$4(1+\varepsilon)$-Approximate Semi-streaming Algorithm for Maximum Weight Matching

Fahad Panolan

Department of Computer Science and Engineering
Indian Institute of Technology Hyderabad, India

10 December 2022

Graph Streaming

- $G = (V, E)$ is an undirected graph
- $n = |V|$ and $m = |E|$
- Edges $e_1, e_2, \ldots, e_m$ seen as a stream, n known

Graph Streaming

- $G = (V, E)$ is an undirected graph
- $n = |V|$ and $m = |E|$
- Edges $e_1, e_2, \ldots, e_m$ seen as a stream, n known
- Massive graphs include social networks, web graph, call graphs, etc

Graph Streaming

- $G = (V, E)$ is an undirected graph
- $n = |V|$ and $m = |E|$
- Edges $e_1, e_2, \ldots, e_m$ seen as a stream, n known
- Massive graphs include social networks, web graph, call graphs, etc

Question:

- What graph problems can be solve with small space?

Semi-streaming Model

- Lower bounds show that we require $\Omega(n)$ memory

Semi-streaming Model

- Lower bounds show that we require $\Omega(n)$ memory
- What can we compute about G in $o(m)$ space?
- Assume we have $\Theta(n \cdot \text{polylog}(n))$ memory. About polylog per vertex of the graph
- Can solve several interesting problems. Essentially reduce dense graphs to sparse graphs.

Matchings

Definition

- A matching $M \subseteq E$ in a graph $G = (V, E)$ is a set of edges that do not intersect.
- A matching $M \subseteq E$ in a graph $G = (V, E)$ is a perfect matching if all the vertices are matched.

Matchings

Definition

- A matching $M \subseteq E$ in a graph $G = (V, E)$ is a set of edges that do not intersect.
- A matching $M \subseteq E$ in a graph $G = (V, E)$ is a perfect matching if all the vertices are matched.

- Given a graph G does it have a perfect matching?
- Find a maximum cardinality matching.
- Find a maximum weight matching.
- Count number of (perfect) matchings.

Streaming Lower bound One pass streaming algorithm for PERFECT MATCHING requires $\Omega(n^2)$ space.

Semi-streaming setting

Edges $e_1, e_2, \ldots, e_m$ come in some order

Questions:

- With $O(n \cdot polylog(n))$ memory approximate maximum cardinality matching
- With $O(n \cdot polylog(n))$ memory approximate maximum weight matching

Maximum cardinality

Definition

A matching M is maximal if for all $e \in E \setminus M$, $M \cup \{e\}$ is not a matching.

Lemma

If M is maximal, then $|M| \geq |M^\star|/2$ for any matching $M^\star$. Hence, a maximal matching is a 2-approximation.

Maximal matching in streams

Algorithm: GREEDY MATCHING

Result: A maximal matching

$M = \emptyset$;

while an edge e arrives **do**

  if $(M \cup \{e\})$ is a matching **then**

    $M \leftarrow M \cup \{e\}$;

  end

end

return M;

Maximum-weight matching

Offline algorithm: greedy after sorting.

Algorithm: Max-weight Matching

Sort edges such that $w(e_1) \geq w(e_2) \geq \ldots \geq w(e_m)$;
$M = \emptyset$;
for $i = 1$ to m **do**
  if $(M \cup \{e_i\})$ is a matching **then**
    $M \leftarrow M \cup \{e_i\}$;
  end
end
return M;

Lemma

$w(M) \geq w(M^\star)/2$.

Maximum-weight matching

Offline algorithm: greedy after sorting.

Algorithm: Max-weight Matching

Sort edges such that $w(e_1) \geq w(e_2) \geq \ldots \geq w(e_m)$;
$M = \emptyset$;
for $i = 1$ to m **do**
  if $(M \cup \{e_i\})$ is a matching **then**
    $M \leftarrow M \cup \{e_i\}$;
  end
end
return M;

Lemma

$w(M) \geq w(M^\star)/2$.

Streaming setting? Cannot sort!

$4(1+\varepsilon)$-Approx Maximum Weight Matching

- Input: $G = (V, E)$ and $w \colon E \mapsto \{1, \ldots, W\}$.

[Crouch and Stubbs, 2014]

$4(1+\varepsilon)$-Approx Maximum Weight Matching

- Input: $G = (V, E)$ and $w \colon E \mapsto \{1, \ldots, W\}$.
- Let $q = \log_{1+\varepsilon} W$.

[Crouch and Stubbs, 2014]

$4(1+\varepsilon)$-Approx Maximum Weight Matching

- Input: $G = (V, E)$ and $w\colon E \mapsto \{1, \ldots, W\}$.
- Let $q = \log_{1+\varepsilon} W$.
- For each $i \in \{0, 1, \ldots, q\}$, $E_i = \{e \,:\, w(e) \geq (1+\varepsilon)^i\}$

[Crouch and Stubbs, 2014]

$4(1+\varepsilon)$-Approx Maximum Weight Matching

- Input: $G = (V, E)$ and $w \colon E \mapsto \{1, \ldots, W\}$.
- Let $q = \log_{1+\varepsilon} W$.
- For each $i \in \{0, 1, \ldots, q\}$, $E_i = \{e \ : \ w(e) \geq (1+\varepsilon)^i\}$
- Notice that $E_0 \supseteq E_1 \supseteq E_2 \supseteq \ldots \supseteq E_q$.

[Crouch and Stubbs, 2014]

$4(1+\varepsilon)$-Approx Maximum Weight Matching

- Input: $G = (V, E)$ and $w \colon E \mapsto \{1, \ldots, W\}$.
- Let $q = \log_{1+\varepsilon} W$.
- For each $i \in \{0, 1, \ldots, q\}$, $E_i = \{e : \; w(e) \geq (1+\varepsilon)^i\}$
- Notice that $E_0 \supseteq E_1 \supseteq E_2 \supseteq \ldots \supseteq E_q$.
- Let C_i be the output of Greedy Matching on (V, E_i).

[Crouch and Stubbs, 2014]

$4(1+\varepsilon)$-Approx Maximum Weight Matching

- Input: $G = (V, E)$ and $w\colon E \mapsto \{1, \ldots, W\}$.
- Let $q = \log_{1+\varepsilon} W$.
- For each $i \in \{0, 1, \ldots, q\}$, $E_i = \{e : \ w(e) \geq (1+\varepsilon)^i\}$
- Notice that $E_0 \supseteq E_1 \supseteq E_2 \supseteq \ldots \supseteq E_q$.
- Let C_i be the output of GREEDY MATCHING on (V, E_i).
- Let M be the output of GREEDY MATCHING on the stream $C_q, C_{q-1}, \ldots, C_0$.

[Crouch and Stubbs, 2014]

$4(1+\varepsilon)$-Approx Maximum Weight Matching

- Input: $G = (V, E)$ and $w\colon E \mapsto \{1, \ldots, W\}$.
- Let $q = \log_{1+\varepsilon} W$.
- For each $i \in \{0, 1, \ldots, q\}$, $E_i = \{e : w(e) \geq (1+\varepsilon)^i\}$
- Notice that $E_0 \supseteq E_1 \supseteq E_2 \supseteq \ldots \supseteq E_q$.
- Let C_i be the output of Greedy Matching on (V, E_i).
- Let M be the output of Greedy Matching on the stream $C_q, C_{q-1}, \ldots, C_0$.
- Output M

[Crouch and Stubbs, 2014]

Space Complexity

The above algorithm can be implemented in the streaming setting with space $O(\log_{1+\varepsilon} W \cdot n \log n)$.

Correctness Proof

- Let $M^\star$ be a maximum weight matching.
  We want to prove that $w(M^\star) \leq 4(1+\varepsilon)w(M)$
- Let $M_i = M \cap E_i$ and $M_i^\star = M^\star \cap E_i$

Claim 1: $|M_i| \geq \frac{1}{4}|M_i^\star|$

- M_i is a maximal matching in $F_i = C_q \cup C_{q-1} \ldots \cup C_i$. And C_i is a matching in F_i.

Claim 1: $|M_i| \geq \frac{1}{4}|M_i^\star|$

- M_i is a maximal matching in $F_i = C_q \cup C_{q-1} \ldots \cup C_i$. And C_i is a matching in F_i.
- Hence, $|M_i| \geq \frac{1}{2}|C_i|$.

Claim 1: $|M_i| \geq \frac{1}{4}|M_i^\star|$

- M_i is a maximal matching in $F_i = C_q \cup C_{q-1} \ldots \cup C_i$. And C_i is a matching in F_i.
- Hence, $|M_i| \geq \frac{1}{2}|C_i|$.
- C_i is a maximal matching in E_i and $M_i^\star$ is a matching in E_i.

Claim 1: $|M_i| \geq \frac{1}{4}|M_i^\star|$

- M_i is a maximal matching in $F_i = C_q \cup C_{q-1} \ldots \cup C_i$. And C_i is a matching in F_i.
- Hence, $|M_i| \geq \frac{1}{2}|C_i|$.
- C_i is a maximal matching in E_i and $M_i^\star$ is a matching in E_i.
- Hence, $|C_i| \geq \frac{1}{2}|M_i^\star|$.

Now, the claim follows.

Claim 2

There is a function $f\colon M^\star \mapsto M$ such that

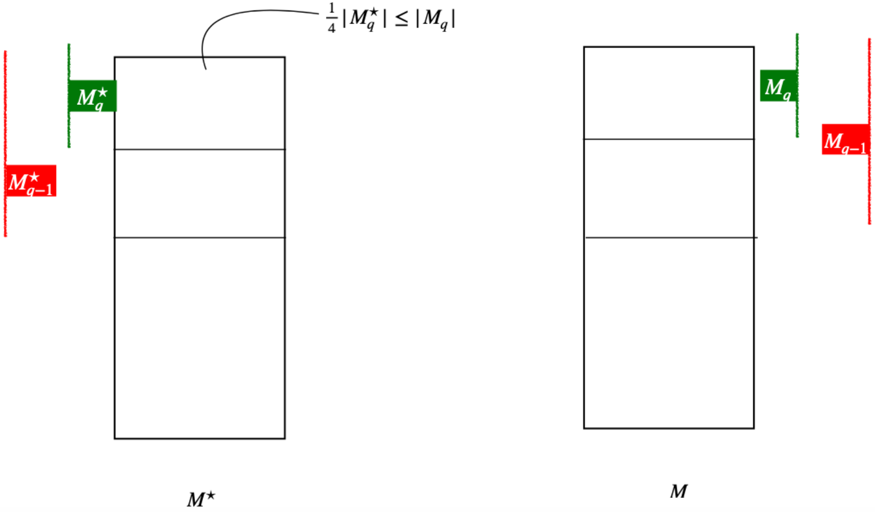
- $\frac{w(e)}{(1+\epsilon)} \leq w(f(e))$ for all $e \in M^\star$
- $|f^{-1}(t)| \leq 4$ for all $t \in M$

Claim 2

There is a function $f\colon M^\star \mapsto M$ such that

- $\frac{w(e)}{(1+\epsilon)} \leq w(f(e))$ for all $e \in M^\star$
- $|f^{-1}(t)| \leq 4$ for all $t \in M$

For each $e \in M^\star$, let j be the largest integer such that $e \in E_j$. Map e to an edge $t \in M_j$ such that at most 4 edges are mapped to t.



$\frac{1}{4}|M_q^\star| \leq |M_q|$
$M_q^\star$
$M_{q-1}^\star$
M_q
M_{q-1}
$M^\star$
M

Final step

There is a function $f\colon M^\star \mapsto M$ such that

- $w(e) \leq (1+\epsilon)w(f(e))$ for all $e \in M^\star$
- $|f^{-1}(t)| \leq 4$ for all $t \in M$

Final step

There is a function $f\colon M^\star \mapsto M$ such that

- $w(e) \leq (1+\epsilon)w(f(e))$ for all $e \in M^\star$
- $|f^{-1}(t)| \leq 4$ for all $t \in M$

Final step

There is a function $f\colon M^\star \mapsto M$ such that

- $w(e) \leq (1+\epsilon)w(f(e))$ for all $e \in M^\star$
- $|f^{-1}(t)| \leq 4$ for all $t \in M$

$$w(M^\star) = \sum_{e \in M^\star} w(e)$$

Final step

There is a function $f\colon M^\star \mapsto M$ such that

- $w(e) \leq (1+\epsilon)w(f(e))$ for all $e \in M^\star$
- $|f^{-1}(t)| \leq 4$ for all $t \in M$

$$
\begin{aligned}
w(M^\star) &= \sum_{e \in M^\star} w(e) \\
&\leq \sum_{e \in M^\star} (1+\varepsilon) w(f(e))
\end{aligned}
$$

Final step

There is a function $f\colon M^\star \mapsto M$ such that

- $w(e) \leq (1+\epsilon)w(f(e))$ for all $e \in M^\star$
- $|f^{-1}(t)| \leq 4$ for all $t \in M$

$$
\begin{aligned}
w(M^\star) &= \sum_{e \in M^\star} w(e) \\
&\leq \sum_{e \in M^\star} (1+\varepsilon) w(f(e)) \\
&\leq (1+\varepsilon) \sum_{t \in M} |f^{-1}(t)| \cdot w(t)
\end{aligned}
$$

Final step

There is a function $f : M^\star \mapsto M$ such that

- $w(e) \leq (1+\epsilon) w(f(e))$ for all $e \in M^\star$
- $|f^{-1}(t)| \leq 4$ for all $t \in M$

$$
\begin{aligned}
w(M^\star) &= \sum_{e \in M^\star} w(e) \\
&\leq \sum_{e \in M^\star} (1+\varepsilon) w(f(e)) \\
&\leq (1+\varepsilon) \sum_{t \in M} |f^{-1}(t)| \cdot w(t) \\
&\leq (1+\varepsilon) \sum_{t \in M} 4 \cdot w(t)
\end{aligned}
$$

Final step

There is a function $f\colon M^\star \mapsto M$ such that

- $w(e) \leq (1+\epsilon)w(f(e))$ for all $e \in M^\star$
- $|f^{-1}(t)| \leq 4$ for all $t \in M$

$$
\begin{aligned}
w(M^\star) &= \sum_{e \in M^\star} w(e) \\
&\leq \sum_{e \in M^\star} (1+\varepsilon) w(f(e)) \\
&\leq (1+\varepsilon) \sum_{t \in M} |f^{-1}(t)| \cdot w(t) \\
&\leq (1+\varepsilon) \sum_{t \in M} 4 \cdot w(t) \\
&\leq 4(1+\varepsilon) \cdot w(M)
\end{aligned}
$$

Conclusion

- We saw a $(4 + \epsilon)$-approx streaming algorithm for weighted matching
- Best known factor: $(2 + \epsilon)$ [Paz and Schwartzman, 2017]

Conclusion

- We saw a $(4+\epsilon)$-approx streaming algorithm for weighted matching
- Best known factor: $(2+\epsilon)$ [Paz and Schwartzman, 2017]

Thank You.