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

Sepehr Assadi
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

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

- Up to $O(n^2)$ edges

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

- To edges $e_1, e_2, e_3, e_4, \ldots, e_m$
A well-studied but not well-understood parameter:

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

Theorem: Best approximation ratio possible by two-pass semi-streaming algorithms for maximum matching is at most:

\[
1 - \Omega\left(\frac{\log RS(n)}{\log n}\right)
\]

If the lower bound on \(RS(n)\) is tight:

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

- (a measure of) density of RS graphs

If the upper bound on \(RS(n)\) 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.

An induced matching
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 \( r \)

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 \(r\)

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 \( r \)

A \((3,4)\)-RS graph
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 \( r = \Theta(n) \)

\[ n^{\Omega\left(\frac{1}{\log \log n}\right)} \leq RS(n) \leq \frac{n}{2O(\log^* n)} \]
Ruzsa-Szemeredi (RS) Graphs

- RS graphs are studied extensively in TCS:
  - Property testing, PCP constructions, Graph sparsification, Streaming algorithms
- Used first in [Goel, Khanna, Kapralov; 2012] for semi-streaming matching problem
  - Subsequently in [Kapralov; 2013][Konrad; 2015][A, Khanna, Li, Yaroslavtsev; 2016][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 $1 - \Omega\left(\frac{\log \text{RS}(n)}{\log n}\right)$.

- **Conditional lower bound:**

\[ n^{\Omega\left(\frac{1}{\log \log n}\right)} \leq \text{RS}(n) \leq \frac{n}{2^{O(\log^* n)}} \]

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

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

- 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
A general goal of my research:

- Develop new techniques for multi-pass streaming lower bounds


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

- 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, Kapralov, Khanna; 2012]:

- First part of the stream is an \((r, t)\)-RS graph for
  \[ r < \frac{n}{4}, \quad t = n^{\Omega(\frac{1}{\log \log n})} \]

- Second part chooses an induced matching of the RS graph uniformly at random
- Connects all other vertices by a matching to outside

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

Semi-streaming algorithms cannot achieve better than \(\frac{2}{3}\) approximation
A Two-Pass Lower Bound?

- What happens to this family of instances in two passes?

  - It becomes super easy - Store all edges outside the RS graph in the first pass
  - 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
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

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 is zero: the unique maximum matching matches the target

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?

- Store all edges of XOR gadgets in the first pass
- 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 the entire gadgets inside two larger RS graphs.
- A new graph product that creates an XOR gadgets using multiple edges of a single induced matching.
- One needs the special induced matching to get a \(1 - \Theta\left(\frac{1}{k}\right)\)-approximation.

* Embed \(\frac{r}{k}\) gadgets of length \(k\) in an induced matching of size \(r\) in an \((r, t)\)-RS graph.
A Two-Pass Lower Bound?

Is this family of instances hard for two pass algorithms?

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 graphs 2 and 3

E. Then, remaining edges of XOR gadgets arrive

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

Identity of this induced matching is known but NOT its content

Even the identity of this induced matching is unknown
A Two-Pass Lower Bound?

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 graphs 2 and 3

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

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

The algorithm still cannot “remember” these edges
A Two-Pass Lower Bound?

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 graphs 2 and 3

The algorithm still cannot “remember” these edges

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

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

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

\(k\): length of XOR-gadgets

RS graph 1
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:
  - Limited information about XOR gadgets leaves their XOR values almost random

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

- [Gavinsky, Kempe, Kerenidis, Raz, de Wolf; 2007][Verbin, Yu; 2011]
Concluding Remarks
Concluding Remarks

- Takeaway: arbitrarily small-constant approximation of matching in two passes of semi-streaming is either quite hard or just impossible.

- Open questions:
  - Tighter lower bounds: can we prove \( \frac{1}{0.9} \) approximation?
  - More passes: can we get \( \Omega(\log(1/\epsilon)) \)-pass lower bound for \((1 - \epsilon)\) approximation?
  - Removing “conditioning” on RS graphs density?

\[
O\left(\frac{1}{\epsilon^2}\right) \text{ passes}
\]

\[
\tilde{O}\left(\frac{1}{\epsilon} \cdot \log n\right) \text{ passes}
\]

\[
[A, Liu, Tarjan; 2021][Ahn, Guha; 2011]
\
[A, Jambulapati, Jin, Sidford, Tian; 2022]
\[Ahn, Guha; 2018\]

\[
\Omega\left(\frac{n}{2O(\log^* n)}\right) \leq \text{RS}(n) \leq \frac{n}{2O(\log^* n)}
\]

Thank you!