Flow-Cut Gaps and Hardness of Directed Cut Problems

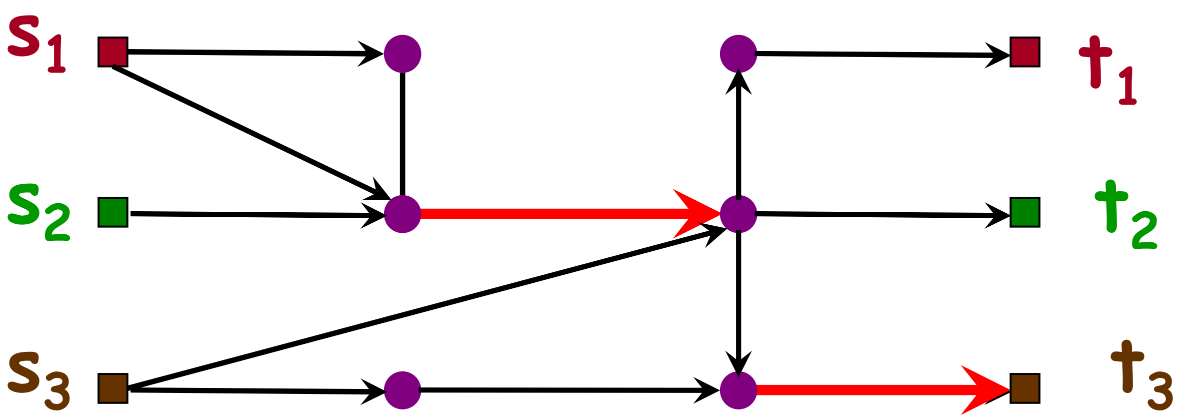
Sanjeev Khanna
University of Pennsylvania

Joint work with: Julia Chuzhoy (TTI)

Minimum Multicut

Input: A graph (directed or undirected) and a collection of k source-sink pairs $(s_1,t_1),\ldots(s_k,t_k)$.

Goal: Find a minimum-size subset of edges whose removal disconnects all s_i-t_i pairs.



Solution cost: 2

Minimum Multicut

- k=1: Minimum s-t cut problem, solvable in polynomial time [Ford, Fulkerson '56].
- k=2: Solvable in polynomial time for undirected graphs [Yannakakis, Kanellakis, Cosmadakis, Papadimitriou '83], but NP-hard for directed graphs [Garg, Vazirani, Yannakakis '94].
- $k \geq 3$: NP-hard for directed and undirected graphs [Dahlhaus, Johnson, Papadimitriou, Seymour, Yannakakis '94].
- Arbitrary k: NP-hard even on undirected star graphs [Garg, Vazirani, Yannakakis '93].

An Integer Program

- For each edge e, a 0/1 indicator variable x_e: x_e is 1 if e is in the solution and 0 otherwise.

- For each source-sink pair s_i-t_i, let P_i be the set of all the paths connecting s_i to t_i.

Constraint: For each path $p \in P_i$, $x_e = 1$ for some $e \in p$.

Goal: Minimize $\sum_e x_e$.

An LP Relaxation

$$\text{Min } \sum_e x_e$$

s.t.

$\forall\, i \in [1..k],\ \forall\, s_i \to t_i$ paths p

$$\sum_{e \in p} x_e \geq 1$$

$\forall\, e \in E \qquad\qquad 0 \leq x_e \leq 1\,.$

Constraint: Assign length to edges such that any source-sink path has length ≥ 1.

Goal: Minimize total length assigned to edges.

Rounding for a Single s-t Pair (k=1)

- Choose $r \in (0,1)$ uniformly at random.
- S = Vertices within distance r from source s.
- Output the cut $(S, V/S)$.

Any edge $e=(u,v)$ belongs to the cut with probability:

$$|dist(s,u) - dist(s,v)| \leq x_e$$

Expected solution cost is $\sum_e x_e = OPT_{LP}$

Arbitrary # of Pairs: Undirected Graphs

Lower bound: $\Omega(\log n)$ [Leighton, Rao '88].

Upper bound: $O(\log n)$ [Garg, Vazirani, Yannakakis '93].

Arbitrary # of Pairs: Directed Graphs

Lower bounds:

- $\Omega(\log n)$.
- $\Omega(k)$ [Saks, Samorodnitsky, Zosin '04].

But $k = O(\log n / \log\log n)$!

Upper bound: $O(n^{11/23})$ [Agarwal, Alon, Charikar' 07]

(Improves $O(n^{1/2})$ bound of [Cheriyan, Karloff, Rabani' 01], [Gupta '03].)

Integrality gap of directed multicut relaxation?

The Cut and Flow Duality

$$\text{Min} \sum_e x_e$$

s.t.

$\forall i \in [1..k], \forall p \in P_i$ $\quad \sum_{e \in P} x_e \geq 1$

$\forall e \in E$ $\quad x_e \geq 0$

Minimum Fractional Multicut

$$\text{Max} \sum_{i,\, p \in P_i} f_p$$

s.t.

$\forall i \in [1..k], \forall p \in P_i$ $\quad f_p \geq 0$

$\forall e \in E$ $\quad \sum_{p:\, e \in p} f_p \leq 1$

Maximum Multicommodity Flow

Flow-Cut Gaps

Max Multicommodity Flow = Min Fractional Multicut

- Integrality gap of the Multicut LP = Gap between Max Multicommodity Flow and Minimum Integral Multicut = Flow-Cut Gap.

- Best known approximation guarantees for many problems are linked to flow-cut gaps.
  - Multicut
  - Well-linked decompositions
  - Oblivious routing

Our Results

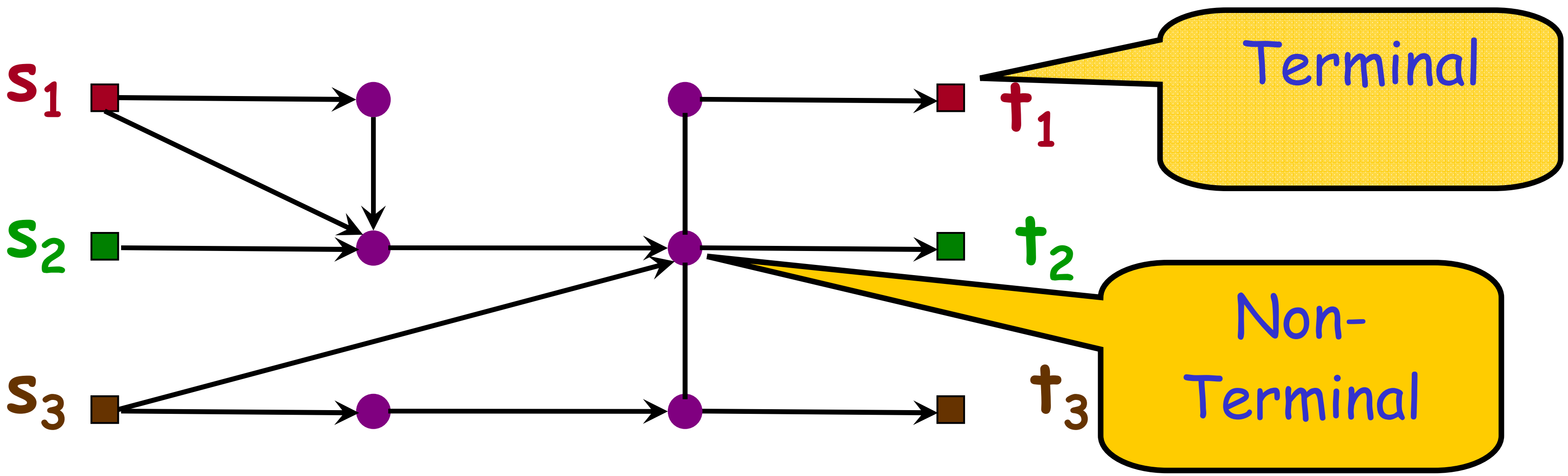
- Flow-cut gap for directed multicut is $\Omega(n^{1/7})$.
  - Improves upon the previous $\Omega(\log n)$ gap.
- An $\Omega(n^{1/7})$ gap between directed sparsest cut and concurrent multicommodity flow.
  - Improves upon the previous $\Omega(\log n)$ gap.
- For any $\varepsilon > 0$, a $2^{\log^{1-\varepsilon} n}$-hardness for directed multicut and directed sparsest cut.
  - Improves upon earlier $\Omega(\log n / \log \log n)$-hardness.

The Multicut Integrality Gap Construction

Vertex Version of Directed Multicut

Input: Same as before.

Goal: Smallest set of non-terminal vertices whose removal disconnects all source-sink pairs.



Integrality Gap Construction

- A parameter n.
- N = Total # of vertices = $O(n^7)$.
- $L \approx n/\log n \approx N^{1/7}$.

A multicut instance where:

(1) # of vertices on any source-sink path is at least L.

(2) Cost of any integral solution is $\Omega(N)$.

Fractional Cost: $O(N/L)$. Integrality Gap $\approx \Omega(N^{1/7})$.

Overview

Step One: A multicut instance H that

- satisfies property (2), and
- satisfies property (1) on a restricted class of paths: the canonical paths are $\Omega(L)$ long.

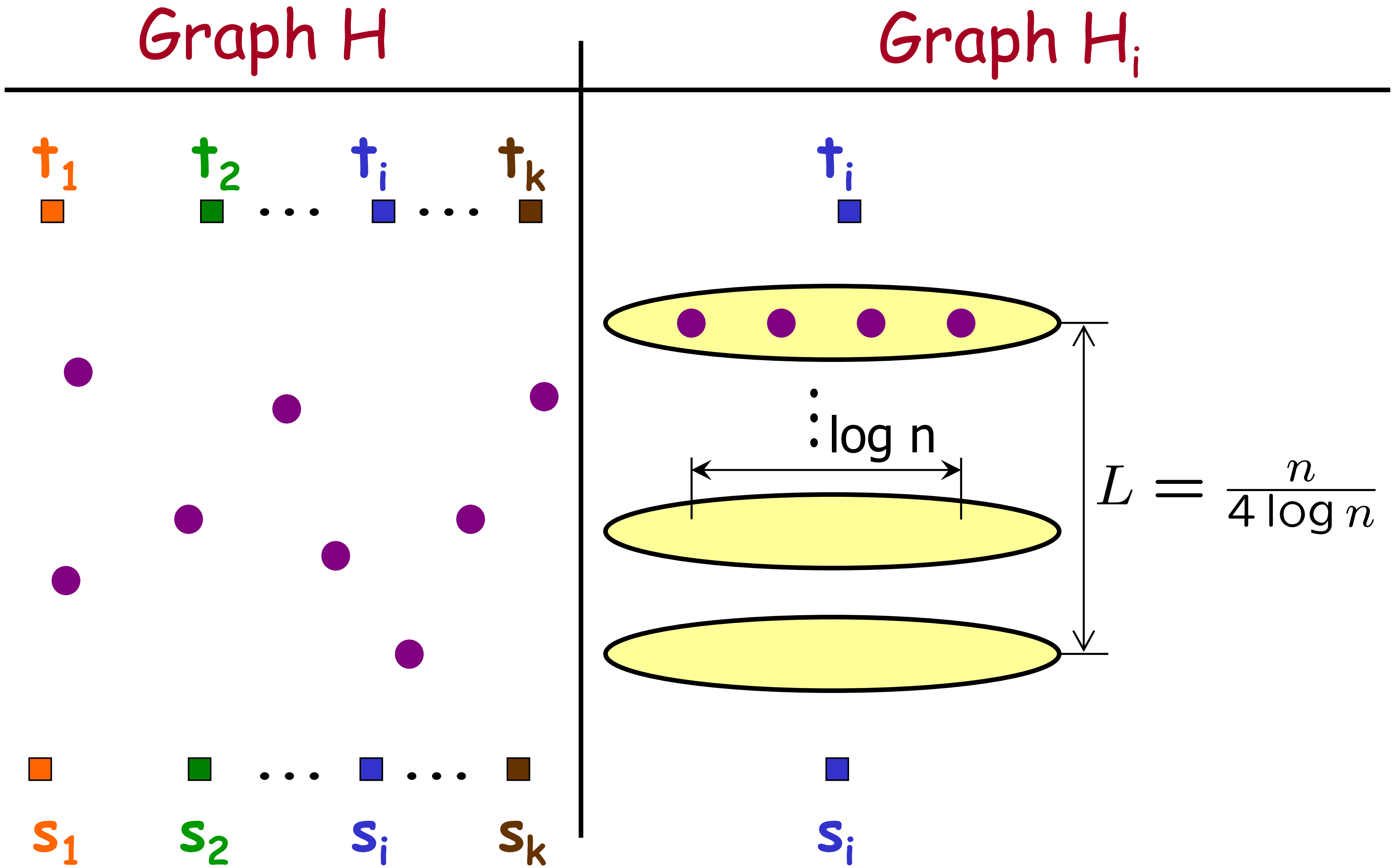
Step Two: An instance B based on a labeling scheme such that only possible source-sink paths are the canonical paths. But these paths may be short.

Step Three: Compose H and B:

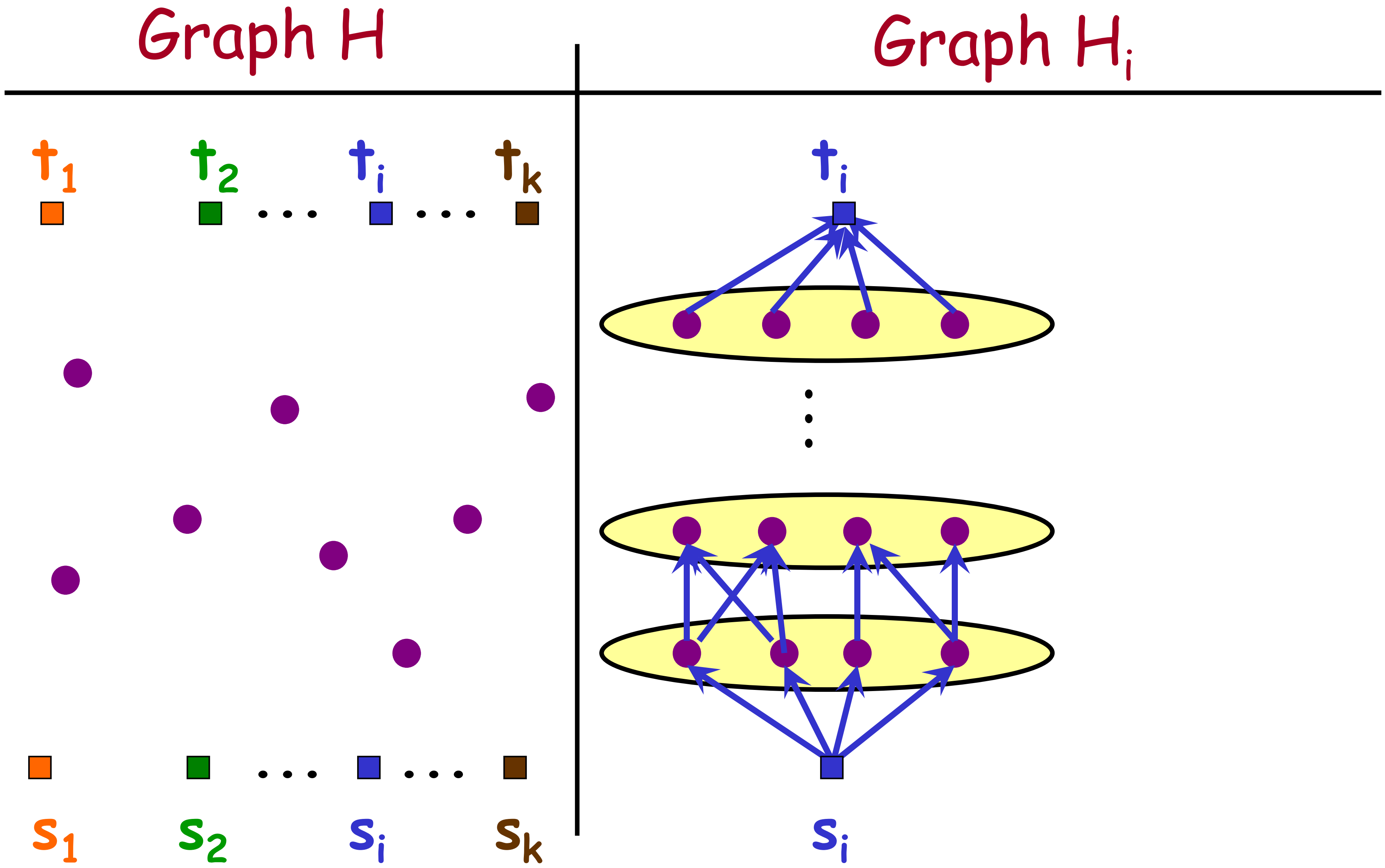
- Only long canonical paths (1).
- Any integral solution has cost $\Omega(N)$ (2).

Step One: The Graph H

- H is a union of k graphs, H_1, H_2, …, H_k
- All graphs H_i share the same set of non-terminal vertices, say, {1,…,n}.
- Each graph H_i has exactly one source-sink pair s_i-t_i.

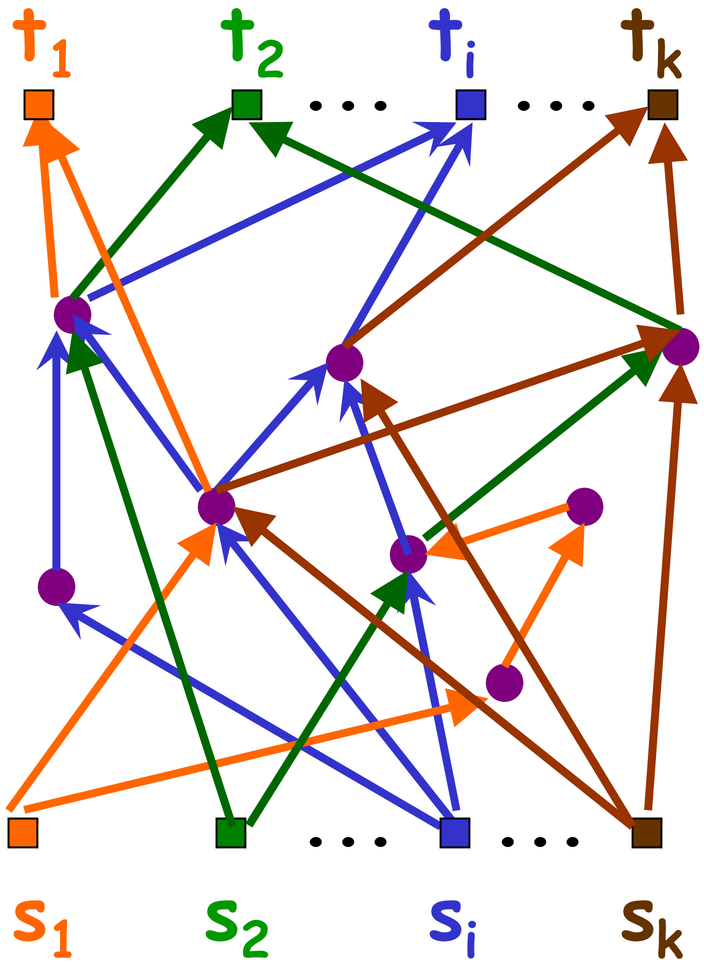


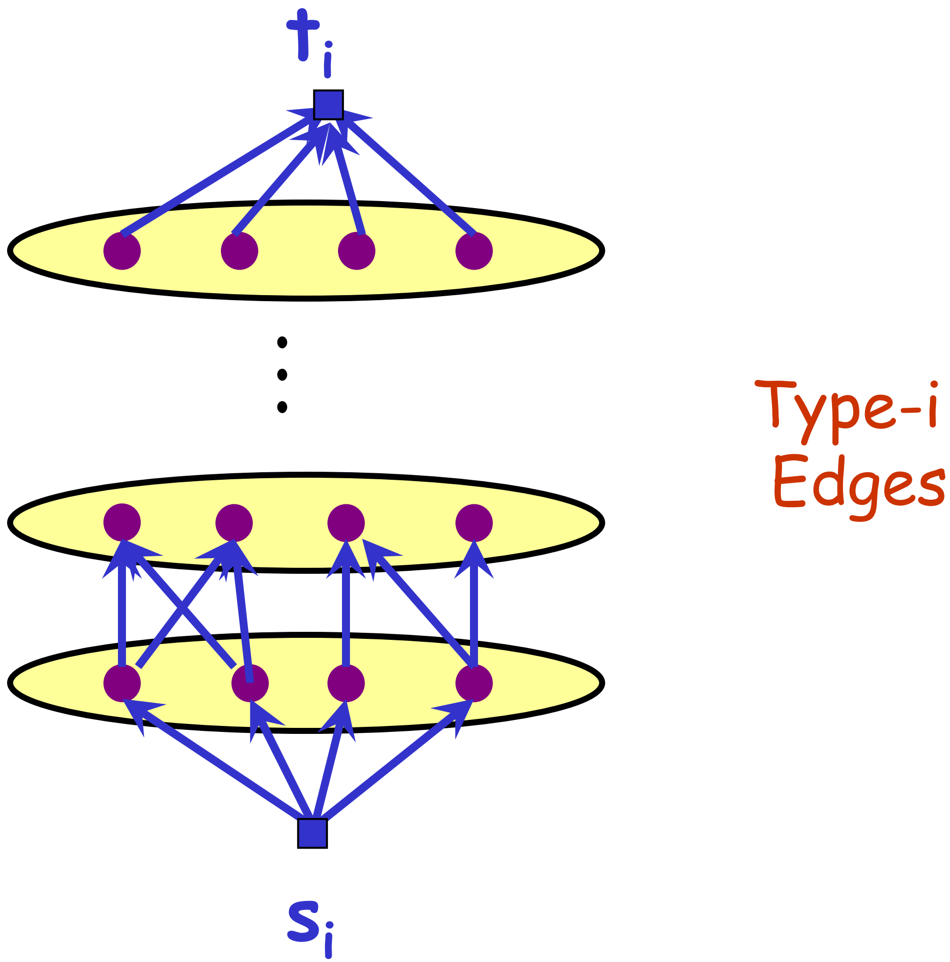
Graph H
Graph H_i
t_1 t_2 t_i t_k
s_1 s_2 s_i s_k
t_i
log n
$L = \frac{n}{4 \log n}$
s_i



Graph H
Graph H_i
t_1 t_2 t_i t_k
s_1 s_2 s_i s_k
t_i
s_i

Graph H	Graph H_i





Properties of Graph H

What we need?

(1) Any source-sink path has length $\geq L = n/(4 \log n)$.

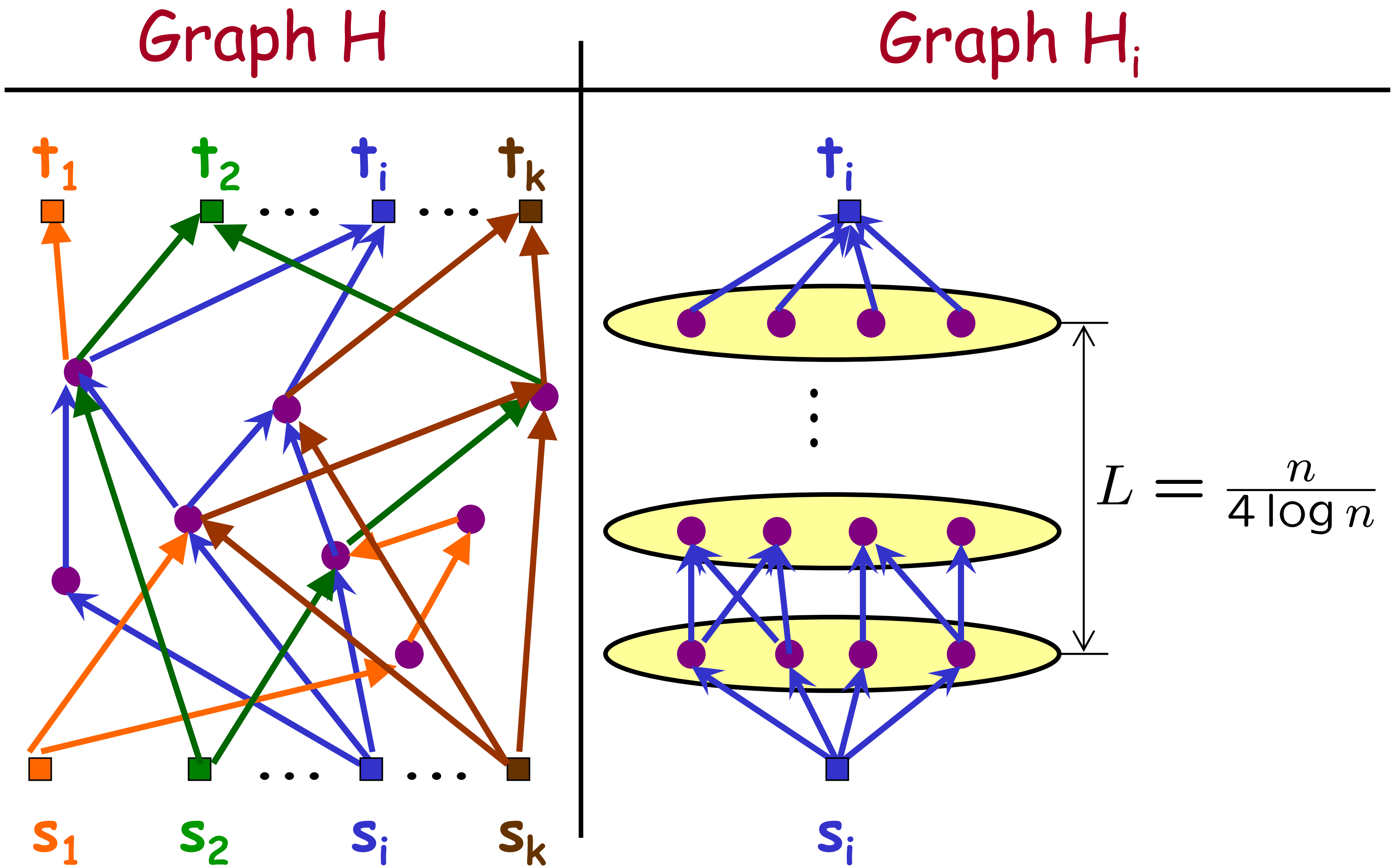
(2) Any integral solution needs to delete almost all vertices.

- To separate a pair s_i–t, need to delete all vertices in one of the L layers in H_i.
- Prob. that a fixed subset S of $n/16$ non-terminal vertices disconnects all pairs is exp. small.
- By union bounds, almost certainly no small solution.

What about Property (1)?

Canonical Paths: An $s_i \rightarrow t_i$ path is canonical iff it contains edges of type-i only.

- A canonical s_i-t_i path uses only edges from H_i.
- Length of any canonical s_i-t_i path in graph H is L.


Graph H
Graph H_i
t_1
t_2
t_i
t_k
s_1
s_2
s_i
s_k
t_i
s_i
L = n / (4 log n)

What about Property (1)?

Canonical Paths: An $s_i \rightarrow t_i$ path is canonical iff it contains edges of type-i only.

- Length of any canonical s_i-t_i path in graph H is L.
- But there are short non-canonical paths between source-sink pairs.

Transform H so that

(1a) Length of any canonical path stays at least L.

(1b) There are no non-canonical source-sink paths.

Step Two

- A multicut instance B, with source-sink pairs and canonical paths for each source-sink pair.
- No non-canonical paths exist in the graph, ensured by using a labeling scheme.
- Any integral solution must remove almost all vertices.
- But canonical paths in graph B can be "short".

Final Step: Composing H and B

Final Graph G: A "composition" of H and B.

- No non-canonical paths in G: pre-images in B don't have non-canonical paths.
- No short canonical paths: pre-images in H don't have short canonical paths.
- Any integral solution must remove almost all vertices.

Putting Things Together …

- G has $N = O(n^7)$ vertices.
- Each source-sink pair is connected only by a canonical path of length at least $L \approx n / \log n$.
- So there is a fractional solution of cost $O(N/L)$.
- Any integral solution must remove $\Omega(N)$ vertices.

Flow-Cut Gap = $\Omega(L) \approx \Omega(N^{1/7})$

Concluding Remarks

- Polynomial flow-cut gaps in directed graphs.
- Almost polynomial inapproximability results.
- Still a large gap remains between best upper and lower bounds on the flow cut gap.

Thank you !