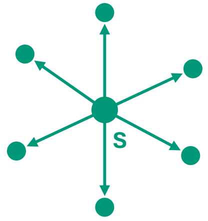
$\rho = k/(k+1)$

clique



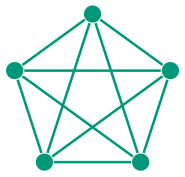
$\rho = (k-1)/2$

directed k-star



$\rho = \sqrt{k}$

directed clique



$\rho = k - 1$

Why study directed densest subgraph (DDS)?

A classical graph primitive

Introduced in 1999

large graph structure

[Kannan, Vinay ’99]

Web graph substructures

local structure in directed graphs

[Kannan, Vinay ’99; Kleinberg ’99]

Spectral properties

structural viewpoint beyond density alone

[Andersen ’10]

Applications

Fraud detection

[Prakash, Sridharan, Seshadri, Machiraju, Faloutsos ’10; Hooi, Song, Beutel, Shah, Shin, Faloutsos ’16]

Community and data mining

[Kleinberg ’99; Ma, Fang, Cheng, Lakshmanan, Zhang, Lin ’21; Chen, Liu, Zhou, Liao, Xu, Li ’23]

Computational biology

[Junker, Schreiber ’08; Saha, Hoch, Khuller, Raschid, Zhang ’10]

Computation of densest subgraph: semi-streaming

Undirected DS

Approx.	Complexity (passes)	Reference
2+ε	$O(\log n)$	Bahmani, Kumar, Vassilvitskii '12
1+ε	1	Esfandiari, Hajiaghayi, Woodruff '16

Directed DS

Approx.	Complexity (passes)	Reference
2+ε	$O(\log n)$	Bahmani, Kumar, Vassilvitskii '12
2+ε	1 **random order**	Mitrović, Pan '24
$\log n$	1	Mitrović, Pan, Qaempanah, Raeisi '25
1+ε	1 $n^{1.5}$ **memory**	Esfandiari, Hajiaghayi, Woodruff '16

DDS captures directed relations, but constant-pass guarantees are weaker.

Computation of densest subgraph: MPC

Undirected DS

Approx.	Complexity (rounds)	Reference
1+ε	$O(\log n)$	Bahmani, Goel, Munagala ’14
1+ε	$O\left(\sqrt{\log n}\right)$	Ghaffari, Lattanzi, Mitrović ’19

Directed DS

Approx.	Complexity (rounds)	Reference
1+ε	$O(\log n)$	Bahmani, Goel, Munagala ’14
2+ε	$O\left(\sqrt{\log n}\right)$	Mitrović, Pan ’24 Mitrović, Pan, Qaempanah, Raeisi ’25

A similar gap appears in near-linear-memory MPC.

Directed densest subgraph: three algorithmic lenses

Peeling

- remove low-degree vertices
- keep the best surviving subgraph
- often needs $\Theta(\log n)$ iterations

[Charikar ‘00;
Bahmani, Kumar, Vassilvitskii ‘12]

LP

Sampling (Sparsification)

Directed densest subgraph: three algorithmic lenses

Peeling

- remove low-degree vertices
- keep the best surviving subgraph
- often needs $\Theta(\log n)$ iterations

[Charikar ‘00;
Bahmani, Kumar, Vassilvitskii ‘12]

LP

- solve or approximate the DS LP
- strong approximation guarantees
- less direct with sublinear working memory

[Charikar ‘00;
Bahmani, Goel, Munagala ‘14]

Sampling (Sparsification)

Directed densest subgraph: three algorithmic lenses

Peeling

- remove low-degree vertices
- keep the best surviving subgraph
- often needs $\Theta(\log n)$ iterations

[Charikar '00;
Bahmani, Kumar, Vassilvitskii '12]

LP

- solve or approximate the DS LP
- strong approximation guarantees
- less direct with sublinear working memory

[Charikar '00;
Bahmani, Goel, Munagala '14]

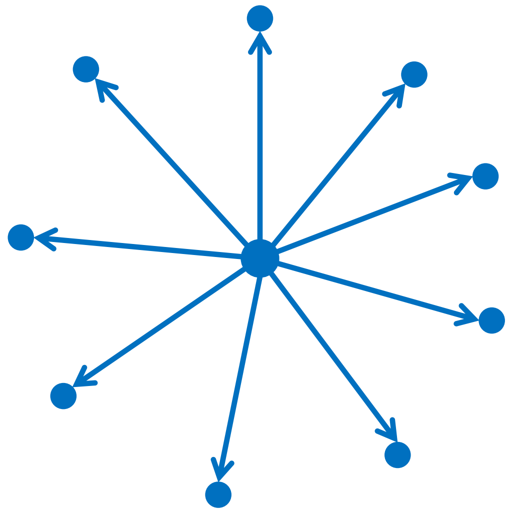
Sampling (Sparsification)

- sample edges to preserve a near-optimal DDS
- works well for undirected DS
- uniform DDS sampling can be large

[Esfandiari, Hajiaghayi, Woodruff '16;
Mitrović, Pan '24]

Limitation of uniform sparsification

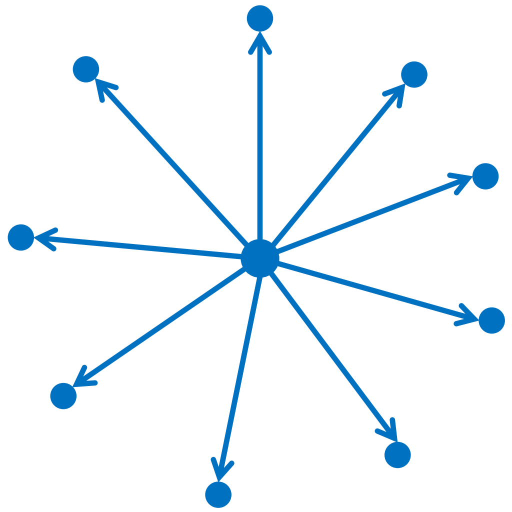
Star: DDS



Limitation of uniform sparsification

Star: DDS

T = leaves
S = the center

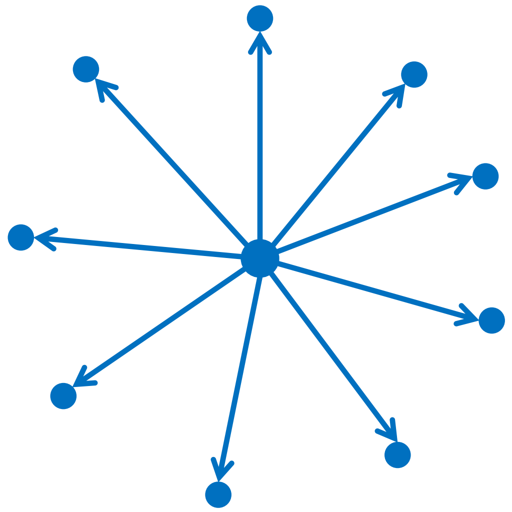


$$\rho(S,T) = \frac{n-1}{\sqrt{(n-1)\cdot 1}} = \sqrt{n-1}$$

Limitation of uniform sparsification

Star: DDS

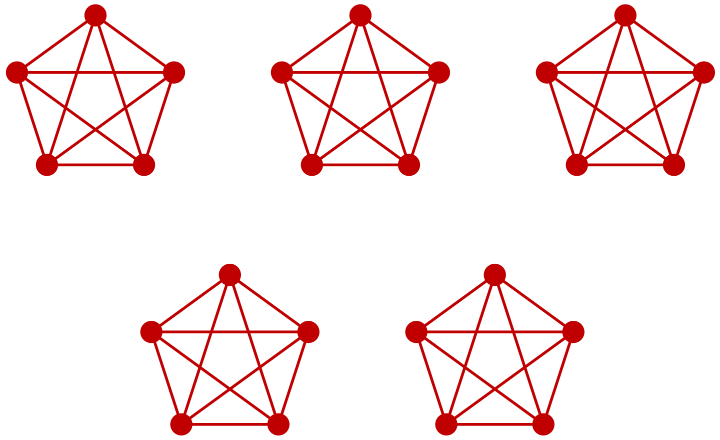
T = leaves
S = the center



$$\rho(S,T) = \frac{n-1}{\sqrt{(n-1)\cdot 1}} = \sqrt{n-1}$$

+

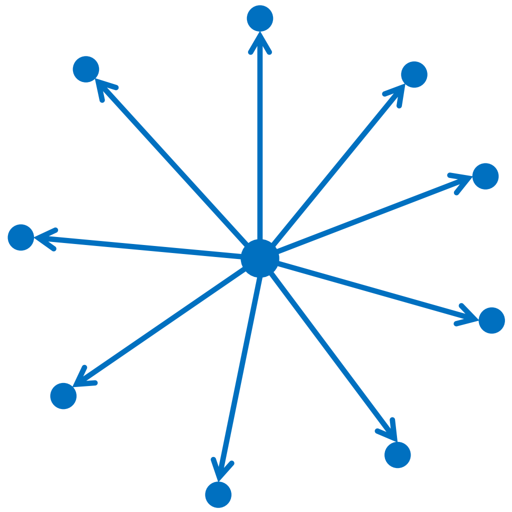
Many dense distractors: $n^{0.5+\beta}$ copies of $n^{0.5-\beta}$ cliques



$\Theta(n)$ vertices; $\Theta\left(n^{1.5-\beta}\right)$ edges

Limitation of uniform sparsification

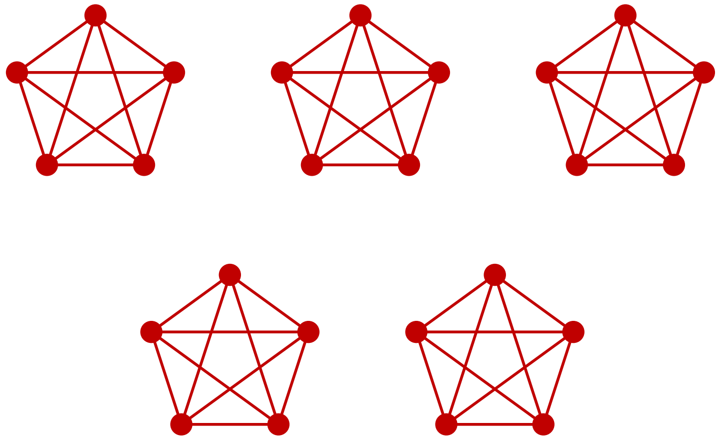
Star: DDS



needs $\Theta(n)$ sampled star edges

+

Many dense distractors: $n^{0.5+\beta}$ copies of $n^{0.5-\beta}$ cliques



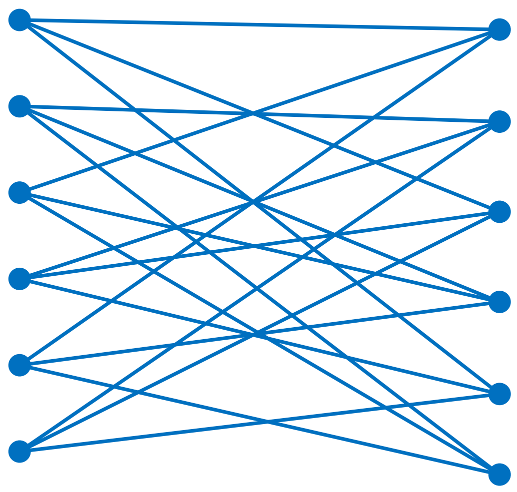
$\Theta(n)$ vertices; $\Theta\left(n^{1.5-\beta}\right)$ edges

Uniform sampling must be large enough to see the star: $\boldsymbol{\Omega(n\sqrt{n})}$ sampled edges in the worst case.

Limitation of peeling

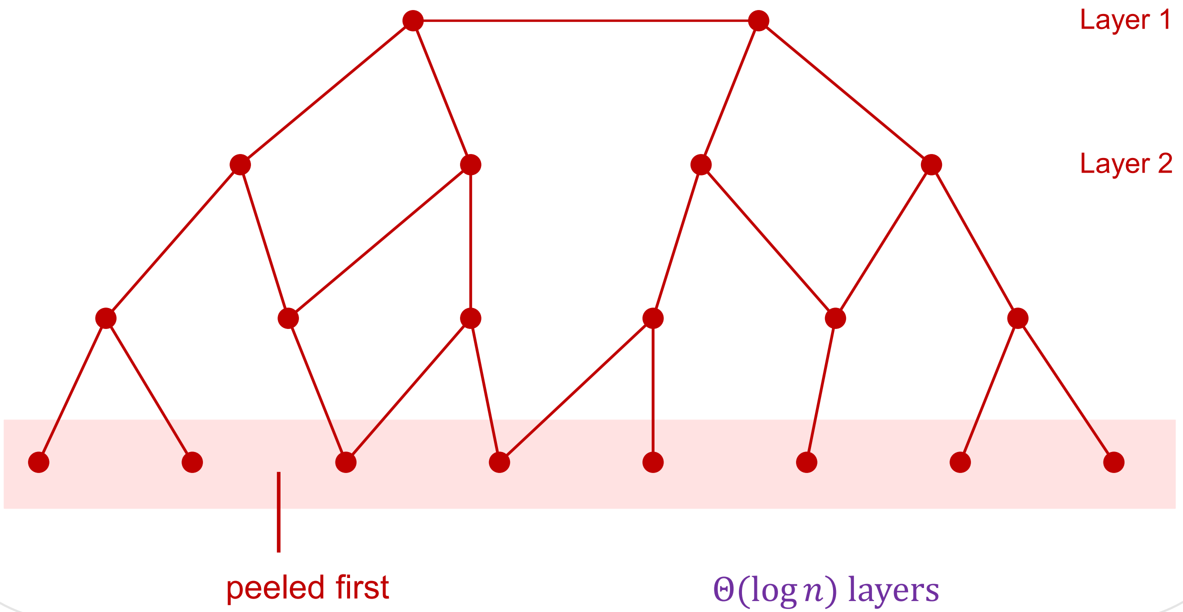
Peeling rule: Remove vertices of degree < D.

(D,D)-biregular DS



density D; all vertices have degree D

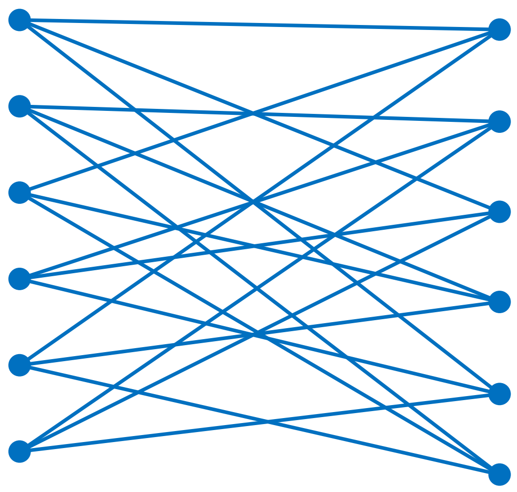
Layered graph: degrees hide the obstruction



Limitation of peeling

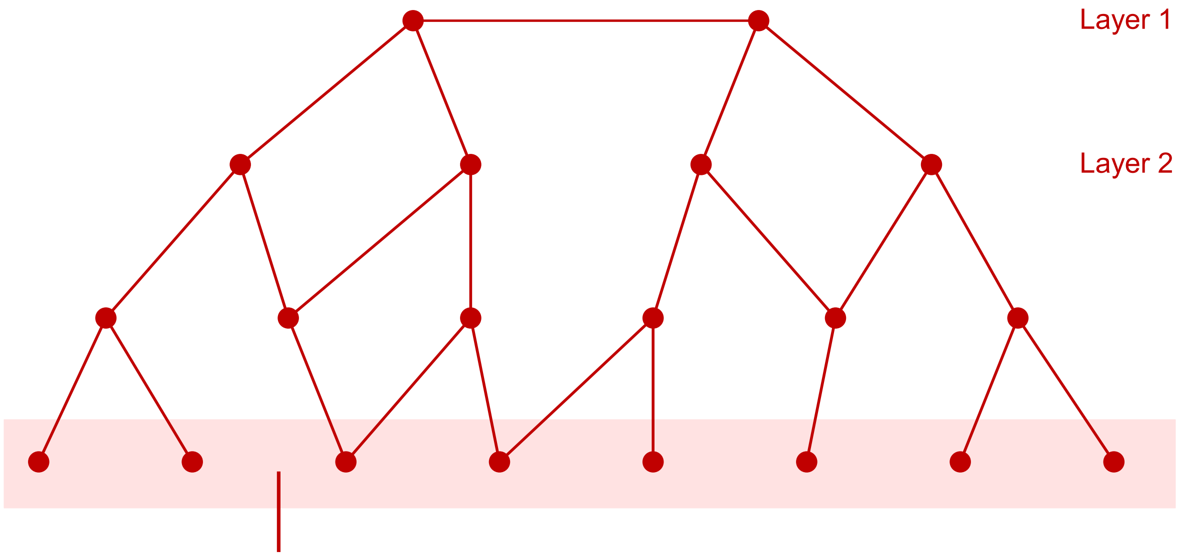
Peeling rule: Remove vertices of degree < D.

(D,D)-biregular DS



density D; all vertices have degree D

Layered graph: degrees hide the obstruction



Any degree-based peeling that preserves the DS removes only one layer per iteration
$\Rightarrow \Theta(\log n)$ iterations.

Take away message

Uniform sparsification

needs $\Omega(n\sqrt{n})$ space
in the worst case

Peeling

needs $\Theta(\log n)$ passes
in the worst case

Take away message

Uniform sparsification

needs $\Omega(n\sqrt{n})$ space
in the worst case

Peeling

needs $\Theta(\log n)$ passes
in the worst case

Can we go beyond the $\Theta(\log n)$-pass
semi-streaming barrier for DDS?

$= O(n)$ space

Our result:

one low-degree peeling step + refined sampling
$\Rightarrow \tilde{O}(n)$ edges preserving a near-optimal DDS

* Peeling and sparsification appear throughout memory-constrained graph algorithms.

Simplify the setup.
Some lies will be said …

"The truth is rarely pure and never simple."

Simplify the setup.

Some lies will be said …

"The truth is rarely pure and never simple."

Let (S^*, T^*) be an optimal DDS.

$D = \rho(S^*, T^*)$ is its density.

imbalance parameter

$$z = \sqrt{\frac{|S^*|}{|T^*|}}$$

balanced vs. one-sided optimum

Assume $z \geq 1$.

Assume D and z are known.

(They can be guessed.)

Idea 1: refined sampling

$n^{1.5}$ **edges**; Esfandiari, Hajiaghayi, Woodruff '16

Idea 1: refined sampling

$n^{1.5}$ **edges**; Esfandiari, Hajiaghayi, Woodruff '16

Use the imbalance parameter z when sampling.

Sampling rule

$$p = \frac{z}{D}$$

- sample each edge with prob p
- recover DDS from the sample

Idea 1: refined sampling

$n^{1.5}$ **edges**; Esfandiari, Hajiaghayi, Woodruff ’16

Use the imbalance parameter z when sampling.

Sampling rule

$$p = \frac{z}{D}$$

- sample each edge with prob p
- recover DDS from the sample

What is preserved?

not hard to show:

the sample preserves
a near-optimal DDS

Idea 1: refined sampling

$n^{1.5}$ **edges**; Esfandiari, Hajiaghayi, Woodruff '16

Use the imbalance parameter z when sampling.

Sampling rule

$$p = \frac{z}{D}$$

- sample each edge with prob p
- recover DDS from the sample

What is preserved?

not hard to show:

the sample preserves a near-optimal DDS

Expected size

$$|E(V,V)| \leq D \cdot \sqrt{|V|^2} = Dn$$

$$\frac{|E(S,T)|}{\sqrt{|S| \cdot |T|}} \leq D$$

Idea 1: refined sampling

$n^{1.5}$ **edges**; Esfandiari, Hajiaghayi, Woodruff '16

Use the imbalance parameter z when sampling.

Sampling rule

$$p = \frac{z}{D}$$

- sample each edge with prob p
- recover DDS from the sample

What is preserved?

not hard to show:

the sample preserves a near-optimal DDS

Expected size

$$|E(V,V)| \leq D \cdot \sqrt{|V|^2} = Dn$$

Expected sample size

$$\leq Dn \cdot p = Dn \cdot (z/D)$$

$$= nz$$

Idea 1: refined sampling

$n^{1.5}$ **edges**; Esfandiari, Hajiaghayi, Woodruff '16

Use the imbalance parameter z when sampling.

Sampling rule

$$p = \frac{z}{D}$$

- sample each edge with prob p
- recover DDS from the sample

What is preserved?

not hard to show:

the sample preserves a near-optimal DDS

Expected size

$$|E(V,V)| \leq D \cdot \sqrt{|V|^2} = Dn$$

Expected sample size

$$\leq Dn \cdot p = Dn \cdot (z/D)$$

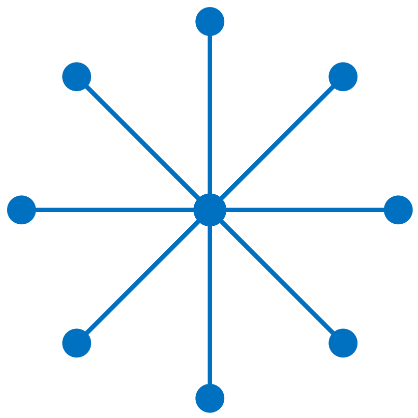
$$= nz$$

Useful when z is moderate;
still costly when z is large.

Refined sampling alone is not enough

For the star obstruction, $z = \sqrt{n}$, so $O(nz) = O(n\sqrt{n})$.

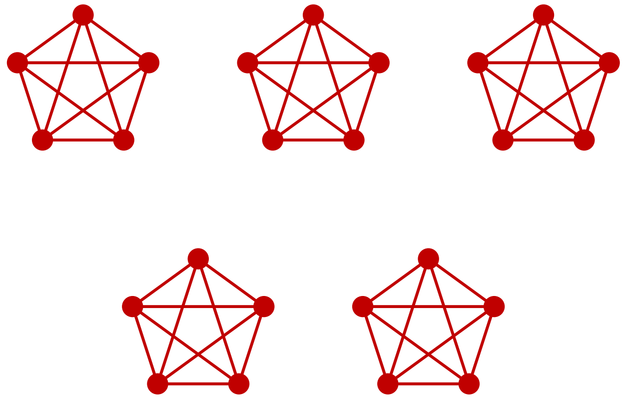
Star: DDS



must sample many star edges

+

Many dense distractors

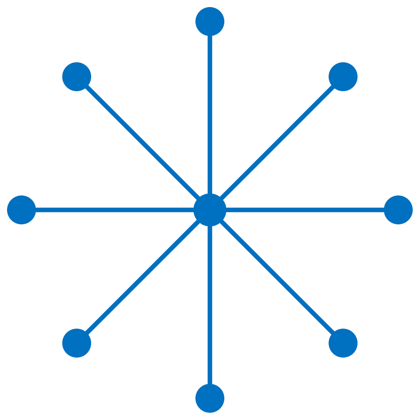


large total edge mass

Refined sampling alone is not enough

For the star obstruction, $z = \sqrt{n}$, so $O(nz) = O(n\sqrt{n})$.

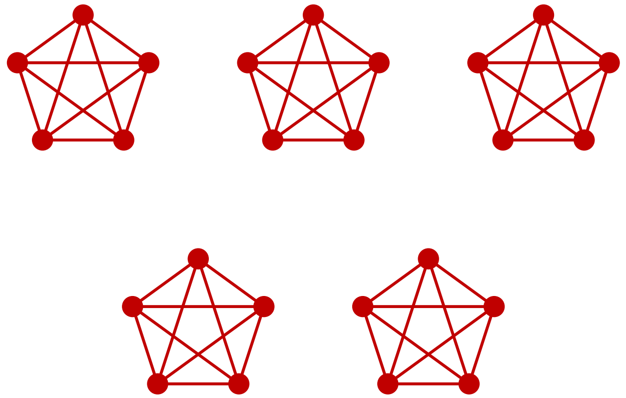
Star: DDS



must sample many star edges

+

Many dense distractors



large total edge mass

The obstruction remains:
sampling must still see the star.

Idea 2: the first peeling step is special

$$\frac{|S^*|}{|T^*|} = z^2$$

Idea 2: the first peeling step is special

$$\frac{|S^*|}{|T^*|} = z^2$$

Before peeling

$$\frac{|S|}{|T|} = 1$$

The input starts as two equal copies of V.

Idea 2: the first peeling step is special

$$\frac{|S^*|}{|T^*|} = z^2$$

Before peeling

$$\frac{|S|}{|T|} = 1$$

The input starts as two equal copies of V.

Peeling step

$$k_T = \frac{Dz}{2}$$

Remove vertices v in T with degree < k_T

Idea 2: the first peeling step is special

$$\frac{|S^*|}{|T^*|} = z^2$$

Before peeling

$$\frac{|S|}{|T|} = 1$$

The input starts as two equal copies of V.

Peeling step

$$k_T = \frac{Dz}{2}$$

$$Dz = \frac{E(S^*, T^*)}{\sqrt{|S^*| \cdot |T^*|}} \cdot \sqrt{\frac{|S^*|}{|T^*|}} = \frac{\boldsymbol{E(S^*, T^*)}}{\boldsymbol{|T^*|}}$$

Remove vertices v in T with degree < k_T

Idea 2: the first peeling step is special

$$\frac{|S^*|}{|T^*|} = z^2$$

Before peeling

$$\frac{|S|}{|T|} = 1$$

The input starts as two equal copies of V.

Peeling step

$$k_T = \frac{Dz}{2}$$

Remove vertices v in T with degree < k_T

Case 1. $|\mathrm{T}| \leq 10 \cdot \frac{|S|}{z^2}$

Idea 2: the first peeling step is special

$$\frac{|S^*|}{|T^*|} = z^2$$

Before peeling

$$\frac{|S|}{|T|} = 1$$

The input starts as two equal copies of V.

Peeling step

$$k_T = \frac{Dz}{2}$$

Remove vertices v in T with degree < k_T

Case 1. $|\mathrm{T}| \leq 10 \cdot \frac{|S|}{z^2}$

$$\begin{aligned} |E(S,T)| &\leq D\sqrt{|S| \cdot |T|} \\ &\leq D\sqrt{|S| \cdot \frac{10|S|}{z^2}} \\ &= O\left(\frac{D|S|}{z}\right) \\ &\leq O\left(\frac{nD}{z}\right) \end{aligned}$$

Idea 2: the first peeling step is special

$$\frac{|S^*|}{|T^*|} = z^2$$

Before peeling

$$\frac{|S|}{|T|} = 1$$

The input starts as two equal copies of V.

Peeling step

$$k_T = \frac{Dz}{2}$$

Remove vertices v in T with degree < k_T

Case 1. $|T| \leq 10 \cdot \frac{|S|}{z^2}$

$$\begin{aligned} |E(S,T)| &\leq D\sqrt{|S| \cdot |T|} \\ &\leq D\sqrt{|S| \cdot \frac{10|S|}{z^2}} \\ &= O\left(\frac{D|S|}{z}\right) \\ &\leq O\left(\frac{nD}{z}\right) \end{aligned}$$

✓ Done: sample remaining edges with $p = \frac{z}{D}$

Expected sample size = $O(n)$

Idea 2: the first peeling step is special

Case 2. $|T| > 10 \cdot \frac{|S|}{z^2}$

After peeling

each remaining vertex v in T has degree at least

$$k_T = \frac{Dz}{2}$$

Idea 2: the first peeling step is special

Case 2. $|\mathrm{T}| > 10 \cdot \frac{|S|}{z^2}$

After peeling

each remaining vertex v in T has degree at least

$$k_T = \frac{Dz}{2}$$

many vertices remain, **and**
each has large degree

Idea 2: the first peeling step is special

Case 2. $|\mathrm{T}| > 10 \cdot \frac{|S|}{z^2}$

After peeling

each remaining vertex v in T has degree at least

$$k_T = \frac{Dz}{2}$$

many vertices remain, **and**
each has large degree

Density lower bound

$$|E(S,T)| \geq k_T\,|T|$$

Idea 2: the first peeling step is special

Case 2. $|\mathrm{T}| > 10 \cdot \frac{|S|}{z^2}$

After peeling

each remaining vertex v in T has degree at least

$$k_T = \frac{Dz}{2}$$

many vertices remain, **and**
each has large degree

Density lower bound

$$|E(S,T)| \geq k_T\ |T|$$

$$\rho(S,T) = \frac{|E(S,T)|}{\sqrt{|S| \cdot |T|}} \geq\ k_T \cdot \sqrt{\frac{|T|}{|S|}}$$

Idea 2: the first peeling step is special

Case 2. $|\mathrm{T}| > 10 \cdot \frac{|S|}{z^2}$

After peeling

each remaining vertex v in T has degree at least

$$k_T = \frac{Dz}{2}$$

many vertices remain, **and**
each has large degree

Density lower bound

$$|E(S,T)| \geq k_T\ |T|$$

$$\rho(S,T) = \frac{|E(S,T)|}{\sqrt{|S| \cdot |T|}} \geq k_T \cdot \sqrt{\frac{|T|}{|S|}}$$

$$= \frac{Dz}{2} \cdot \sqrt{\frac{|T|}{|S|}} > \frac{Dz}{2} \cdot \frac{\sqrt{10}}{z}$$

Idea 2: the first peeling step is special

Case 2. $|T| > 10 \cdot \frac{|S|}{z^2}$

After peeling

each remaining vertex v in T has degree at least

$$k_T = \frac{Dz}{2}$$

many vertices remain, **and**
each has large degree

Density lower bound

$$|E(S,T)| \geq k_T\ |T|$$

$$\rho(S,T) = \frac{|E(S,T)|}{\sqrt{|S| \cdot |T|}} \geq k_T \cdot \sqrt{\frac{|T|}{|S|}}$$

$$= \frac{Dz}{2} \cdot \sqrt{\frac{|T|}{|S|}} > \frac{Dz}{2} \cdot \frac{\sqrt{10}}{z}$$

$$\geq D$$

Idea 2: the first peeling step is special

Case 2. $|T| > 10 \cdot \frac{|S|}{z^2}$

After peeling

each remaining vertex v in T has degree at least

$$k_T = \frac{Dz}{2}$$

many vertices remain, **and** each has large degree

Density lower bound

$$|E(S,T)| \geq k_T\, |T|$$

$$\rho(S,T) = \frac{|E(S,T)|}{\sqrt{|S| \cdot |T|}} \geq k_T \cdot \sqrt{\frac{|T|}{|S|}}$$

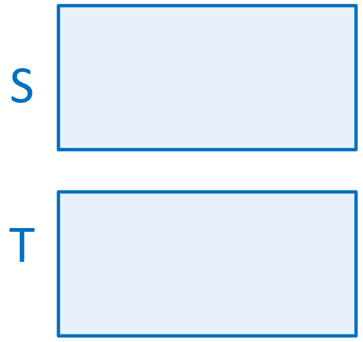
$$= \frac{Dz}{2} \cdot \sqrt{\frac{|T|}{|S|}} > \frac{Dz}{2} \cdot \frac{\sqrt{10}}{z}$$

$$\geq D$$

✓ dense enough: return (S,T)

Summary (Algorithm)

Initially



$$\frac{|S|}{|T|} = 1$$

Summary (Algorithm)

Initially

S

T

$$\frac{|S|}{|T|} = 1$$

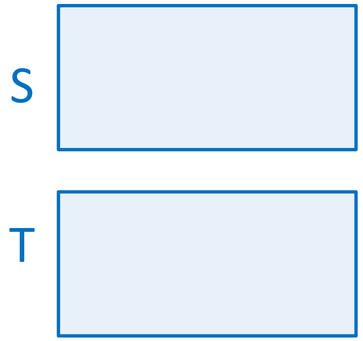
Peeling step

remove v in T
with degree < k_T

$$k_T = \frac{Dz}{2}$$

Summary (Algorithm)

Initially



$$\frac{|S|}{|T|} = 1$$

Peeling step

remove v in T
with degree < k_T

$$k_T = \frac{Dz}{2}$$

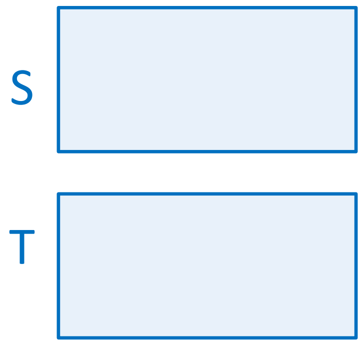
Case 1. T drops

sample with $p = \frac{z}{D}$

✓ O(n) edges

Summary (Algorithm)

Initially



$$\frac{|S|}{|T|} = 1$$

Peeling step

remove v in T with degree < k_T

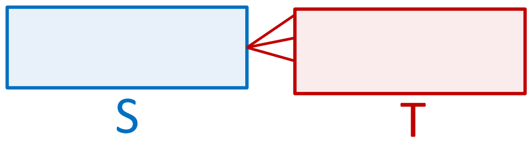
$$k_T = \frac{Dz}{2}$$

Case 1. T drops

sample with $p = \frac{z}{D}$

✓ O(n) edges

Case 2. T does not drop



many vertices remain

each has degree $\geq k_T$

✓ dense enough

Applications

Massively Parallel Computation

Algorithm:

1. One peeling step

O(1) rounds

Massively Parallel Computation

Algorithm:

1. One peeling step — O(1) rounds

2. Sparsification — O(1) rounds

Massively Parallel Computation

Algorithm:

1. One peeling step — O(1) rounds
2. Sparsification — O(1) rounds
3. Gather the sparsified graph on one machine and solve DDS locally — O(1) rounds

Massively Parallel Computation

Algorithm:

1. One peeling step — O(1) rounds
2. Sparsification — O(1) rounds
3. Gather the sparsified graph on one machine and solve DDS locally — O(1) rounds

Round complexity: $O(\sqrt{\log n})$ to $O(1)$ in near-linear-memory MPC

Semi-streaming: two passes

Peeling and sampling can be **separated**.

Algorithm:

1. First pass — compute degrees; mark vertices below the peeling threshold
2. Second pass — sample edges from the remaining graph

Direct implementation of Peel-Sample in two passes.

Semi-streaming: one pass

Peeling and sampling must happen **together**.

Obstacle

In the middle of the stream,
we do not know which vertices
should be peeled.

Semi-streaming: one pass

Peeling and sampling must happen **together**.

Obstacle

In the middle of the stream,
we do not know which vertices
should be peeled.

Idea: **ignore** first $k_T = \frac{Dz}{2}$ edges at v; then sample later edges with $p = \frac{z}{D}$.

This is not literal peeling; some DDS edges may be lost, but the loss is controlled.

Semi-streaming: one pass

Peeling and sampling must happen **together**.

Obstacle

In the middle of the stream,
we do not know which vertices
should be peeled.

Idea: **ignore** first $k_T = \frac{Dz}{2}$ edges at v; then sample later edges with $p = \frac{z}{D}$.

This is not literal peeling; some DDS edges may be lost, but the loss is controlled.

Pass complexity: $O(\log n)$ to 1

Sublinear-time

High level: similar to MPC.

- Mark low-degree vertices in O(n) time -- with adjacency lists, just check deg(v).

Sublinear-time

High level: similar to MPC.

- Mark low-degree vertices in O(n) time -- with adjacency lists, just check deg(v).

Non-trivial part

How do we sample edges whose endpoints are both unmarked?

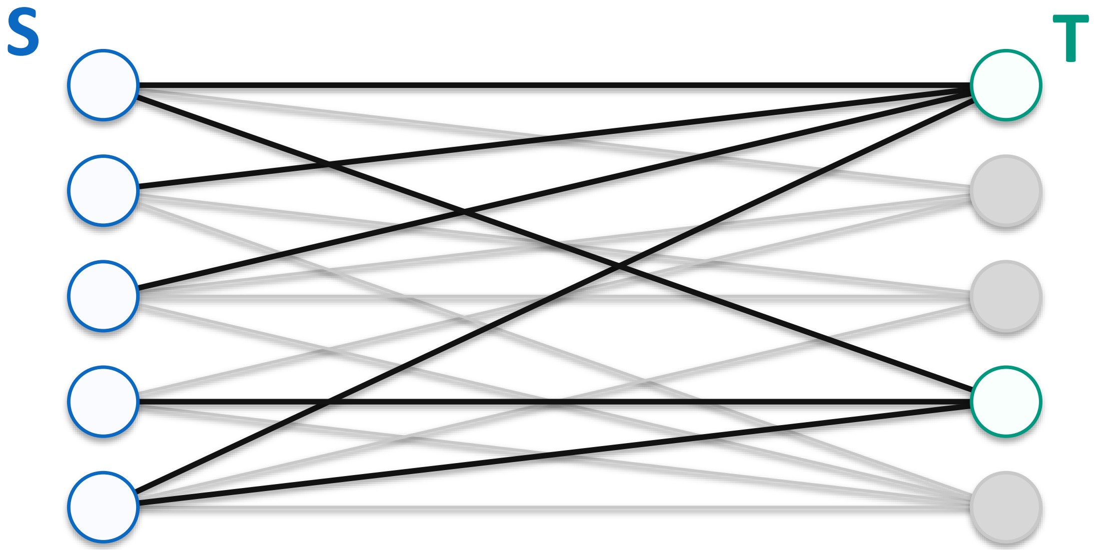
Sublinear-time

High level: similar to MPC.

- Mark low-degree vertices in O(n) time -- with adjacency lists, just check deg(v).

Non-trivial part

How do we sample edges whose endpoints are both unmarked?



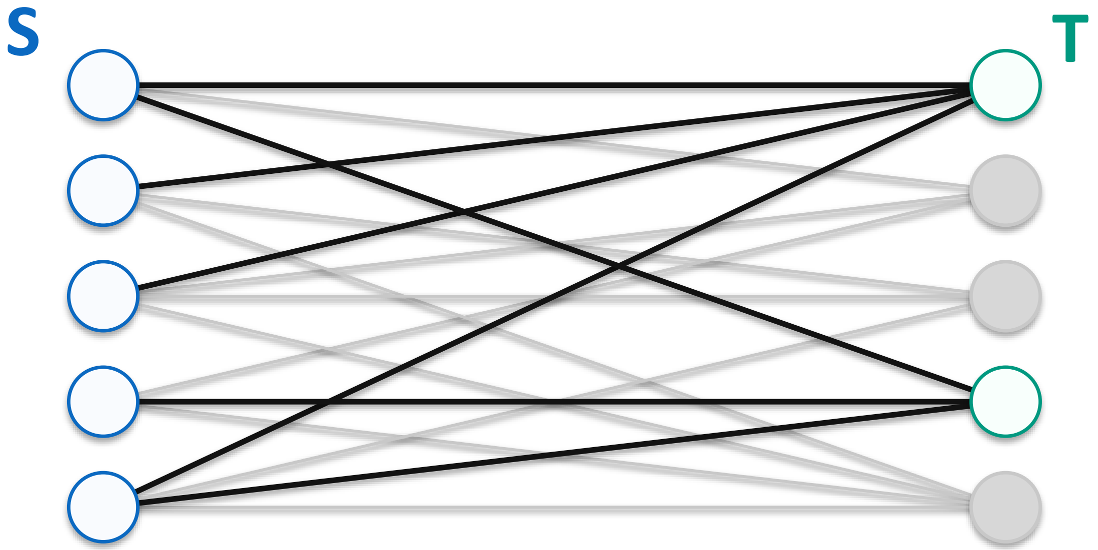
Sublinear-time

High level: similar to MPC.

- Mark low-degree vertices in O(n) time -- with adjacency lists, just check deg(v).

Non-trivial part

How do we sample edges whose endpoints are both unmarked?



Observation: view the graph as bipartite and peel **only** from **T**.
Sample neighbors of the **remaining** vertices in **T**.

Open questions

Other problems

Undirected DS in $poly \log \log n$ MPC rounds.

(log log n approximate orientation in poly log log n rounds [Ghaffari, Grunau '25])