## Constructing Large Matchings via Query Access to a Maximal Matching Oracle FSTTCS 2020

#### Lidiya Khalidah binti Khalil and Christian Konrad



Let G = (V, E) be a graph

- Let G = (V, E) be a graph
  - A matching  $M \subseteq E$  is a subset of non-adjacent edges







not a matching

- Let G = (V, E) be a graph
  - A matching  $M \subseteq E$  is a subset of non-adjacent edges



• *M* is *maximal* if  $M \cup \{e\}$  is not a matching, for every  $e \in E \setminus M$ 

maximal

not maximal

- Let G = (V, E) be a graph
  - A matching  $M \subseteq E$  is a subset of non-adjacent edges



• *M* is *maximal* if  $M \cup \{e\}$  is not a matching, for every  $e \in E \setminus M$ 





maximal

not maximal

•  $M^*$  is maximum if for every other matching  $M \subseteq E$ :  $|M^*| \ge |M|$ 



maximum

- Let G = (V, E) be a graph
  - A matching  $M \subseteq E$  is a subset of non-adjacent edges



• *M* is *maximal* if  $M \cup \{e\}$  is not a matching, for every  $e \in E \setminus M$ 





maximal

not maximal

•  $M^*$  is maximum if for every other matching  $M \subseteq E$ :  $|M^*| \ge |M|$ 



maximum

**Property:** 

 $|\mathsf{maximal\ matching}| \geq \frac{1}{2} |\mathsf{maximum\ matching}|$ 

**Goal:** Maximum Matching approximation (better than 1/2-approx.)

Goal: Maximum Matching approximation (better than 1/2-approx.)

#### In many computational models... (e.g. streaming, distributed models)

- computing maximal matchings is easy
- computing maximum matching approximations is more difficult

Goal: Maximum Matching approximation (better than 1/2-approx.)

#### In many computational models... (e.g. streaming, distributed models)

- computing maximal matchings is easy
- computing maximum matching approximations is more difficult

#### Edge-arrival Streaming Model:



Goal: Maximum Matching approximation (better than 1/2-approx.)

In many computational models... (e.g. streaming, distributed models)

- computing maximal matchings is easy
- computing maximum matching approximations is more difficult

#### Edge-arrival Streaming Model:



• Input stream: Sequence of edges of input graph G = (V, E) with n = |V| in arbitrary order

$$S = e_2 e_1 e_4 e_3$$



Goal: Maximum Matching approximation (better than 1/2-approx.)

In many computational models... (e.g. streaming, distributed models)

- computing maximal matchings is easy
- computing maximum matching approximations is more difficult

#### Edge-arrival Streaming Model:



• Input stream: Sequence of edges of input graph G = (V, E) with n = |V| in arbitrary order

$$S = e_2 e_1 e_4 e_3$$



• Goal: Few passes algorithms with small space

Goal: Maximum Matching approximation (better than 1/2-approx.)

In many computational models... (e.g. streaming, distributed models)

- computing maximal matchings is easy
- computing maximum matching approximations is more difficult

#### Edge-arrival Streaming Model:



• Input stream: Sequence of edges of input graph G = (V, E) with n = |V| in arbitrary order

$$S = e_2 e_1 e_4 e_3$$



- Goal: Few passes algorithms with small space
- Streaming Maximal Matching Algorithm: Insert current edge into initially empty matching if possible (GREEDY), using space Õ(n)

## State of the Art Streaming Matching Algorithms

# passes	Approximation	det/rand	Reference			
Bipartite Graphs						
1	$\frac{1}{2}$	det	$\operatorname{GREEDY}$ , folklore			
2	$\overline{2} - \sqrt{2} pprox 0.5857$	rand	Konrad '18			
3	0.6067	rand	Konrad '18			
$O(\frac{1}{\epsilon^2})$	$1-\epsilon$	det	Assadi, Liu, Tarjan '21			
General Graphs						
1	$\frac{1}{2}$	det	GREEDY, folklore			
2	Ō.53125	det	Kale and Tirodkar '17			
$\frac{1}{\epsilon} O(\frac{1}{\epsilon})$	$1-\epsilon$	det	Tirodkar '18			

## State of the Art Streaming Matching Algorithms

# passes	Approximation	det/rand	Reference			
Bipartite Graphs						
1	$\frac{1}{2}$	det	$\operatorname{GREEDY}$ , folklore			
2	$2-\sqrt{2}pprox 0.5857$	rand	Konrad '18			
3	0.6067	rand	Konrad '18			
$O(\frac{1}{\epsilon^2})$	$1-\epsilon$	det	Assadi, Liu, Tarjan '21			
General Graphs						
1	$\frac{1}{2}$	det	GREEDY, folklore			
2	Ō.53125	det	Kale and Tirodkar '17			
$\frac{1}{\epsilon} O(\frac{1}{\epsilon})$	$1-\epsilon$	det	Tirodkar '18			

Most of these algorithms (including previous works) solely run GREEDY in carefully selected subgraphs in each pass, thereby collecting edges and outputting the largest matching among the edges stored.

## State of the Art Streaming Matching Algorithms

# passes	Approximation	det/rand	Reference			
Bipartite Graphs						
1	$\frac{1}{2}$	det	$\operatorname{GREEDY}$ , folklore			
2	$\overline{2} - \sqrt{2} pprox 0.5857$	rand	Konrad '18			
3	0.6067	rand	Konrad '18			
$O(\frac{1}{\epsilon^2})$	$1-\epsilon$	det	Assadi, Liu, Tarjan '21			
General Graphs						
1	$\frac{1}{2}$	det	GREEDY, folklore			
2	Ō.53125	det	Kale and Tirodkar '17			
$\frac{1}{\epsilon} O(\frac{1}{\epsilon})$	$1-\epsilon$	det	Tirodkar '18			

Most of these algorithms (including previous works) solely run GREEDY in carefully selected subgraphs in each pass, thereby collecting edges and outputting the largest matching among the edges stored.

## How large a matching can we compute if we solely invoke Greedy in each pass?

## Maximal Matching Oracle

## Maximal Matching Oracle



- Player and oracle play r rounds of a "matching game"
- In each round r:
  - **1** Player queries a subset of vertices  $V_i \subseteq V$
  - 3 Oracle returns maximal matching  $M_i$  in induced subgraph  $G[V_i]$
- Player outputs largest matching in ∪<sub>1≤i≤r</sub> M<sub>i</sub>

## Maximal Matching Oracle



- Player and oracle play r rounds of a "matching game"
- In each round r:
  - **1** Player queries a subset of vertices  $V_i \subseteq V$
  - 3 Oracle returns maximal matching  $M_i$  in induced subgraph  $G[V_i]$
- Player outputs largest matching in ∪<sub>1≤i≤r</sub> M<sub>i</sub>

**Research Question:** What is the trade-off between the number of rounds and the approximation ratio?



# $\begin{array}{c} query(V_i) \\ \hline \\ Player \\ maximal matching in G[V_i] \end{array}$

#### Upper Bounds for Bipartite Graphs:

• 1 round: <sup>1</sup>/<sub>2</sub>-approximation query(V) yields maximal matching in input graph



- 1 round: <sup>1</sup>/<sub>2</sub>-approximation query(V) yields maximal matching in input graph
- 2 rounds:  $\geq \frac{1}{2}$ -approximation



- 1 round: <sup>1</sup>/<sub>2</sub>-approximation query(V) yields maximal matching in input graph
- 2 rounds:  $\geq \frac{1}{2}$ -approximation
- **3 rounds:** 3/5-approximation 3-pass streaming algorithm analysed by Kale and Tirodkar '17



- 1 round: <sup>1</sup>/<sub>2</sub>-approximation query(V) yields maximal matching in input graph
- 2 rounds:  $\geq \frac{1}{2}$ -approximation
- **3 rounds:** 3/5-approximation 3-pass streaming algorithm analysed by Kale and Tirodkar '17
- $\Theta(\frac{1}{\epsilon^6})$  rounds:  $(1 \epsilon)$ -approximation Streaming algorithm that runs in  $\Theta(\frac{1}{\epsilon^5})$  passes by Eggert et al. '12 can be adapted to the model



- 1 round: <sup>1</sup>/<sub>2</sub>-approximation query(V) yields maximal matching in input graph
- 2 rounds:  $\geq \frac{1}{2}$ -approximation
- **3 rounds:** 3/5-approximation 3-pass streaming algorithm analysed by Kale and Tirodkar '17
- $\Theta(\frac{1}{\epsilon^6})$  rounds:  $(1 \epsilon)$ -approximation Streaming algorithm that runs in  $\Theta(\frac{1}{\epsilon^5})$  passes by Eggert et al. '12 can be adapted to the model

#### **Upper Bounds for General Graphs:**



- 1 round: <sup>1</sup>/<sub>2</sub>-approximation query(V) yields maximal matching in input graph
- 2 rounds:  $\geq \frac{1}{2}$ -approximation
- **3 rounds:** 3/5-approximation 3-pass streaming algorithm analysed by Kale and Tirodkar '17
- $\Theta(\frac{1}{\epsilon^6})$  rounds:  $(1 \epsilon)$ -approximation Streaming algorithm that runs in  $\Theta(\frac{1}{\epsilon^5})$  passes by Eggert et al. '12 can be adapted to the model

**Upper Bounds for General Graphs:** X (except 1 round  $\frac{1}{2}$ -approx.)

We give the following Lower Bound Results:

We give the following Lower Bound Results:

#### **Bipartite Graphs**

• 1 round  $\frac{1}{2}$ -approximation:

We give the following Lower Bound Results:

#### **Bipartite Graphs**

• 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$ 

We give the following Lower Bound Results:

- 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$
- 2 rounds  $\geq \frac{1}{2}$ -approximation:

We give the following Lower Bound Results:

- 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$
- 2 rounds  $\geq \frac{1}{2}$ -approximation:  $\frac{1}{2}$  is best possible  $\checkmark$

We give the following Lower Bound Results:

- 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$
- 2 rounds  $\geq \frac{1}{2}$ -approximation:  $\frac{1}{2}$  is best possible  $\checkmark$
- 3 rounds  $\frac{3}{5}$ -approximation:

We give the following Lower Bound Results:

- 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$
- 2 rounds  $\geq \frac{1}{2}$ -approximation:  $\frac{1}{2}$  is best possible  $\checkmark$
- 3 rounds  $\frac{3}{5}$ -approximation: optimal  $\checkmark$

We give the following Lower Bound Results:

- 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$
- 2 rounds  $\geq \frac{1}{2}$ -approximation:  $\frac{1}{2}$  is best possible  $\checkmark$
- 3 rounds  $\frac{3}{5}$ -approximation: optimal  $\checkmark$
- $\Theta(\frac{1}{\epsilon^6})$  rounds  $(1 \epsilon)$ -approximation:

We give the following Lower Bound Results:

- 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$
- 2 rounds  $\geq \frac{1}{2}$ -approximation:  $\frac{1}{2}$  is best possible  $\checkmark$
- 3 rounds  $\frac{3}{5}$ -approximation: optimal  $\checkmark$
- $\Theta(\frac{1}{\epsilon^6})$  rounds  $(1 \epsilon)$ -approximation:  $\Omega(\frac{1}{\epsilon})$  rounds are needed for a  $(1 \epsilon)$ -approximation

We give the following Lower Bound Results:

#### **Bipartite Graphs**

- 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$
- 2 rounds  $\geq \frac{1}{2}$ -approximation:  $\frac{1}{2}$  is best possible  $\checkmark$
- 3 rounds  $\frac{3}{5}$ -approximation: optimal  $\checkmark$
- $\Theta(\frac{1}{\epsilon^6})$  rounds  $(1 \epsilon)$ -approximation:  $\Omega(\frac{1}{\epsilon})$  rounds are needed for a  $(1 \epsilon)$ -approximation

General Graphs (1 round 1/2-approximation)

We give the following Lower Bound Results:

#### **Bipartite Graphs**

- 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$
- 2 rounds  $\geq \frac{1}{2}$ -approximation:  $\frac{1}{2}$  is best possible  $\checkmark$
- 3 rounds  $\frac{3}{5}$ -approximation: optimal  $\checkmark$
- $\Theta(\frac{1}{\epsilon^6})$  rounds  $(1 \epsilon)$ -approximation:  $\Omega(\frac{1}{\epsilon})$  rounds are needed for a  $(1 \epsilon)$ -approximation

**General Graphs** (1 round 1/2-approximation)  $\Omega(n)$  rounds are needed for an approximation ratio  $\frac{1}{2} + \epsilon$ , for any  $\epsilon > 0$
We give the following Lower Bound Results:

#### **Bipartite Graphs**

- 1 round  $\frac{1}{2}$ -approximation: optimal  $\checkmark$
- 2 rounds  $\geq \frac{1}{2}$ -approximation:  $\frac{1}{2}$  is best possible  $\checkmark$
- 3 rounds  $\frac{3}{5}$ -approximation: optimal  $\checkmark$
- $\Theta(\frac{1}{\epsilon^6})$  rounds  $(1 \epsilon)$ -approximation:  $\Omega(\frac{1}{\epsilon})$  rounds are needed for a  $(1 \epsilon)$ -approximation

**General Graphs** (1 round 1/2-approximation)  $\Omega(n)$  rounds are needed for an approximation ratio  $\frac{1}{2} + \epsilon$ , for any  $\epsilon > 0$ 

#### **Outline:**

- $\Omega(\frac{1}{\epsilon})$  rounds are needed for a  $(1 \epsilon)$ -approximation
- **2**  $\Omega(n)$  rounds needed for  $(\frac{1}{2} + \epsilon)$ -approx. in general graphs,  $\epsilon > 0$
- O.6-approximation lower bound for 3 rounds (main technical result)



**Proof.** Consider semi-complete graph on 2c vertices

• Unique perfect matching  $M^*$  of size c

**Proof.** Consider semi-complete graph on 2c vertices



• Sub-optimal matching constitutes at best a  $\frac{c-1}{c} = (1 - \frac{1}{c})$ -approximation



- Unique perfect matching  $M^*$  of size c
- Sub-optimal matching constitutes at best a  $\frac{c-1}{c} = (1 \frac{1}{c})$ -approximation
- Insight: Any query gives at most one edge from  $M^*$



- Unique perfect matching  $M^*$  of size c
- Sub-optimal matching constitutes at best a  $\frac{c-1}{c} = (1 \frac{1}{c})$ -approximation
- Insight: Any query gives at most one edge from  $M^*$



- Unique perfect matching  $M^*$  of size c
- Sub-optimal matching constitutes at best a  $\frac{c-1}{c} = (1 \frac{1}{c})$ -approximation
- Insight: Any query gives at most one edge from  $M^*$
- Hence, to achieve a  $(1 \epsilon)$ -approximation, for  $\epsilon = \frac{1}{c+1}$ ,  $c = \frac{1}{c} 1$  queries are needed



- Unique perfect matching  $M^*$  of size c
- Sub-optimal matching constitutes at best a  $\frac{c-1}{c} = (1 \frac{1}{c})$ -approximation
- Insight: Any query gives at most one edge from M\*
- Hence, to achieve a  $(1 \epsilon)$ -approximation, for  $\epsilon = \frac{1}{c+1}$ ,
  - $c = rac{1}{\epsilon} 1$  queries are needed
- Use multiple disjoint gadgets for arbitrary n



- Unique perfect matching  $M^*$  of size c
- Sub-optimal matching constitutes at best a  $\frac{c-1}{c} = (1 \frac{1}{c})$ -approximation
- Insight: Any query gives at most one edge from M\*
- Hence, to achieve a  $(1 \epsilon)$ -approximation, for  $\epsilon = \frac{1}{c+1}$ ,
  - $c = \frac{1}{\epsilon} 1$  queries are needed
- Use multiple disjoint gadgets for arbitrary n

















#### Deterministic / Randomized Query Algorithms:

- Lower bounds on previous slides hold even if the input graph is known by the player
- They also hold for randomized query algorithms

#### Lower Bound for 3 Rounds on Bipartite Graphs:

- More subtle argument
- Oracle builds graph that depends on the queries
- Lower bound therefore only holds for deterministic algorithms

#### **Algorithm** (input graph G = (A, B, E))

**Algorithm** (input graph G = (A, B, E))

- $M \leftarrow \operatorname{query}(A \cup B)$
- $M_L \leftarrow \operatorname{query}(M(A) \cup \overline{M(B)})$
- $\ \, {\bf S} \ \, B'\subseteq B(M)\leftarrow {\rm endpoints} \ \, {\rm of} \ \, {\rm path} \ \, {\rm of} \ \, {\rm length} \ \, {\rm two} \ \, {\rm in} \ \, M\cup M_L$
- $M_R \leftarrow \operatorname{query}(B' \cup \overline{M(A)})$
- **5** return largest matching using edges  $M \cup M_L \cup M_R$

**Algorithm** (input graph G = (A, B, E))

- $M \leftarrow \operatorname{query}(A \cup B)$
- $M_L \leftarrow \operatorname{query}(M(A) \cup \overline{M(B)})$
- $\ \, {\bf S} \ \, B'\subseteq B(M)\leftarrow {\rm endpoints} \ \, {\rm of} \ \, {\rm path} \ \, {\rm of} \ \, {\rm length} \ \, {\rm two} \ \, {\rm in} \ \, M\cup M_L$
- $M_R \leftarrow \operatorname{query}(B' \cup \overline{M(A)})$
- **5** return largest matching using edges  $M \cup M_L \cup M_R$



Input graph G = (A, B, E) with perfect matching  $M^*$ 

**Algorithm** (input graph G = (A, B, E))

- $M \leftarrow query(A \cup B)$
- $M_L \leftarrow \operatorname{query}(M(A) \cup \overline{M(B)})$
- **③**  $B' \subseteq B(M)$  ← endpoints of path of length two in  $M \cup M_L$
- $M_R \leftarrow \operatorname{query}(B' \cup \overline{M(A)})$
- **5** return largest matching using edges  $M \cup M_L \cup M_R$



1st query: Matching M

**Algorithm** (input graph G = (A, B, E))

- $M \leftarrow query(A \cup B)$
- $M_L \leftarrow \operatorname{query}(M(A) \cup \overline{M(B)})$
- **③**  $B' \subseteq B(M)$  ← endpoints of path of length two in  $M \cup M_L$
- $M_R \leftarrow \operatorname{query}(B' \cup \overline{M(A)})$
- **5** return largest matching using edges  $M \cup M_L \cup M_R$



**Algorithm** (input graph G = (A, B, E))

- $M \leftarrow query(A \cup B)$
- $M_L \leftarrow \operatorname{query}(M(A) \cup \overline{M(B)})$
- **③**  $B' \subseteq B(M)$  ← endpoints of path of length two in  $M \cup M_L$
- $M_R \leftarrow \operatorname{query}(B' \cup \overline{M(A)})$
- **5** return largest matching using edges  $M \cup M_L \cup M_R$



2nd query: Matching  $M_L$ 

**Algorithm** (input graph G = (A, B, E))

- $M \leftarrow query(A \cup B)$
- $M_L \leftarrow \operatorname{query}(M(A) \cup \overline{M(B)})$
- $\ \, {\bf S} \ \, B'\subseteq B(M)\leftarrow {\rm endpoints} \ \, {\rm of} \ \, {\rm path} \ \, {\rm of} \ \, {\rm length} \ \, {\rm two} \ \, {\rm in} \ \, M\cup M_L$
- $M_R \leftarrow \operatorname{query}(B' \cup \overline{M(A)})$
- **5** return largest matching using edges  $M \cup M_L \cup M_R$



**Algorithm** (input graph G = (A, B, E))

- $M \leftarrow query(A \cup B)$
- $M_L \leftarrow \operatorname{query}(M(A) \cup \overline{M(B)})$
- **③**  $B' \subseteq B(M)$  ← endpoints of path of length two in  $M \cup M_L$
- $M_R \leftarrow \operatorname{query}(B' \cup \overline{M(A)})$
- **5** return largest matching using edges  $M \cup M_L \cup M_R$



Matching  $M_R$ 

**Algorithm** (input graph G = (A, B, E))

- $M \leftarrow \operatorname{query}(A \cup B)$
- $M_L \leftarrow \operatorname{query}(M(A) \cup \overline{M(B)})$
- $\ \, {\bf S} \ \, B'\subseteq B(M)\leftarrow {\rm endpoints} \ \, {\rm of} \ \, {\rm path} \ \, {\rm of} \ \, {\rm length} \ \, {\rm two} \ \, {\rm in} \ \, M\cup M_L$
- $M_R \leftarrow \operatorname{query}(B' \cup \overline{M(A)})$
- **5** return largest matching using edges  $M \cup M_L \cup M_R$



Largest matching in  $M \cup M_L \cup M_R$  (*M* augmented with  $M_L \cup M_R$ )

Analysis:

**Analysis:**  $\frac{3}{5}$ -approximation algorithm [Kale and Tirodkar, '17]

Analysis:  $\frac{3}{5}$ -approximation algorithm [Kale and Tirodkar, '17]

Worst-case Example:

Analysis:  $\frac{3}{5}$ -approximation algorithm [Kale and Tirodkar, '17]

Worst-case Example:



- $M \leftarrow query(A \cup B)$
- $M_L \leftarrow \operatorname{query}(M(A) \cup \overline{M(B)})$
- **③** B' ⊆ B(M) ← endpoints of path of length two in  $M ∪ M_L$
- $M_R \leftarrow \mathsf{query}(B' \cup \overline{M(A)})$

#### Lower Bound Construction - First Query

**Strategy:** Bound "knowledge" about input graph after each query ("structure graph"); ensure perfect matching can be added

#### Lower Bound Construction - First Query

**Strategy:** Bound "knowledge" about input graph after each query ("structure graph"); ensure perfect matching can be added

#### First Query:

- Oracle commits to structure below and returns subset of edges M (no edges between  $A_{out}$  and  $B_{out}$ )
- A perfect matching (blue edges) can be added, which implies that approximation factor is 3/5 at best after first query



#### Second Query:

- Information can be bounded by structure below grey edges indicate that edges are not present in output graph
- Again, perfect matching can be added, which implies that approximation factor is 3/5 at best after second query


## Third Query:

- Structure cannot easily be captured using a single "structure graph"
- Instead, case distinctions with cleverly grouping cases together

## Third Query:

- Structure cannot easily be captured using a single "structure graph"
- Instead, case distinctions with cleverly grouping cases together

**Example Case:** Query includes  $\{b_1, b_2, b_3\}$ 



## Third Query:

- Structure cannot easily be captured using a single "structure graph"
- Instead, case distinctions with cleverly grouping cases together

**Example Case:** Query includes  $\{b_1, b_2, b_3\}$ 



## Third Query:

- Structure cannot easily be captured using a single "structure graph"
- Instead, case distinctions with cleverly grouping cases together

**Example Case:** Query includes  $\{b_1, b_2, b_3\}$ 



Key Technique: Structural properties that allow eliminating cases

• Can we compute a Maximum Matching in  $o(n^2)$  rounds?

- Can we compute a Maximum Matching in  $o(n^2)$  rounds?
- Can we prove that  $\Omega(1/\epsilon^2)$  rounds are required for computing a  $(1-\epsilon)$ -approximation?

- Can we compute a Maximum Matching in  $o(n^2)$  rounds?
- Can we prove that  $\Omega(1/\epsilon^2)$  rounds are required for computing a  $(1-\epsilon)$ -approximation?

## **Outlook:**

- Can we compute a Maximum Matching in  $o(n^2)$  rounds?
- Can we prove that  $\Omega(1/\epsilon^2)$  rounds are required for computing a  $(1-\epsilon)$ -approximation?

# Outlook:

• Extensions: Edge queries instead of vertex queries

- Can we compute a Maximum Matching in  $o(n^2)$  rounds?
- Can we prove that  $\Omega(1/\epsilon^2)$  rounds are required for computing a  $(1-\epsilon)$ -approximation?

# Outlook:

- Extensions: Edge queries instead of vertex queries
- Randomization?

# Thank you for your attention.