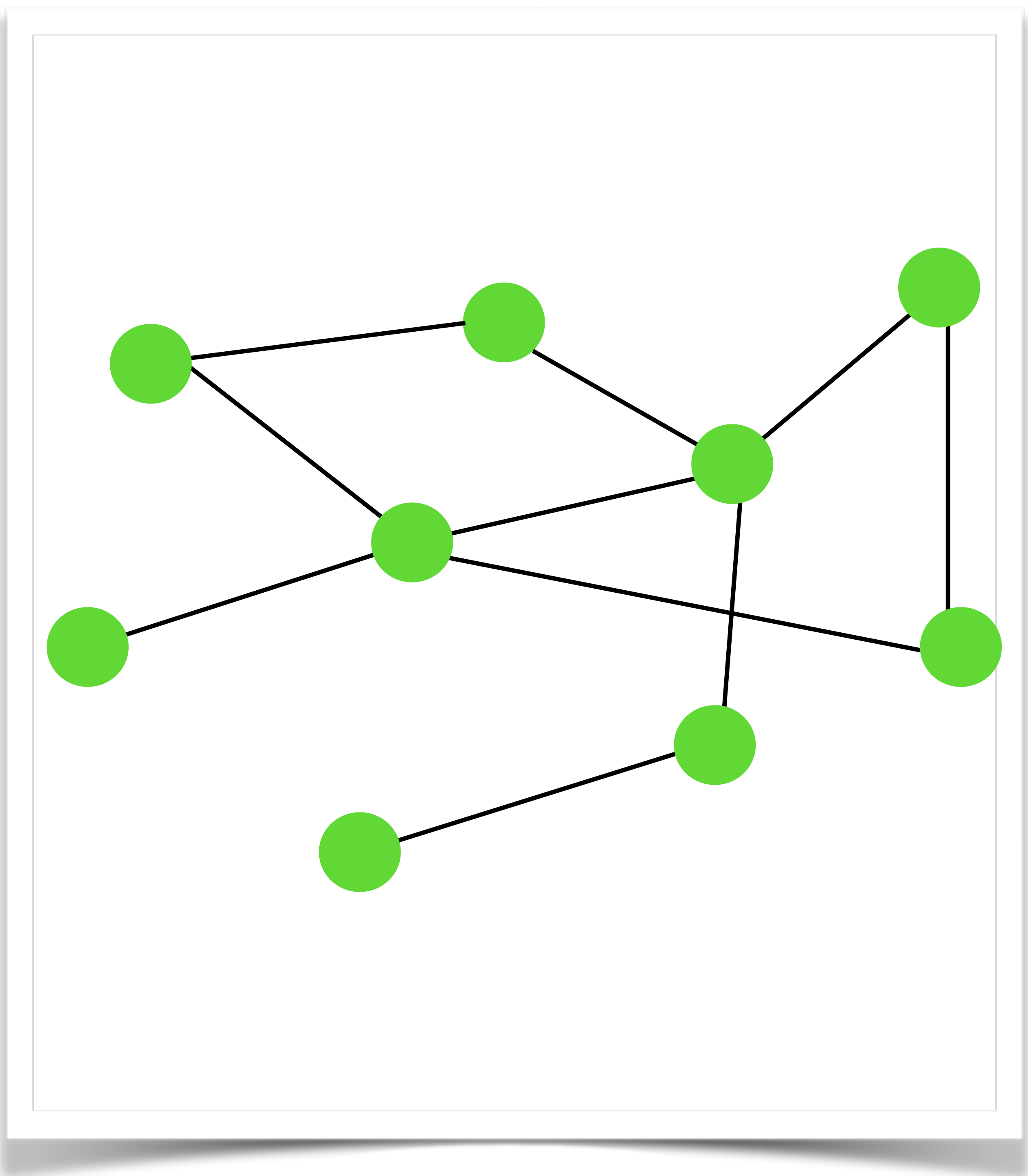
A Two-Pass (Conditional) Lower Bound for Semi-Streaming Maximum Matching

Sepehr Assadi

RUTGERS

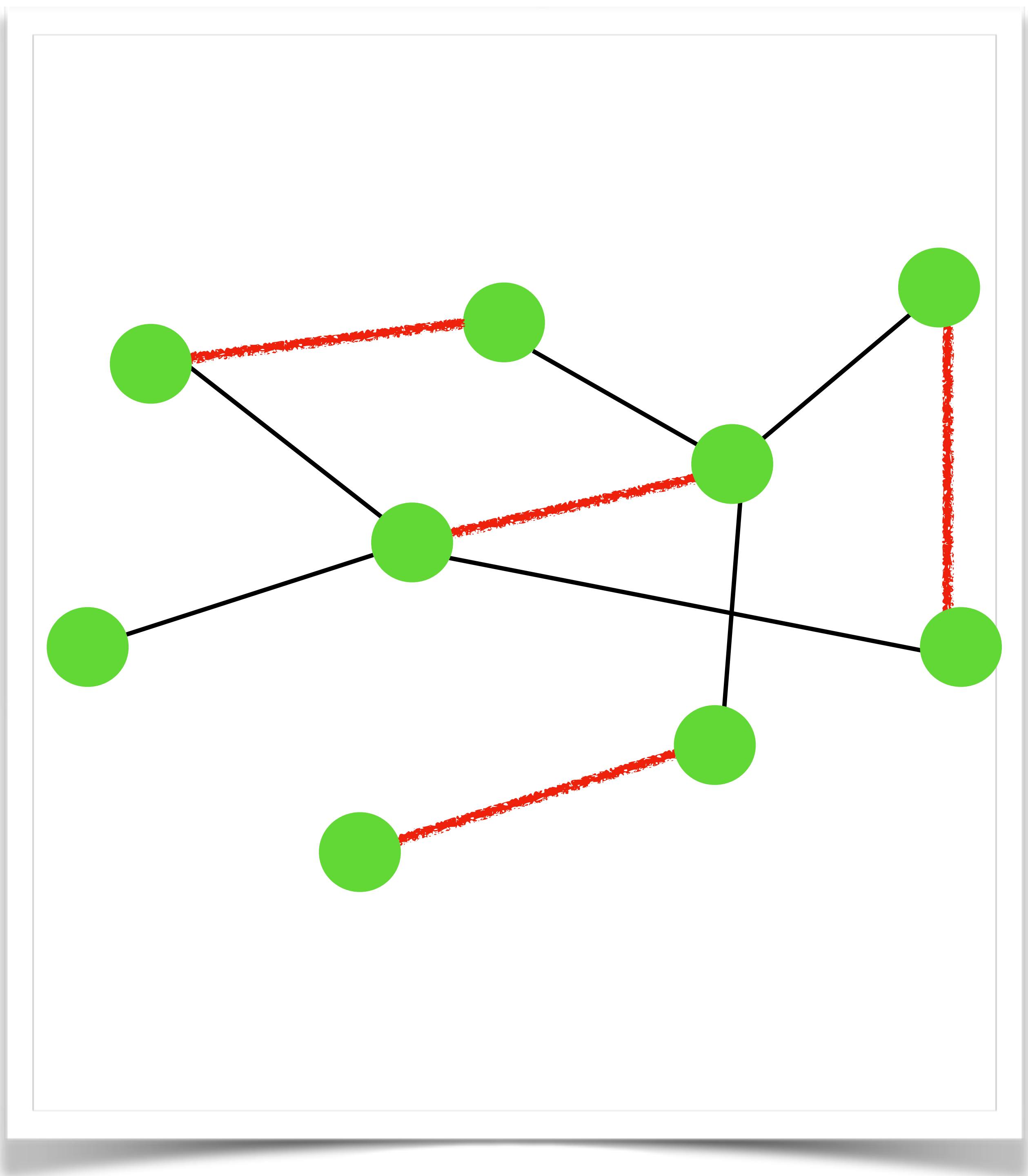
Main Result

Maximum Matching



Matching:
Any set of vertex-disjoint edges

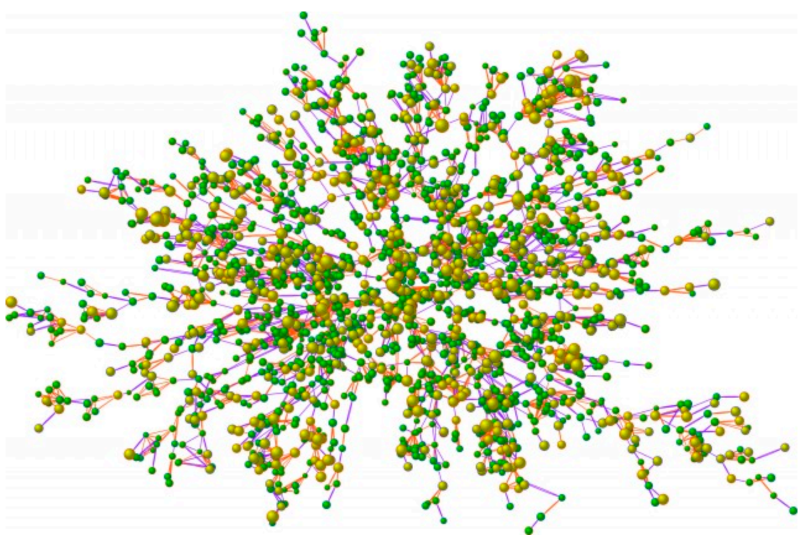
Maximum Matching

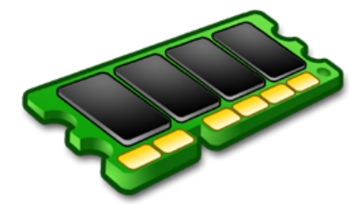


Matching:
Any set of vertex-disjoint edges

Maximum Matching Problem:
Find a matching of largest size

Semi-Streaming Model





$\tilde{O}(n)$ size memory

e_1	e_2	e_3	e_4	• • •	e_m

Up to $O(n^2)$ edges

- Semi-streaming model of computation
  - [Feigenbaum, Kannan, McGregor, Suri, Zhang; 2005]

- A lower ... semi-streaming a...

Theorem: Best appr... ation ratio possible by two-pass semi-streaming algorithms for max...u... matching is at most:

$$1 - \Omega\left(\frac{\log \mathsf{RS}(n)}{\log n}\right)$$

...ings ... size $\Theta(n)$ in

- (a measure of) density of

A well-studied but not well-understood parameter:

$$n^{\Omega(\frac{1}{\log\log n})} \leq \mathsf{RS}(n) \leq \frac{n}{2^{O(\log^* n)}}$$

If the lower bound on $\mathsf{RS}(n)$ is tight:

No $1 - o(\frac{1}{\log\log n})$ approximation

If the upper bound on $\mathsf{RS}(n)$ is tight:

No 0.98 approximation

Rest of this Talk

Question 1

How to interpret?

Question 2

Why do we care?

Question 3

How to prove?

Rest of this Talk

Question 1

How to interpret?

Question 2

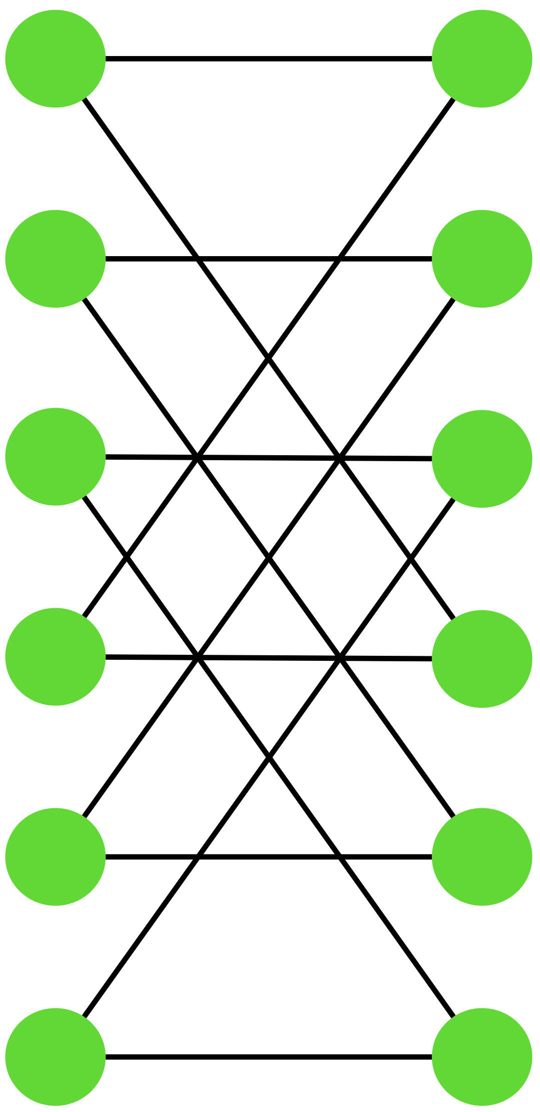
Why do we care?

Question 3

How to prove?

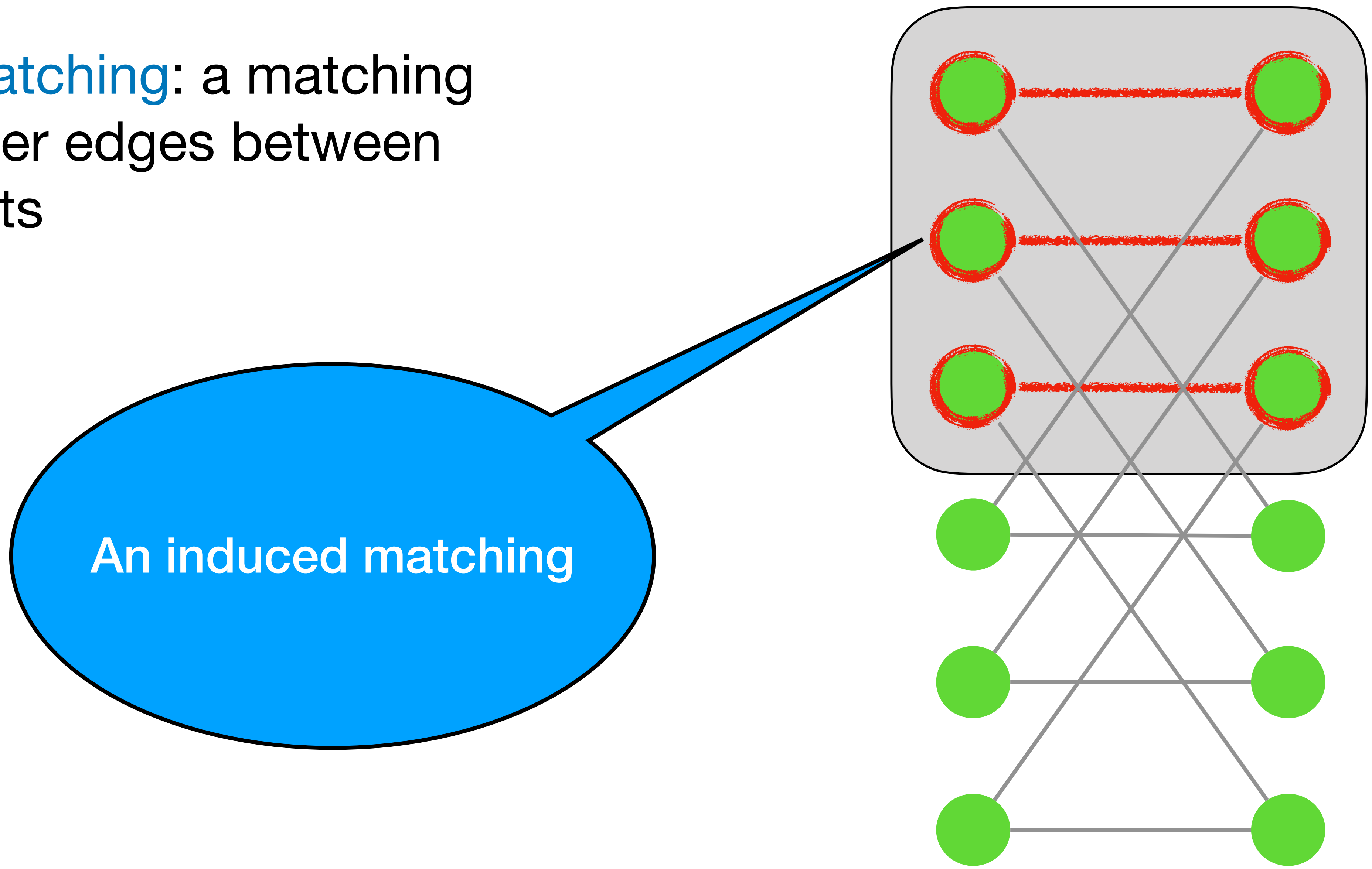
Ruzsa-Szemeredi (RS) Graphs

- Induced matching: a matching with no other edges between its endpoints



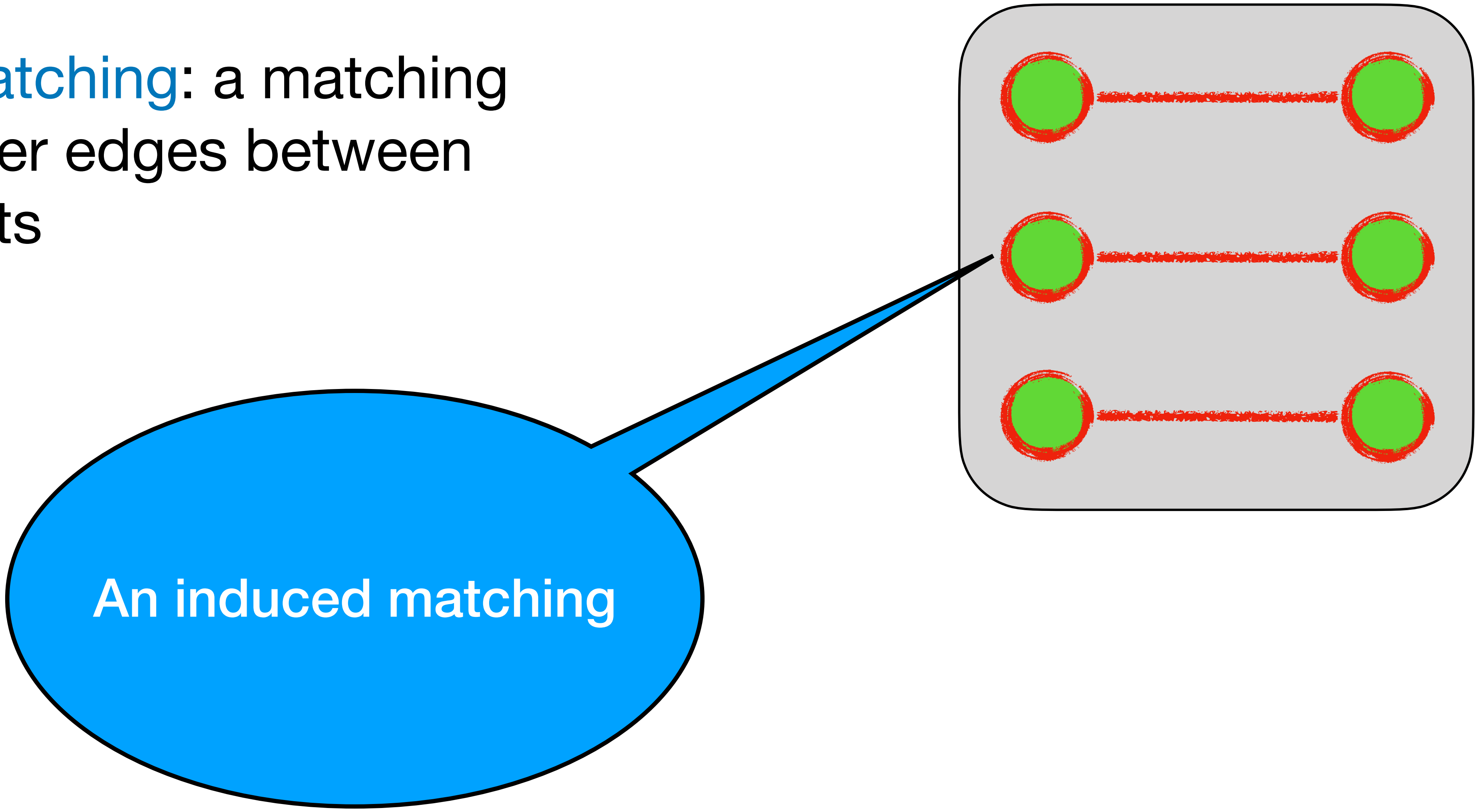
Ruzsa-Szemeredi (RS) Graphs

- Induced matching: a matching with no other edges between its endpoints



Ruzsa-Szemeredi (RS) Graphs

- Induced matching: a matching with no other edges between its endpoints

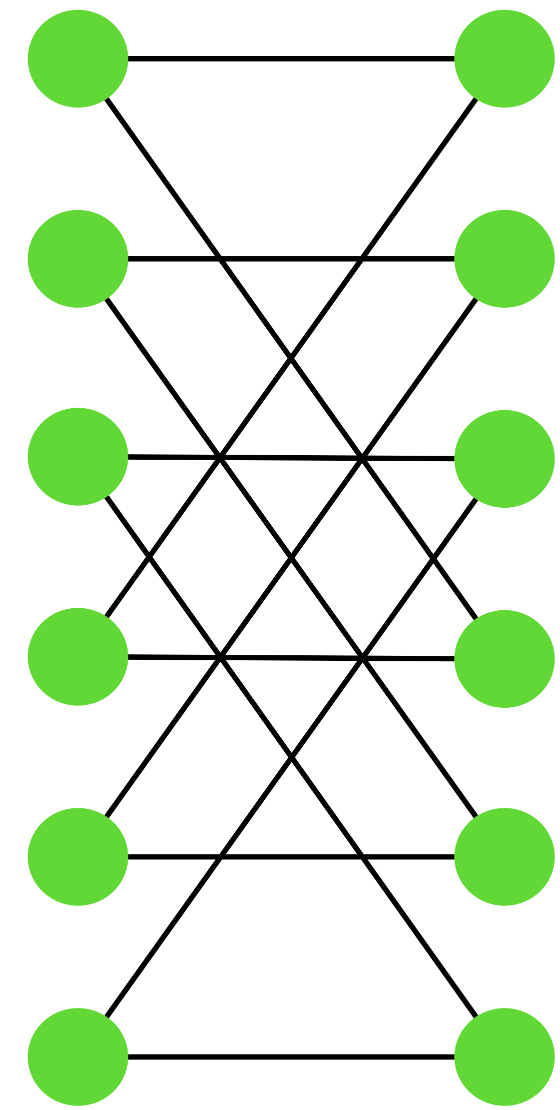


Ruzsa-Szemeredi (RS) Graphs

- Induced matching: a matching with no other edges between its endpoints

- (r, t)-RS graph: A graph with t edge-disjoint induced matchings of size r

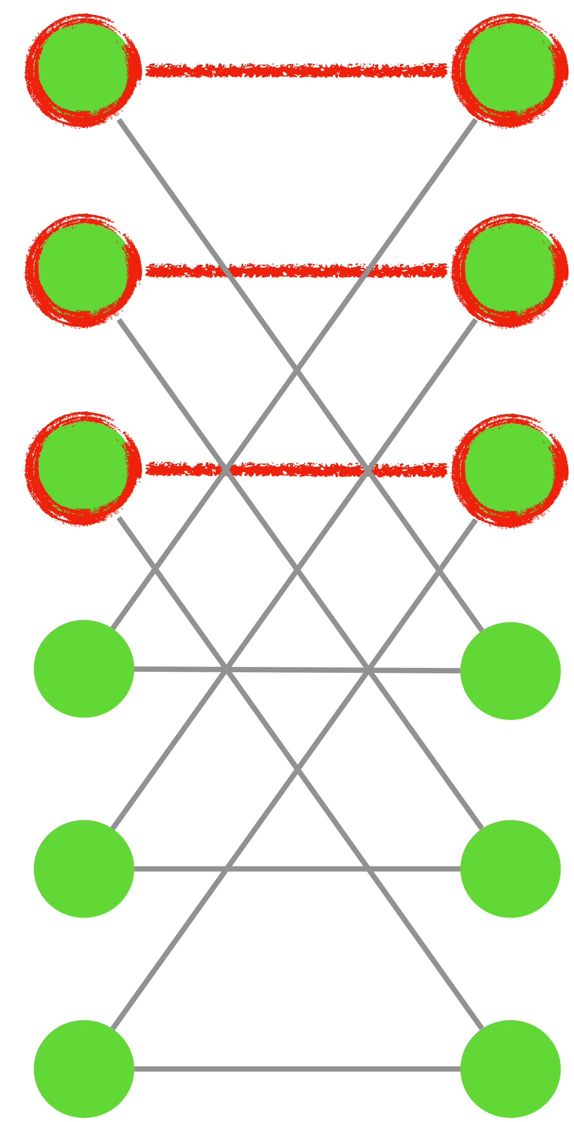
A (3,4)-RS graph



Ruzsa-Szemeredi (RS) Graphs

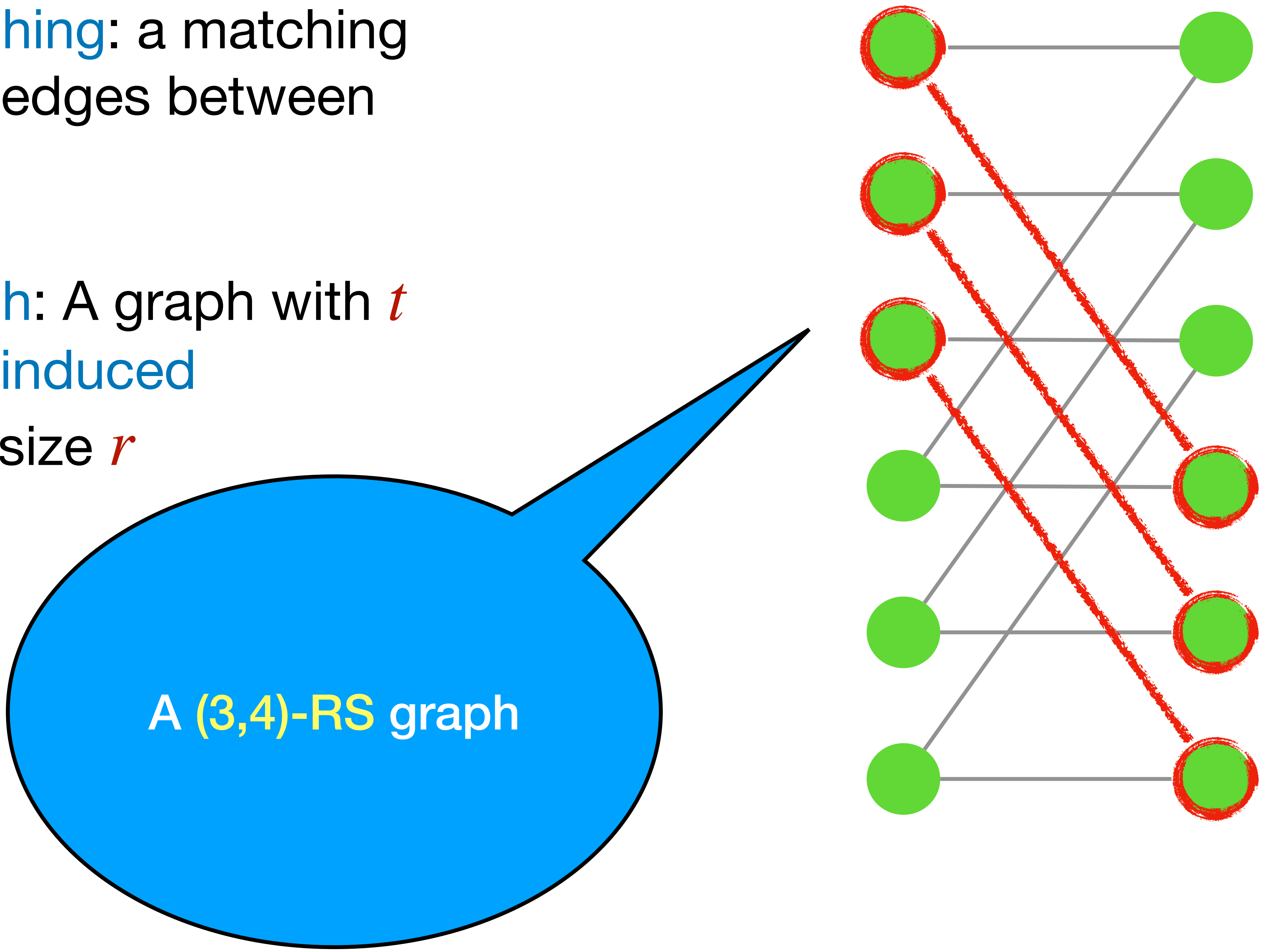
- Induced matching: a matching with no other edges between its endpoints

- (r, t)-RS graph: A graph with t edge-disjoint induced matchings of size r



Ruzsa-Szemeredi (RS) Graphs

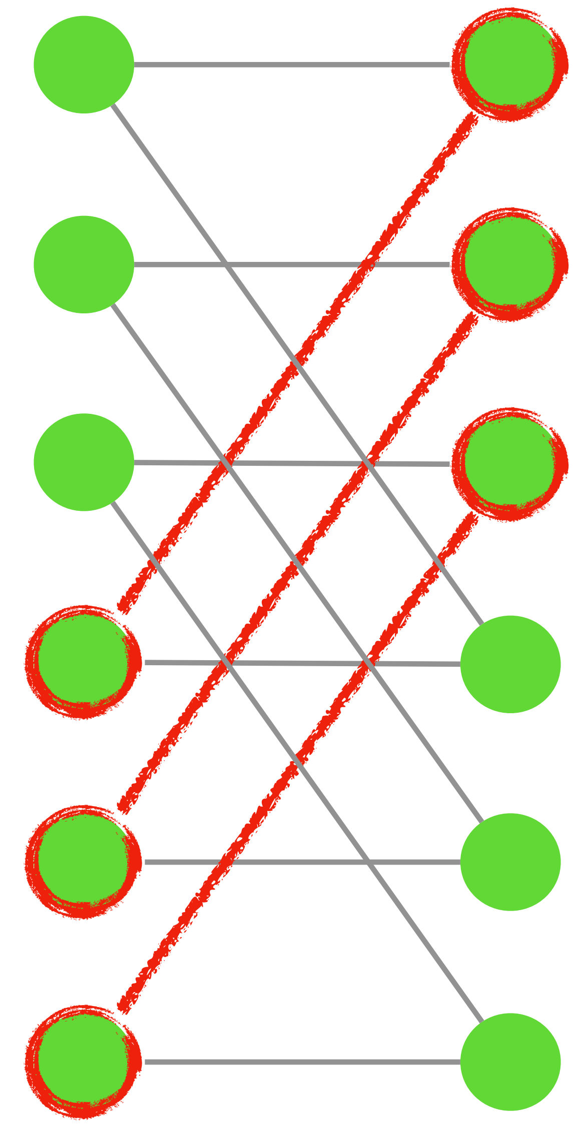
- Induced matching: a matching with no other edges between its endpoints
- (r, t)-RS graph: A graph with t edge-disjoint induced matchings of size r



Ruzsa-Szemeredi (RS) Graphs

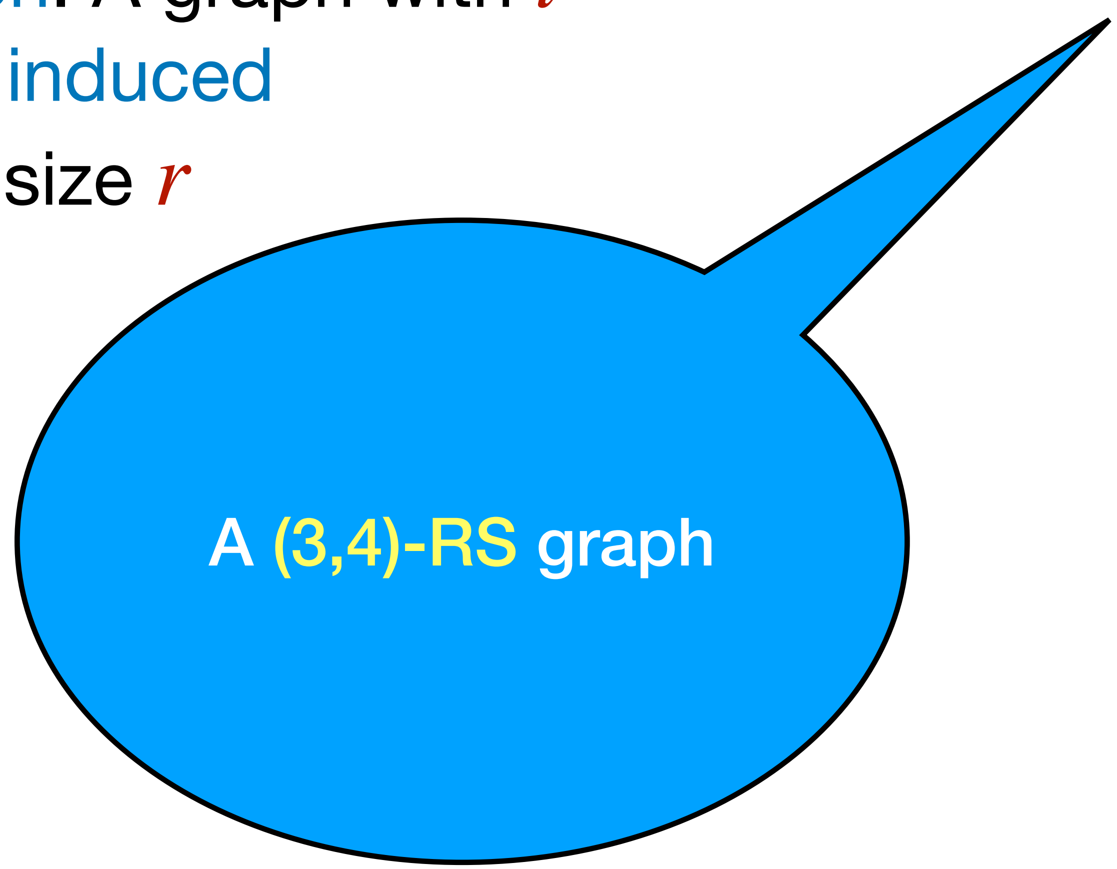
- Induced matching: a matching with no other edges between its endpoints
- (r, t)-RS graph: A graph with t edge-disjoint induced matchings of size r

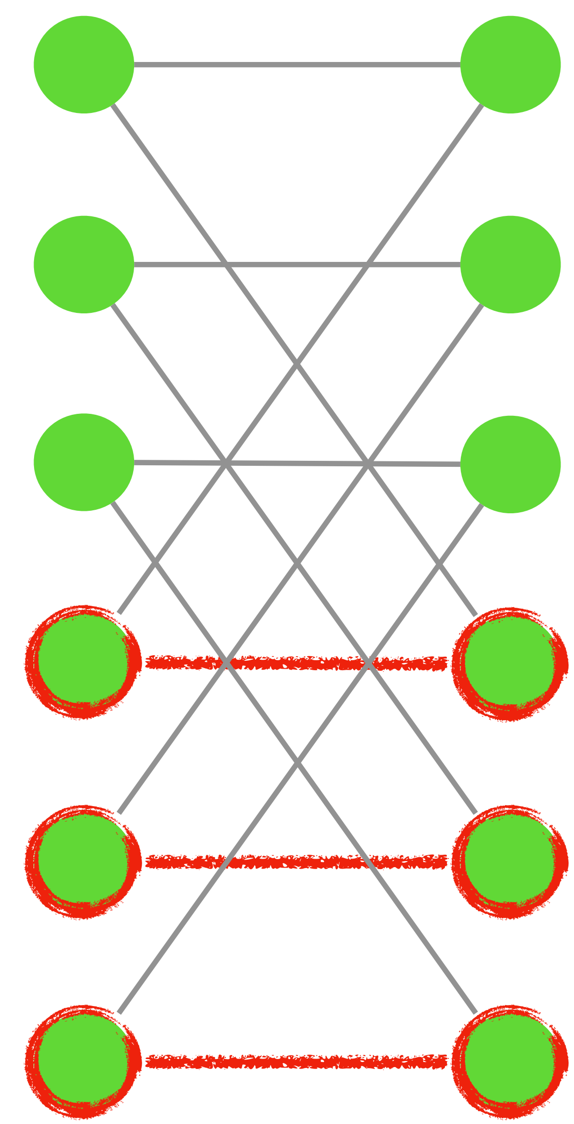
A (3,4)-RS graph



Ruzsa-Szemeredi (RS) Graphs

- Induced matching: a matching with no other edges between its endpoints
- (r, t)-RS graph: A graph with t edge-disjoint induced matchings of size r



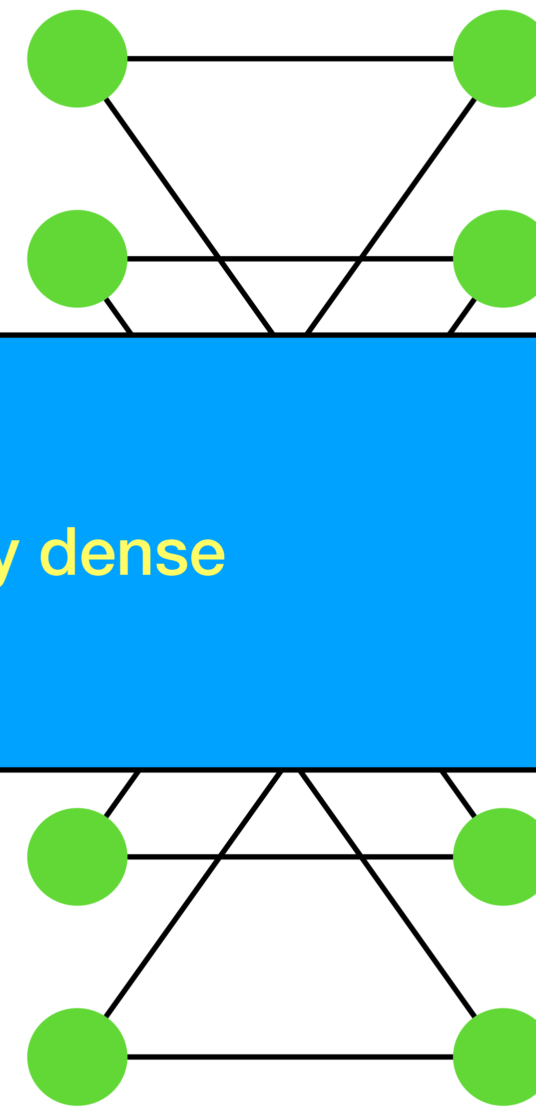


Ruzsa-Szemeredi (RS) Graphs

- Induced matching: a matching with no other edges between its endpoints
-

RS graphs are locally sparse but globally dense

- We are interested in (r, t)-RS graphs with **large** r and t



Ruzsa-Szemeredi (RS) Graphs

RS(n): largest value of t in an (r,t)-RS graph on n vertices with $r = \Theta(n)$

$$n^{\Omega\left(\frac{1}{\log\log n}\right)} \leq \textbf{RS}(n) \leq \frac{n}{2^{O(\log^* n)}}$$

Ruzsa-Szemeredi (RS) Graphs

This work:
we use them separately for both purposes

- RS graphs are studie
  - Property testing, P
    Streaming algorithms
- Used first in [Goel, Khanna, K alov; 2012] for semi-streaming matching problem
  - Subsequently in [K apralov; 2013][Konrad; 2015][**A**, Khanna, Li, Yarostlatsev; 20 6][**A**, Khanna, Li; 2017][Kapralov; 2021] …
- Used first in [**A**, Raz; 2020] for “hiding” information from multi-pass streaming algorithms
  - Subsequently in [Chen, Kol, Paramonov, Saxena, Song, Yu; 2021]

Interpreting Our Result

Theorem: Best approximation ratio of by two-pass semi-streaming matching is at most $1 - \Omega(\frac{\log \mathsf{RS}(n)}{\log n})$

$$n^{\Omega(\frac{1}{\log\log n})} \leq \mathsf{RS}(n) \leq \frac{n}{2^{O(\log^* n)}}$$

- Conditional lower bound:

Moral of the Story

An arbitrarily small-constant factor approximation to matching via two-pass semi-streaming algorithms is either quite hard or even impossible

bound of $\mathsf{RS}(n)$ from $\frac{n}{2^{O(\log^* n)}}$ all the way to $n^{o(1)}$

Interpreting Our Result

Theorem: Best approximation ratio of by two-pass semi-streaming matching is at most $1 - \Omega(\frac{\log \mathsf{RS}(n)}{\log n})$

$$n^{\Omega(\frac{1}{\log \log n})} \leq \mathbf{RS}(n) \leq \frac{n}{2^{O(\log^* n)}}$$

- Currently, the best two-pass semi-streaming algorithm achieves a $(2 - \sqrt{2}) \approx 0.58$ approximation [Konrad, Naidu; 2021]
  - Following [Konrad, Magniez, Matheu; 2012][Esfandiari, Hajiaghayi, Monemizadeh; 2016][Kale, Tirodkar; 2017][Konrad; 2018]
- Previously, best two-pass semi-streaming lower bound ruled out $(1 - \frac{1}{n^{o(1)}})$ approximation [Chen, Kol, Paramonov, Saxena, Song, Yu; 2021]
  - Following [Guruswami, Onak; 2013][**A**, Raz; 2020]

Rest of this Talk

Question 1

How to interpret?

Question 2

Why do we care?

Question 3

How to prove?

Rest of this Talk

Question 1

How to interpret?

Question 2

Why do we care?

Question 3

How to prove?

Two-Pass Algorithms for Matching

- Maximum matching is among the most studied problems in the semi-streaming model

- A long line of work studied two-pass algorithms for this problem

  - [Konrad, Magniez, Matheu; 2012][Esfandiari, Hajiaghayi, Monemizadeh; 2016][Kale, Tirodkar; 2017][Konrad; 2018][Konrad, Naidu; 2021]

- Yet, no non-trivial lower bound for constant-factor approximation algorithms were known*

- Our result is thus the first to address this regime

* [Konrad, Naidu; 2021] independently and concurrently proved a lower bound for special case of algorithms that only run the greedy algorithm in their first pass

Detour: Bigger Picture

- For single-pass algorithms, the state-of-the-art upper and lower bounds go hand in hand
  - We have the tools to prove pretty strong lower bounds!
- For multi-pass algorithms, the state-of-the-art upper and lower bounds are quite far from each other
  - Lower bound techniques are lacking considerably!
  - Two passes is already where this gap emerges

Detour: Bigger Picture

- A general goal of my research:
  - Develop new techniques for multi-pass streaming lower bounds
  - [**A**, Chen, Khanna; 2019][**A**, Raz; 2020][**A**, Kol, Saxena, Yu; 2020], [**A**, Vishvajeet; 2021]
- Our result in this work is a proof of concept for these techniques:
  - At least in the ballpark of current algorithms…

Rest of this Talk

Question 1

How to interpret?

Question 2

Why do we care?

Question 3

How to prove?

Rest of this Talk

Question 1

How to interpret?

Question 2

Why do we care?

Question 3

How to prove?

Disclaimer: Technical details will be imprecise for conveying the intuition

Our Approach

Theorem: Best approximation ratio of by two-pass semi-streaming matching is at most $1 - \Omega(\frac{\log \mathrm{RS}(n)}{\log n})$

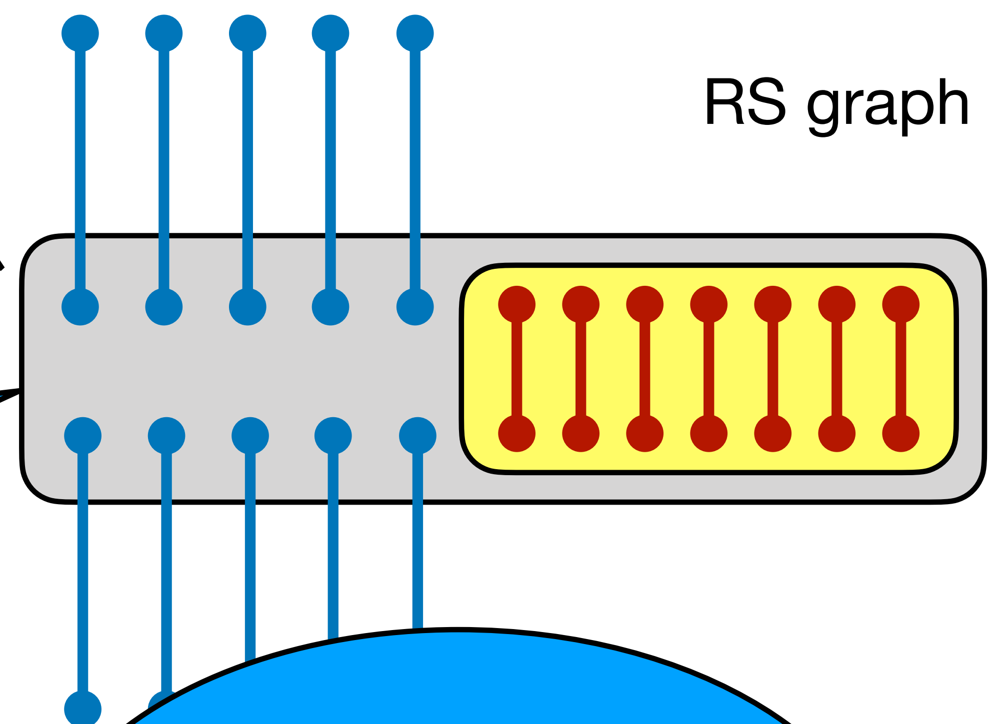
- Combination of several techniques:
  - Single-pass lower bound of [Goel, Kapralov, Khanna; 2012]
  - Two-pass lower bound framework of [**A**, Raz; 2020]
  - The "XOR-gadget" approach of [**A**, Behnezhad; 2021], [Chen, Kol, Paramonov, Saxena, Song, Yu; 2021]
  - "XOR-lemmas" for analyzing XOR-gadgets [**A**, Vishvajeet; 2021] [Gavinsky, Kempe, Kerenidis, Raz, de Wolf; 2007][Verbin, Yu; 2011]

Single-Pass Lower Bound

[Goel, Kapralo

- First part of the strea
  (r, t)-RS graph for
  $r < \frac{n}{4}, t = n^{\Omega(\frac{1}{\log \log n})}$

This graph has a perfect matching



A streaming algorithm that “remembers” $o(1)$ fraction of the RS graph only “remembers” $o(1)$ fraction of the special induced matching

y

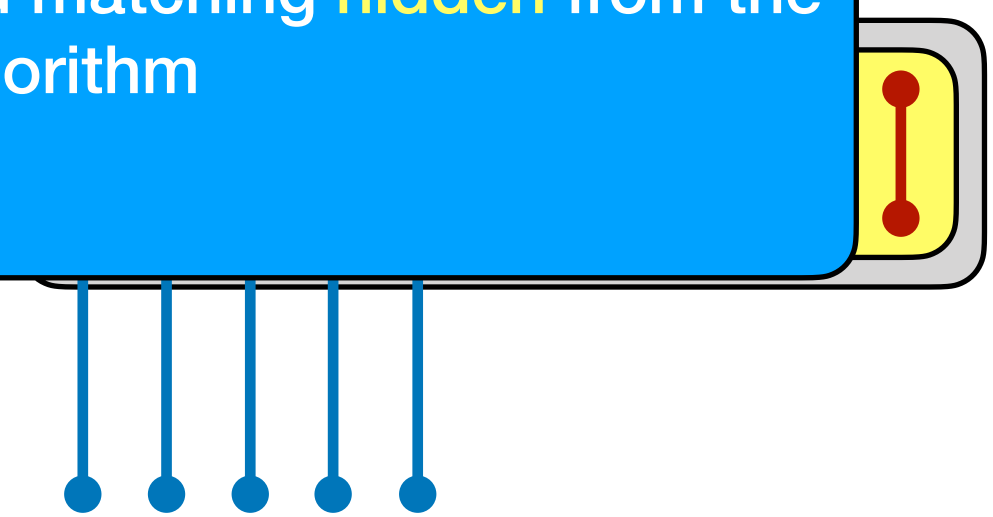
Semi-streaming algorithms cannot achieve better than $2/3$ approximation

A Two-Pass Lower Bound?

- What happens to this family of instances in two passes?
- It b… …aph
  -
  - Find the special induced matching
  - Store all its edges in the second pass

Takeaway:

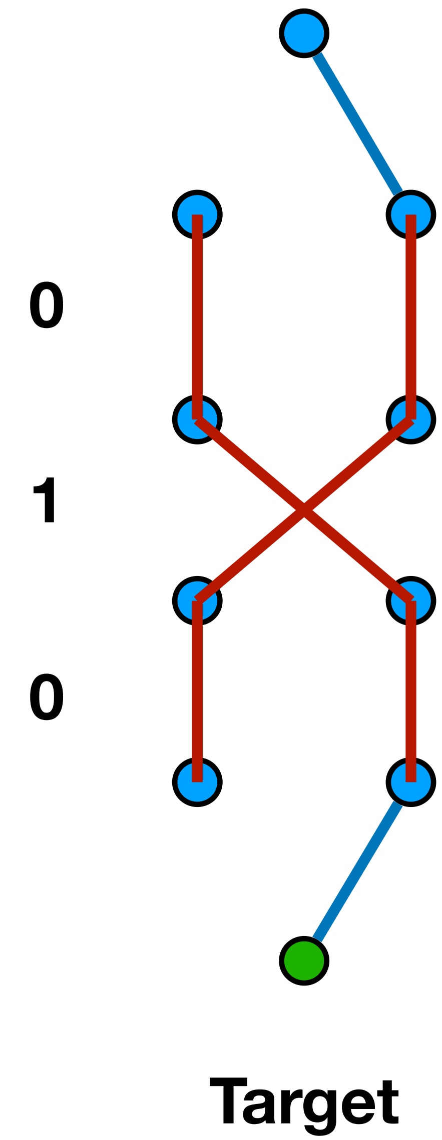
Keep the identity of the special induced matching hidden from the first pass of the algorithm



XOR Gadgets

XOR-gadget of [**A**, Behnezhad; 2021]:

- Straight connection represents zero
- Cross connection represents one

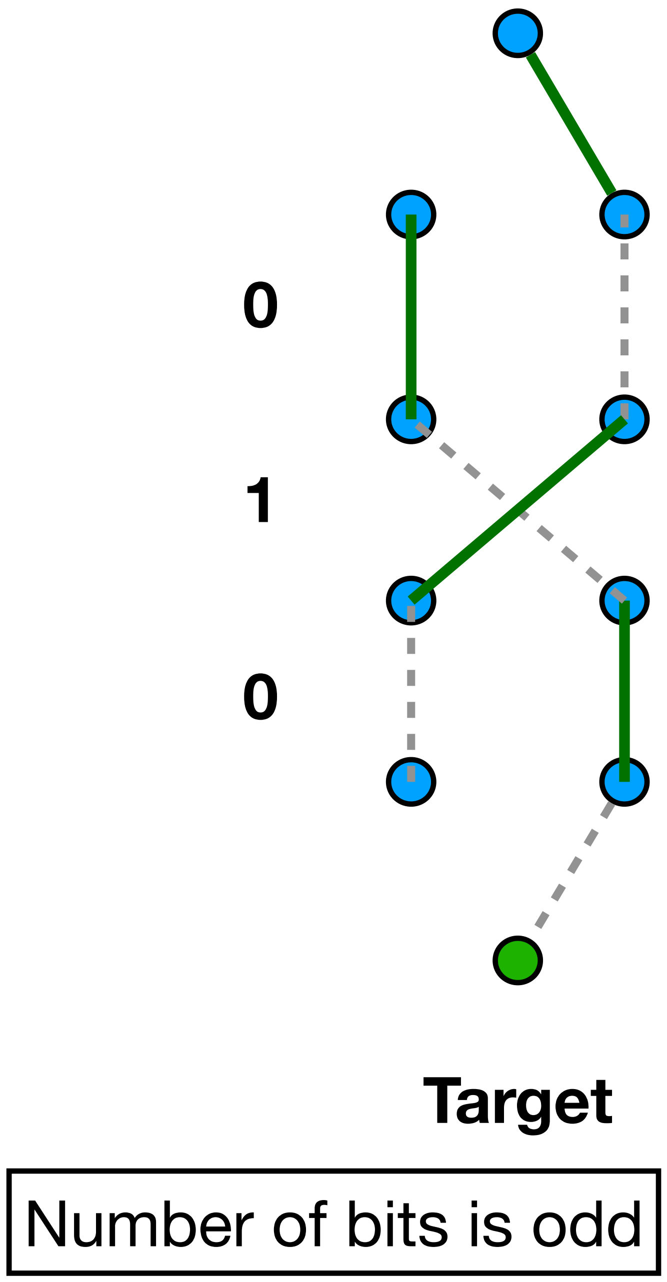


Number of bits is odd

XOR Gadgets

XOR-gadget of [**A**, Behnezhad; 2021]:

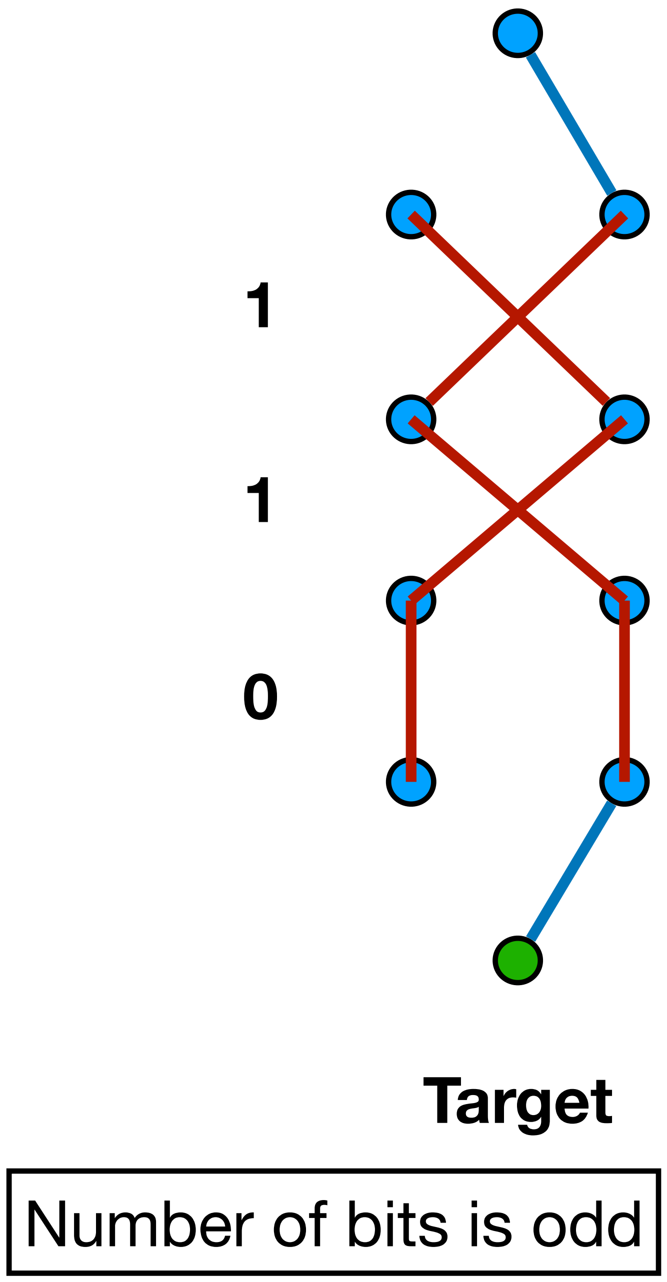
- Straight connection represents zero
- Cross connection represents one
- XOR is one: there is a maximum matching leaving target unmatched



XOR Gadgets

XOR-gadget of [**A**, Behnezhad; 2021]:

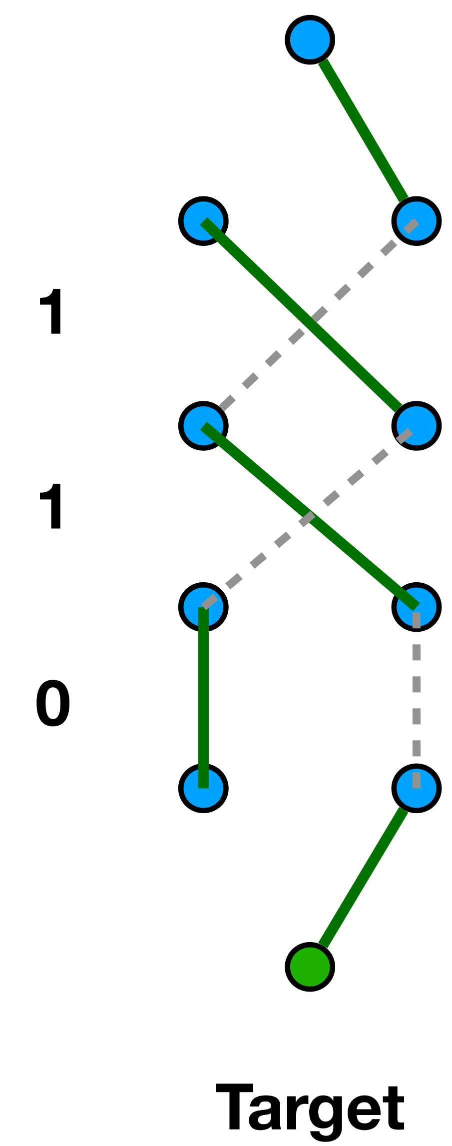
- Straight connection represents zero
- Cross connection represents one
- XOR is one: there is a maximum matching leaving target unmatched
- XOR is zero: the unique maximum matching matches the target



XOR Gadgets

XOR-gadget of [**A**, Behnezhad; 2021]:

- Straight connection represents zero
- Cross connection represents one
- XOR is one: there is a maximum matching leaving target unmatched
- XOR is zero: the unique maximum matching matches the target

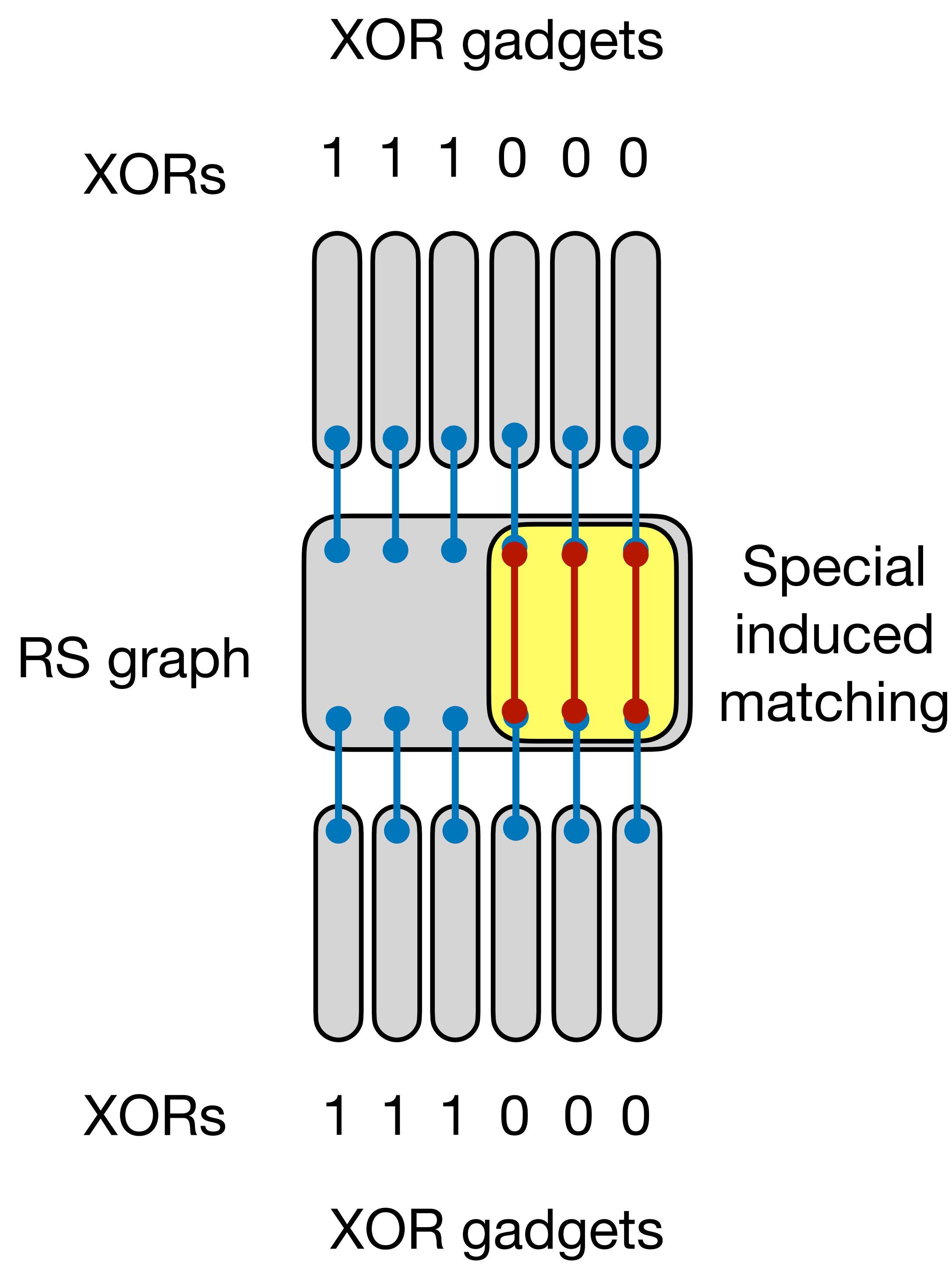


Number of bits is odd

XOR Gadgets

XOR-gadget of [A, Behnezhad; 2021]:

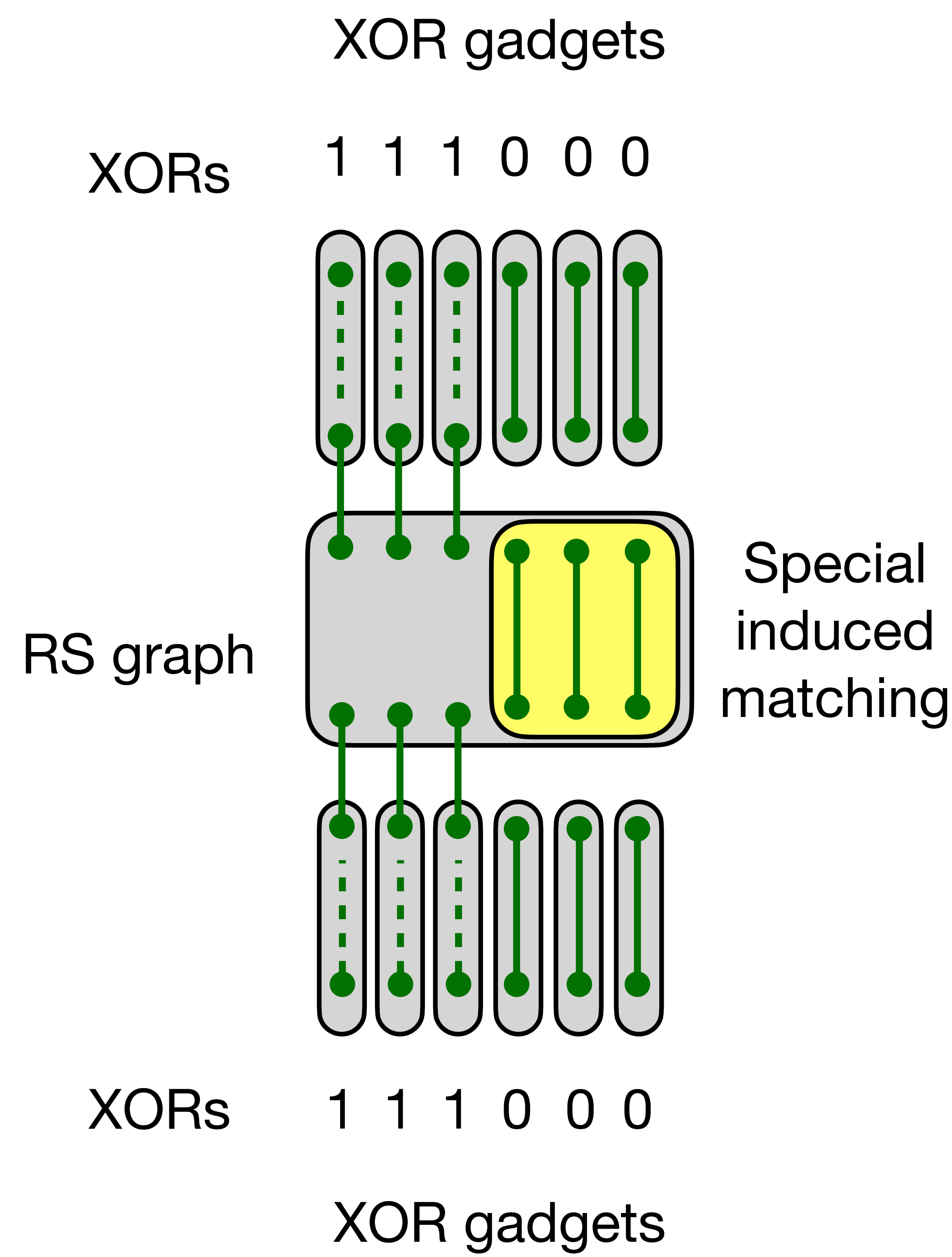
- Use as a switch for hiding the special induced matching



XOR Gadgets

XOR-gadget of [A, Behnezhad; 2021]:

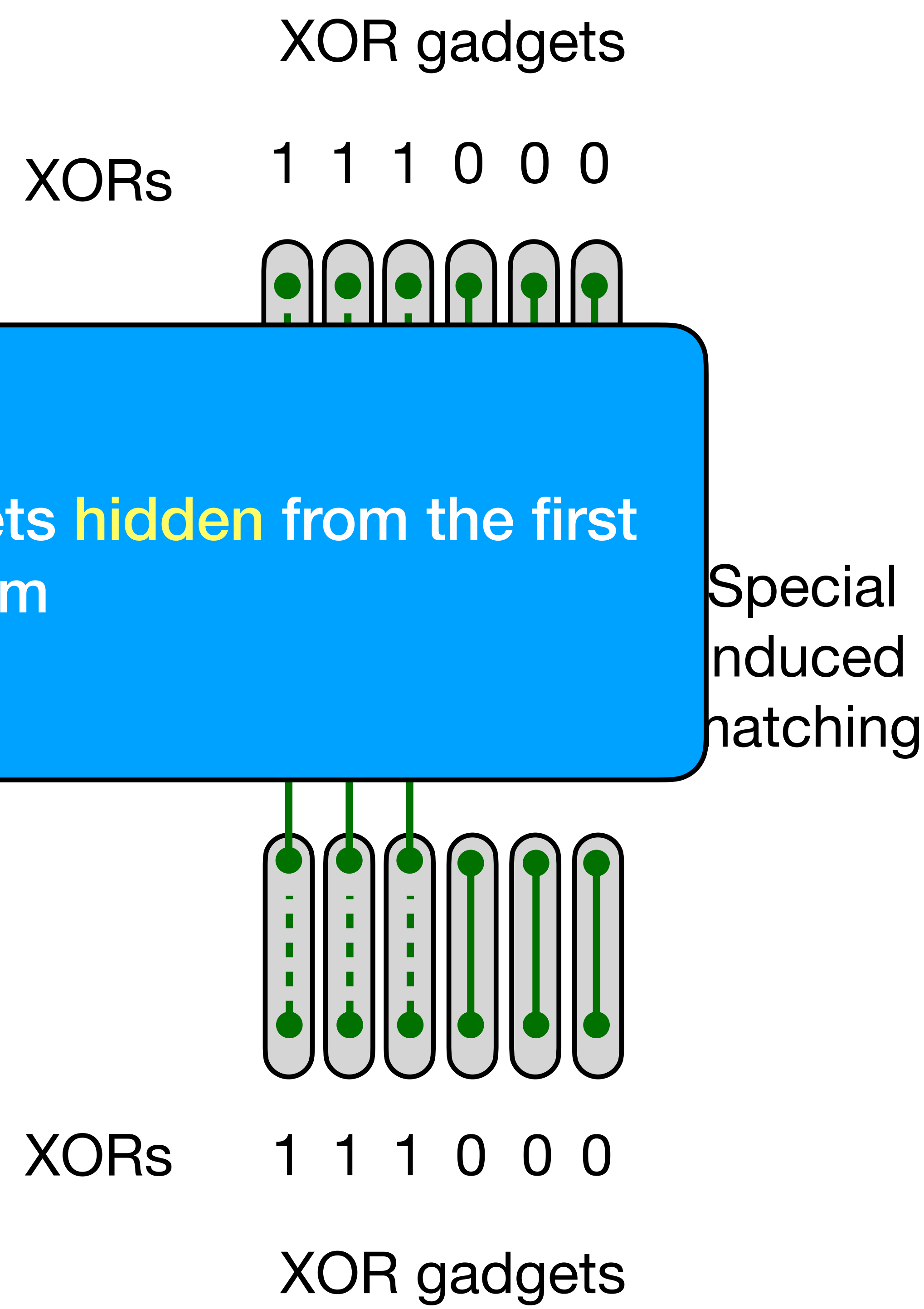
- Use as a switch for hiding the special induced matching
- Any (near) maximum matching has to pick edges of the special induced matching



A Two-Pass Lower Bound?

- Is this family of instances hard for two pass algorithms?

- St
  -
  - Find the special induced matching
  - Store all its edges in the second pass



Takeaway:

Keep the "XOR value" of the XOR-gadgets hidden from the first pass of the algorithm

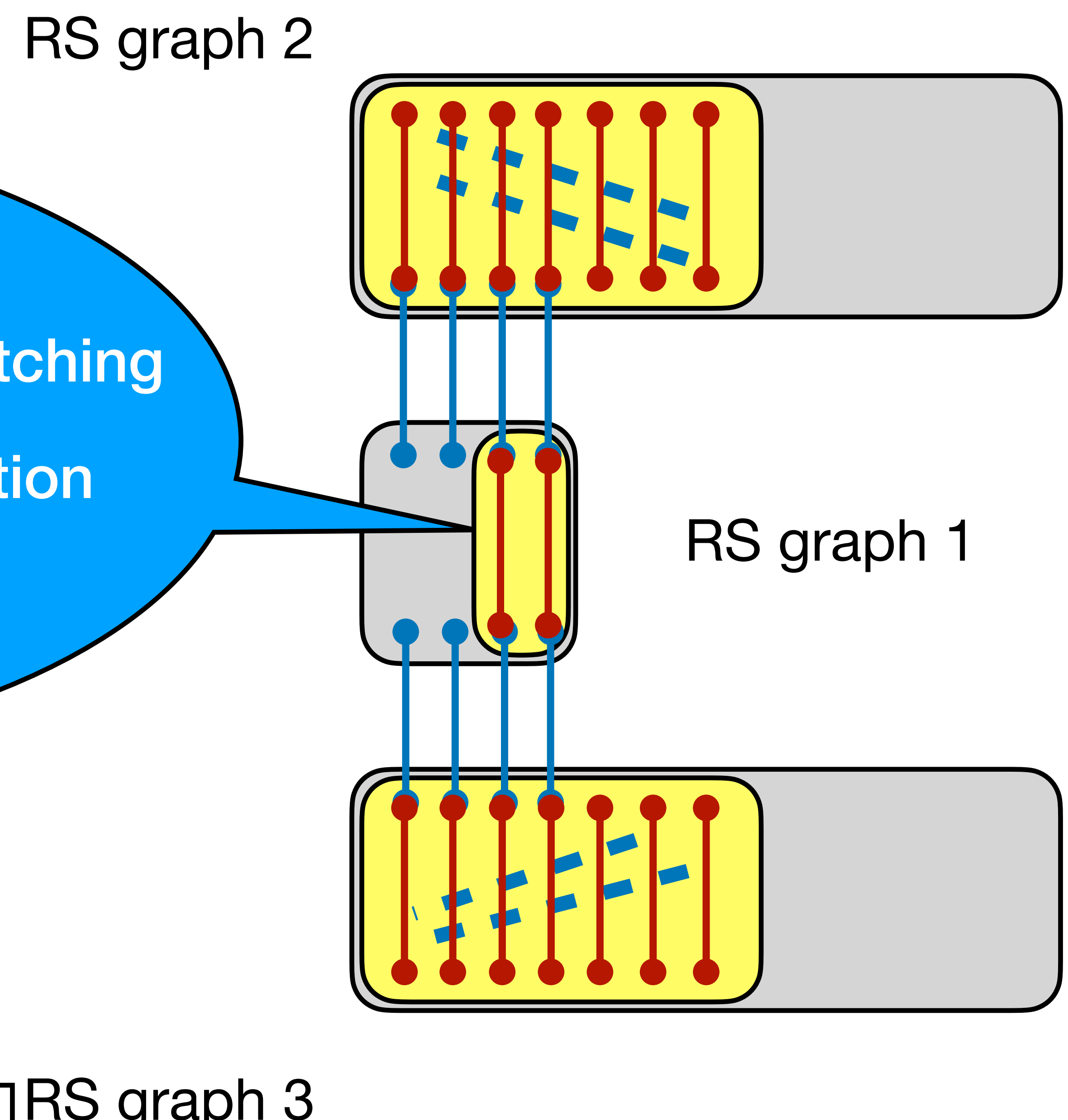
Two-Pass Lower Bound Framework

An adaptation of [**A**, Raz; 2020]:

- Hide th…
  larg…

- A…
  XO…
  a sing…

One needs the special induced matching to get a $(1 - \Theta(\frac{1}{k}))$-approximation



* Embed $\approx \frac{r}{k}$ gadgets of length k in an induced matching of size r in an (r, t)-RS graph

… Bound?

Many induced matchings: not much is "remembered" about a random one

Many induced … "remembered" about a random one

A. First RS graph 1 arrives

B. Then RS graphs 2 and 3 arrive

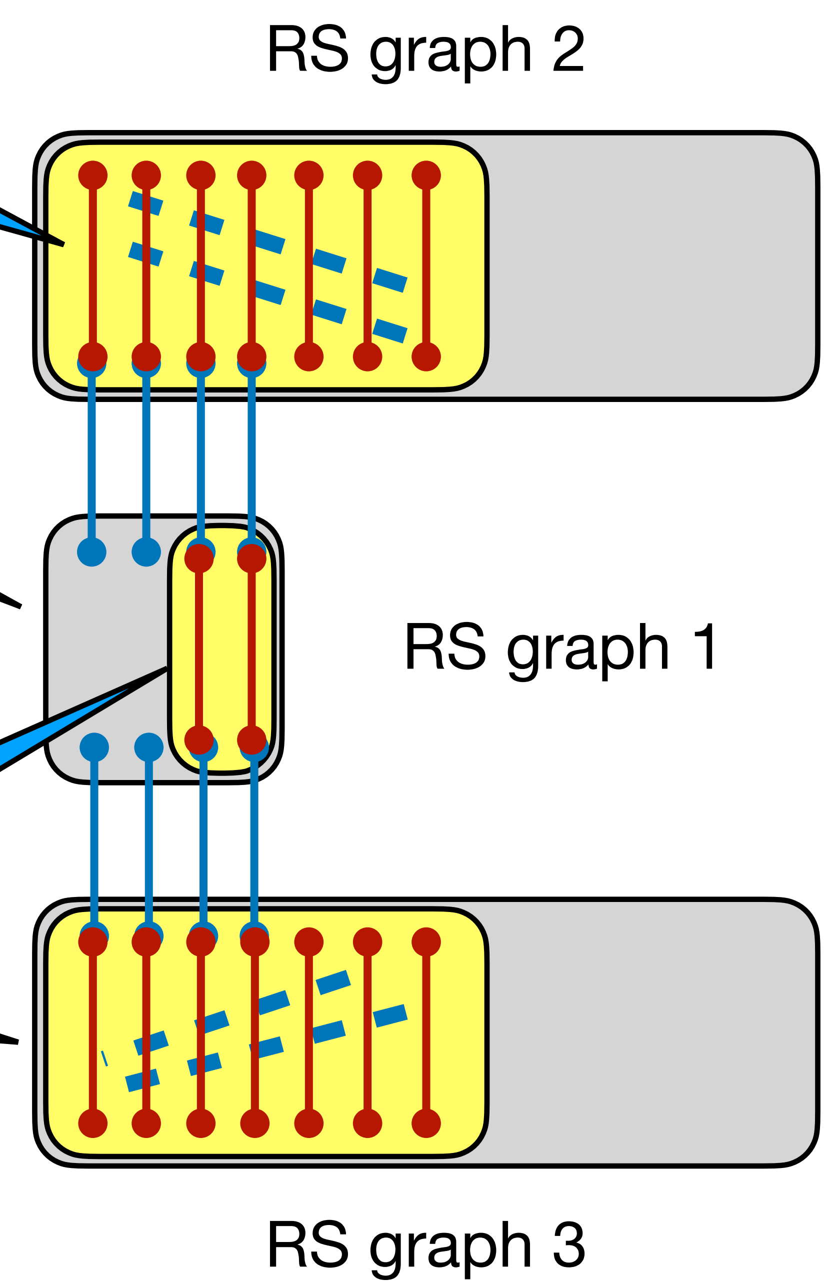
…ing

Identity of this induced matching is known but NOT its content

D. …

matching… graph… d 3

Even the identity of this induced matching is unknown



A Two-Pass Lower Bound?

Many induced matchings: not much
“remembered” about a random one

Identity of this induced matching is still
“random” even after the first pass

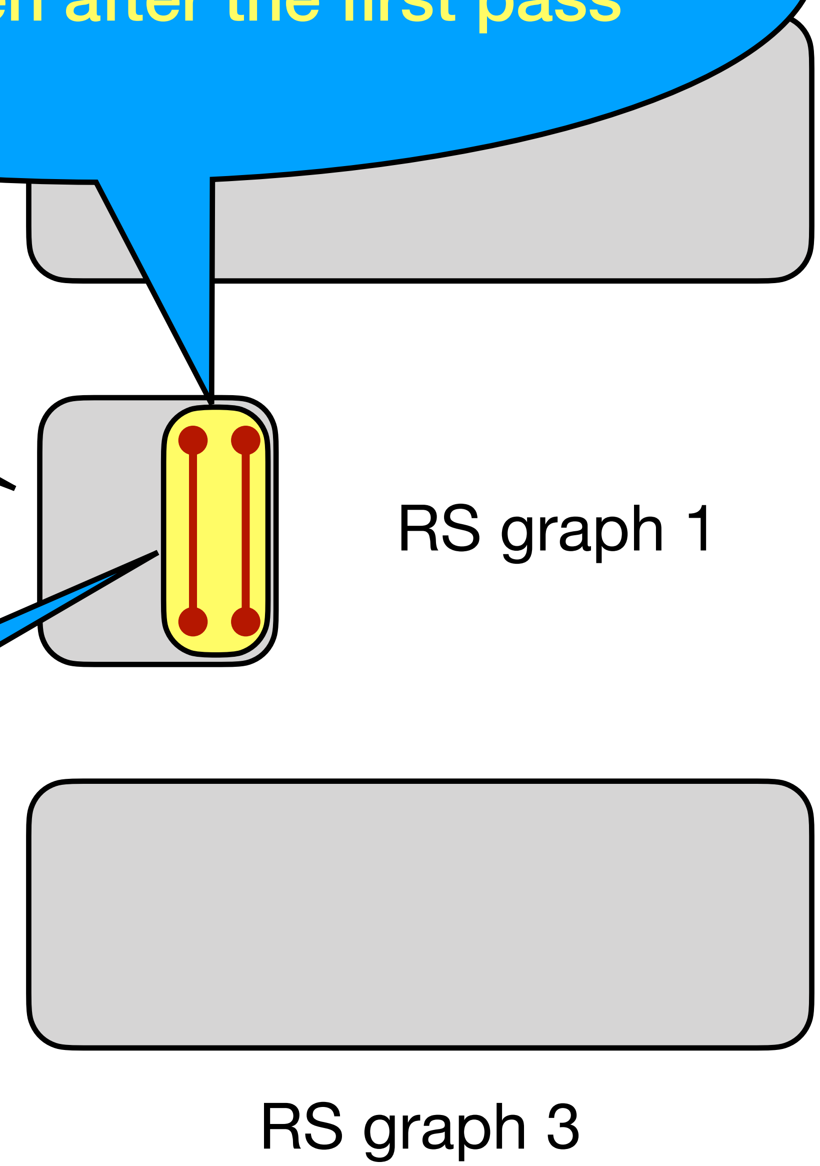
A. First RS graph 1 arrives

B. Then RS graphs 2 and 3 arrive

C. We pick a special induced matching from RS graph 1

D. We pick two special induced matchings from RS graph and 3

The algorithm still cannot “remember” these edges



A Two-Pass Lower Bound?

Many induced matchings: not muc
"remembered" about a random on

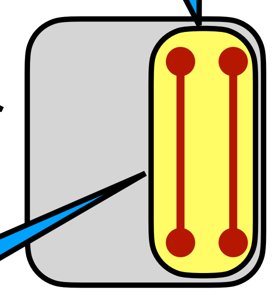
Identity of this induced matching is still "random" even after the first pass

A. First RS graph 1 arrives

B. Then RS graphs 2 and 3 arrive

C. We pick a special induced matching from RS graph 1

D. We pick two special induced matchings from RS grap nd 3



RS graph 1

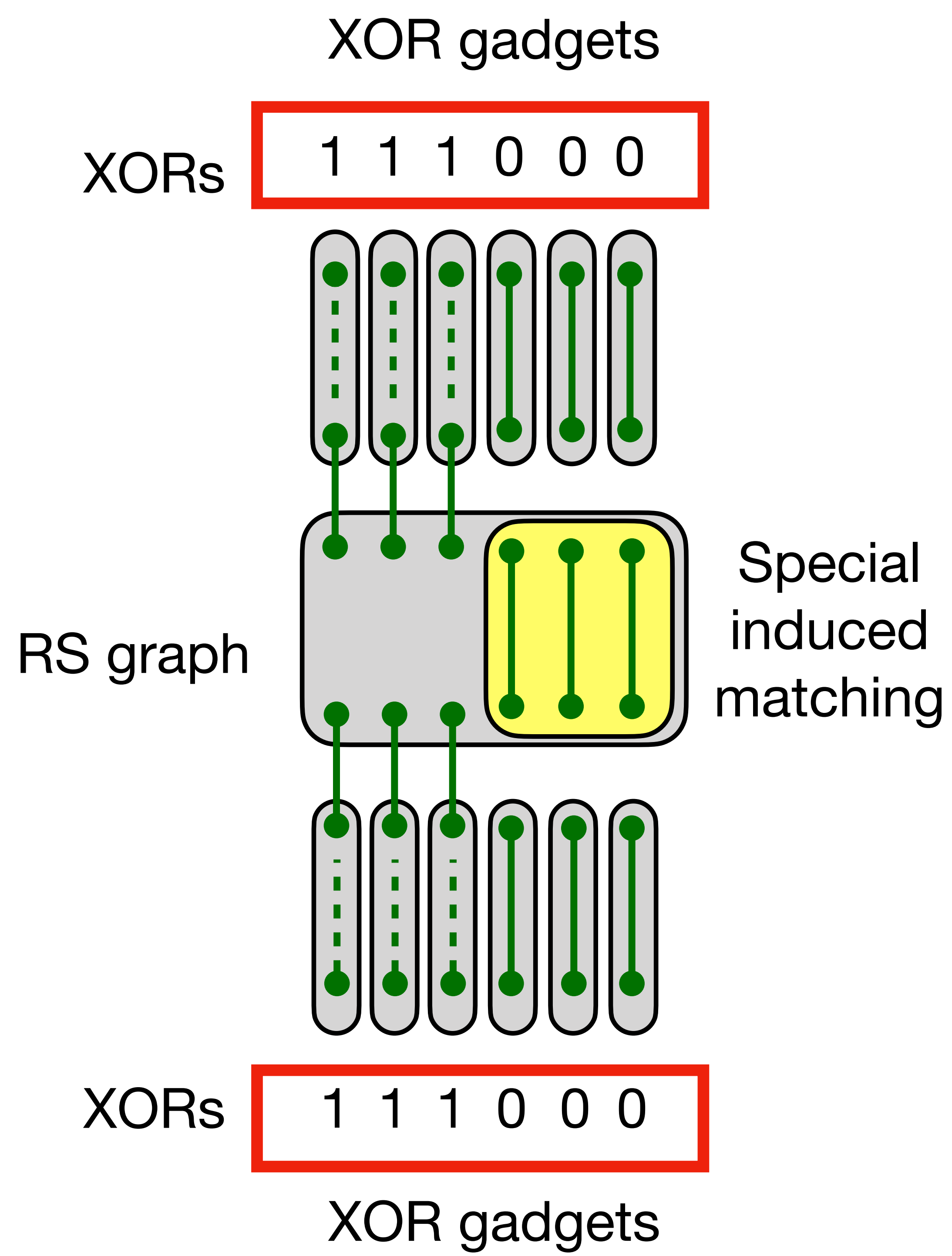
The algorithm still cannot "remember" these edges

The algorithm cannot get a $(1 - \Theta(\frac{1}{k}))$-approximation

k: length of XOR-gadgets

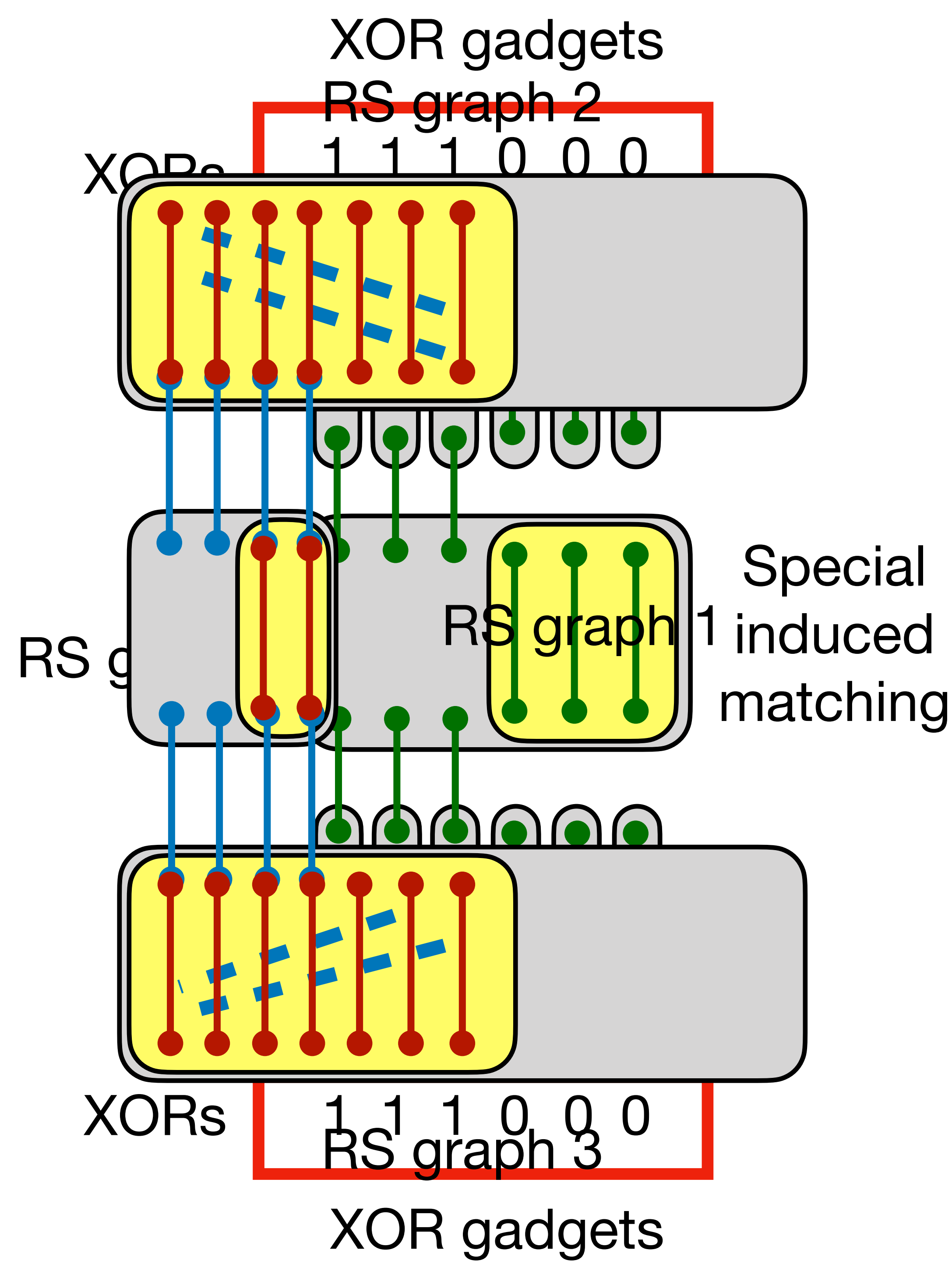
Challenge?

- We need a very strong lower bound for XOR gadgets
- Vector of values of XORs should remain almost random even after the first pass



Challenge?

- Proven using "XOR Lemmas"
  - Solving XOR of many independent problems becomes much harder
- Qualitatively different from prior approaches [**A**, Vishvajeet; 2021][Chen, Kol, Paramonov, Saxena, Song, Yu; 2021]
  - Our XOR problems are correlated by the choice of a single induced matching

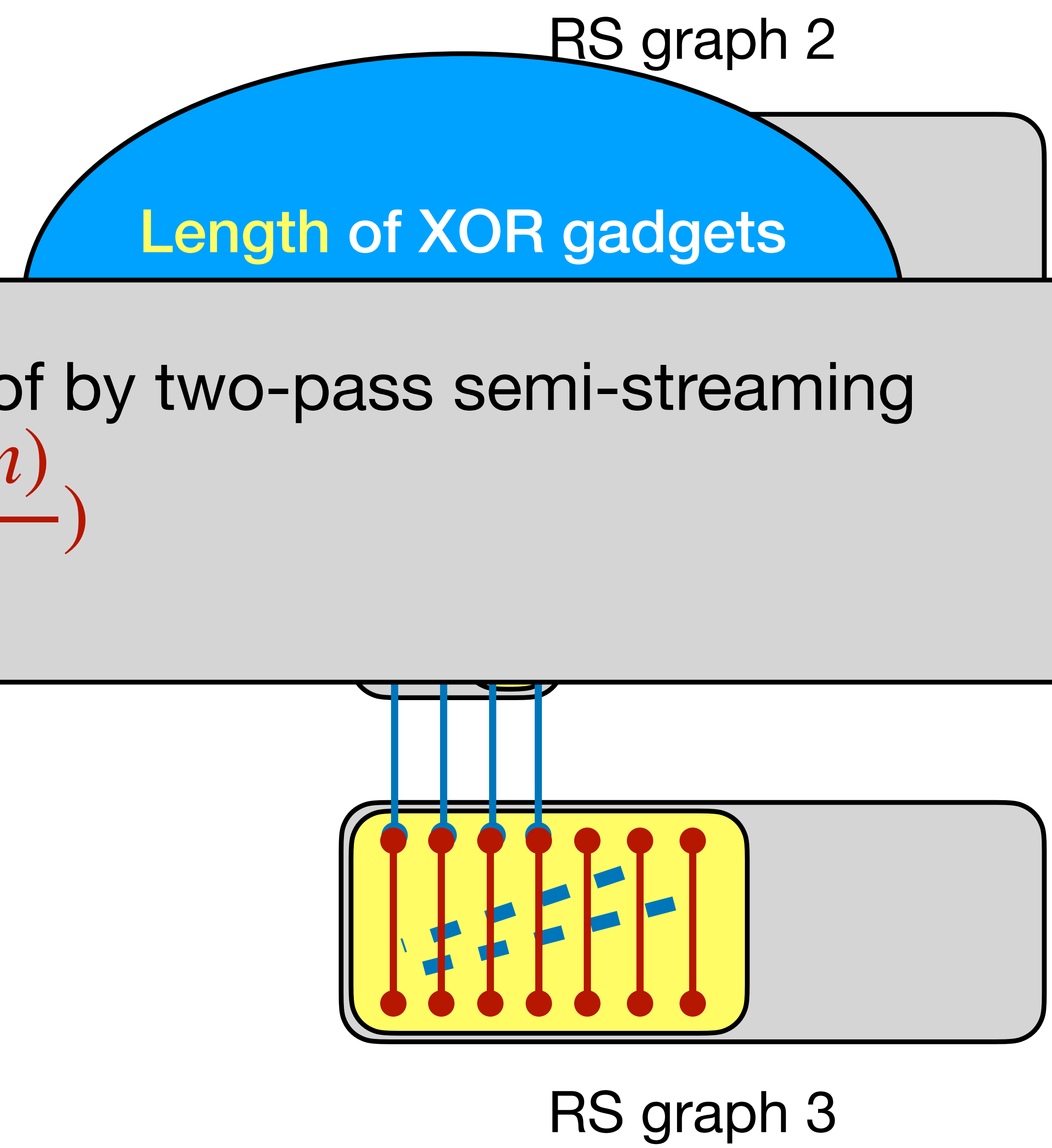


Proof Technique

- Information-complexity direct-sum arguments:

Theorem: Best approximation ratio of by two-pass semi-streaming matching is at most $1 - \Omega(\frac{\log \text{RS}(n)}{\log n})$

  - limited information about XOR gadgets leaves their XOR values almost random
  - [Gavinsky, Kempe, Kerenidis, Raz, de Wolf; 2007][Verbin, Yu; 2011]



Concluding Remarks

Concluding Remarks

$O(\frac{1}{\epsilon^2})$ passes

[A, Liu, Tarjan; 2021][Ahn, Guha; 2011]

$\tilde{O}(\frac{1}{\epsilon} \cdot \log n)$ passes

[A, Jambulapati, Jin, Sidford, Tian; 2022]

[Ahn, Guha; 2018]

...s semi-streaming

...$) \leq$ **RS**$(n) \leq \frac{n}{2^{O(\log^* n)}}$

...ation of matching in two pass... hard or just impossible

- **Open questions:**
  - Tighter lower bounds: can we prove < 0.9 approximation?
  - More passes: can we get $\Omega(\log(1/\epsilon))$-pass lower bound for $(1-\epsilon)$ -approximation?
  - Removing "conditioning" on RS graphs density?

Thank you!