

# Approximate Maximum Matching in Random Streams

Alireza Farhadi\*      MohammadTaghi Hajiaghayi\*      Tung Mai†  
Anup Rao†      Ryan A. Rossi†

## Abstract

In this paper, we study the problem of finding a maximum matching in the semi-streaming model when edges arrive in a random order. In the semi-streaming model, an algorithm receives a stream of edges and it is allowed to have a memory of  $\tilde{O}(n)$ <sup>1</sup> where  $n$  is the number of vertices in the graph. A recent inspiring work by Assadi et al. [1] shows that there exists a streaming algorithm with the approximation ratio of  $\frac{2}{3}$  that uses  $\tilde{O}(n^{1.5})$  memory. However, the memory of their algorithm is much larger than the memory constraint of the semi-streaming algorithms. In this work, we further investigate this problem in the semi-streaming model, and we present simple algorithms for approximating maximum matching in the semi-streaming model. Our main results are as follows.

- We show that there exists a single-pass deterministic semi-streaming algorithm that finds a  $\frac{3}{5}$  ( $= 0.6$ ) approximation of the maximum matching in bipartite graphs using  $\tilde{O}(n)$  memory. This result significantly outperforms the state-of-the-art result of Konrad [15] that finds a 0.539 approximation of the maximum matching using  $\tilde{O}(n)$  memory.
- By giving a black-box reduction from finding a matching in general graphs to finding a matching in bipartite graphs, we show there exists a single-pass deterministic semi-streaming algorithm that finds a  $\frac{6}{11}$  ( $\approx 0.545$ ) approximation of the maximum matching in general graphs, improving upon the state-of-art result 0.506 approximation by Gamlath et al. [9].

## 1 Introduction

As noted by Lovász and Plummer in their book [19] “Matching Theory is a central part of graph theory, not only because of its applications, but also because it is the source of important ideas developed during the rapid growth of combinatorics during the last several decades”. Classical algorithms for finding a matching assume that we have enough memory to store all of the vertices and edges of the graph in the memory. However, in the era of big data, we often face with massive datasets that are too large to fit into the computer’s memory. *Semi-streaming* is one of the models introduced to deal with massive inputs [7]. In this model, the algorithm has access to a sequence of edges and it is allowed to have a memory of  $\tilde{O}(n)$  where  $n$  is the number of vertices in the graph. Finding a large matching as a fundamental problem in graph theory has been extensively studied in the semi-streaming model.

When edges of the graph arrive in an arbitrary order, it is known that a simple algorithm that adds edges to the matching in a greedy manner, can find a maximal matching, and in fact,  $1/2$  approximation of the maximum matching using  $\tilde{O}(n)$  memory. However, whether or not we can get better than  $1/2$  approximation with  $\tilde{O}(n^{2-\epsilon})$  memory for a constant  $\epsilon > 0$  has remained an

---

\*University of Maryland. Email: {farhadi,hajiagha}@cs.umd.edu.

†Adobe Research. Email: {tumai,anuprao,rrossi}@adobe.com.

<sup>1</sup>We use  $\tilde{O}(\cdot)$  notion to hide logarithmic factors.

	APPROXIMATION FACTOR		SPACE
	Bipartite graphs	General graphs	
Kapralov et al. [14]	$\frac{1}{\text{polylog } n}$	$\frac{1}{\text{polylog } n}$	$O(\text{polylog } n)$
Konrad et al. [16]	0.5005	0.5003	$\tilde{O}(n)$
Gamlath et al. [9]	0.512	0.506	$\tilde{O}(n)$
Konrad [15]	0.539	-	$\tilde{O}(n)$
<b>This paper</b>	<b>0.6</b>	<b>0.545</b>	$\tilde{O}(n)$

Table 1: Single-pass semi-streaming algorithms known for the maximum matching when edges arrive in a random order.

important open question since the introduction of the semi-streaming model [7]. It is shown in a series of works [11, 13] that there is no semi-streaming algorithm for the maximum matching problem with a approximation ratio better than  $1 - 1/e$ . Specifically, a work by Kapralov [13] shows that any streaming algorithm with the approximation ratio of better than  $1 - 1/e$  requires  $n^{1+\Omega(1/\log \log n)}$  memory. However, this still leaves room between  $1/2$  and  $1 - 1/e$ . In order to make progress on this long-standing open problem, an influential work by Konrad et al. [16] studied the problem when edges arrive in a uniformly random order. Despite the fact that the approximation ratio of the greedy algorithm could be  $1/2 + o(1)$  in this setting, they surprisingly showed there exists an algorithm that achieves the approximation ratio of  $1/2 + 0.003$  using  $\tilde{O}(n)$  memory. A recent inspiring work of Assadi et al. [1] uses *edge degree constrained subgraphs* (EDCS) introduced by Bernstein and Stein [2] to design streaming algorithms and they show that there exists a streaming algorithm for the maximum matching problem with the approximation ratio of  $2/3$  that uses  $\tilde{O}(n^{1.5})$  memory. However, the memory of their algorithm is not within memory bounds of semi-streaming algorithms.

In this work, we further study the problem of finding a maximum matching in random arrival of edges. In this model, Konrad et al. [16] showed that the  $1/2$  barrier can be broken, and they designed a semi-streaming algorithm with the approximation ratio of  $0.5003$  in general graphs and  $0.5005$  in bipartite graphs. These approximation factors are later improved to  $0.539$  in bipartite graphs and  $0.506$  in general graphs [9, 15]. Also, a surprising result by Kapralov et al. [14] showed that when edges arrive in a random order, there exists a streaming algorithm that finds  $(1/\text{polylog } n)$  approximation of the size of a maximum matching using only poly-logarithmic space. (Table 1) gives a brief survey of known results in a random stream of edges as well as a comparison to our results.

We improve the state-of-the-art results in both bipartite and general graphs when edges arrive in a random order. We show that there exists a single-pass semi-streaming algorithm that finds a  $0.6$  approximation of the maximum matching in bipartite graphs.

**Theorem 1.1.** *There exists a deterministic single-pass semi-streaming algorithm that w.h.p.<sup>2</sup> finds a  $(\frac{3}{5} - o(1))$  approximate of the maximum matching in bipartite graphs when edges arrive in a uniformly random order.*

This result significantly improves the randomized algorithm of [15] that finds a  $0.539$  approximation of the maximum matching in bipartite graphs. We also give a black-box reduction from finding a matching in general graphs in the semi-streaming model to the problem of finding a matching in bipartite graphs. Our reduction result is as follows.

<sup>2</sup>With high probability

**Theorem 1.2.** *Given any single-pass semi-streaming algorithm with the approximation ratio of  $p$  for finding a maximum matching in bipartite graphs in random arrival of edges, there exists a semi-streaming algorithm that finds w.h.p. a  $(\frac{2p}{2p+1} - o(1))$  approximation of the maximum matching when edges arrive in a random order.*

Using this reduction and Theorem 1.1, we can immediately get the following result.

**Theorem 1.3.** *There exists a deterministic single-pass semi-streaming algorithm that finds w.h.p. a  $(\frac{6}{11} - o(1)) (\approx 0.545)$  approximation of the maximum matching in general graphs when edges arrive in a random order.*

## 1.1 Related Work

In case of weighted graphs, the problem of finding a maximum weighted matching first considered by Feigenbaum et al. [8]. They showed that a modification of the greedy algorithm can find a  $1/6$  approximation of the maximum weighted matching in the semi-streaming model. This bound later improved in a series of works [17, 21, 5, 4, 18, 10] to  $(1/2 - \epsilon)$  approximation. A recent work by Gamlath et al. [9] shows that when edges arrive in a random order, there exists a semi-streaming algorithm that finds a  $(1/2 + \Omega(1))$  approximation of the maximum weighted matching, beating the  $1/2$  barrier for the weighted case.

Another line of research considers the problem of finding an approximate matching in the semi-streaming model using a very small number of passes [6, 12].

## 1.2 Overview of Our Algorithm

Our algorithm starts by using a very small fraction of the edges and gives these edges to the greedy algorithm to find a matching  $\mathcal{M}_0$ . Let  $R$  be the set of edges in the graph that do not share a vertex with the edges in  $\mathcal{M}_0$ . Since the order of edges is uniformly random, it can be shown that (formalized as in Lemma 3.2) the number of edges in  $R$  should be small and we can fit all of them in the semi-streaming memory of our algorithm. Thus, we can find the maximum matching of  $R$  using any offline algorithm. For now, just suppose that we want to break the 0.5-approximation barrier. Let  $\mu$  be the size of maximum matching in the graph. There are at most  $2|\mathcal{M}_0|$  matching edges that share a vertex with the edges in  $\mathcal{M}_0$ . Therefore, the size of the maximum matching in  $R$  is at least  $\mu - 2|\mathcal{M}_0|$ . Combining this matching with the matching  $\mathcal{M}_0$  gives us a matching of size at least  $\mu - |\mathcal{M}_0|$ . Therefore, we can get a matching of size  $\max\{|\mathcal{M}_0|, \mu - |\mathcal{M}_0|\}$  using this simple algorithm. The only case that we do not break the 0.5-approximation barrier is when we roughly have  $|\mathcal{M}_0| = \mu/2$ . Also, the size of a maximum matching in  $R$  should be almost zero. This implies that  $\mathcal{M}_0$  is almost a maximal matching.

Therefore, in the worst case scenario of the mentioned algorithm, the algorithm can find a maximal matching  $\mathcal{M}_0$  by looking at a very small fraction of the edges. Since  $|\mathcal{M}_0| = \mu/2$ , there are  $|\mathcal{M}_0|$  edge disjoint 3-augmenting paths for the edges in  $\mathcal{M}_0$ . All of the previous works in the semi-streaming model use the remaining edges in the streaming to find augmenting paths for  $\mathcal{M}_0$  and increase the size of the matching. Our work is different from them from several perspectives.

- All of the previous works, only use the fact that order of the edges is random to argue that the size of  $\mathcal{M}_0$  is very close to the half of the size of a maximum matching. For finding augmenting paths, they use algorithms that also work when order of edges is adversarial. However, finding augmenting paths is much trickier than finding a matching in graph, and in fact, finding a matching in a graph can easily be reduced to finding augmenting paths. It seems the assumption that order of edges is uniformly at random makes the problem of finding a matching easier. Therefore, this assumption should also helps us to find a better set

of augmenting paths. Our algorithm is a recursive algorithm that extensively uses the fact that the order of edges is random, and it finds much better set of augmenting paths.

- Unlike the previous works on this problem, our algorithm finds augmenting paths with a length larger than 3. We first propose an algorithm for finding 3-augmenting paths, and we show that the approximation ratio of this algorithm is roughly  $4/7$  in bipartite graphs. We then modify this algorithm to find a small set of 5-augmenting paths, and we show that using this modification, the approximation ratio of our algorithm improves to  $3/5$ .
- The state-of-the-art algorithm for finding semi-streaming matching is randomized [15], however our algorithm is deterministic and its expected performance only depends on the random order of the edges, not the random choices of the algorithm.

## 2 Preliminaries

Let  $G = (V, E)$  be a graph. A subset of edges  $\mathcal{M} \subseteq E$  is a matching if no two different edges in  $\mathcal{M}$  share a same endpoint. For any set of edges  $E$ , we use  $\mu(E)$  to denote the size of a maximum matching in  $E$ . We may also abuse the notation through the paper and use  $\mu(E)$  to refer to the set of edges in a maximum matching. We use  $\mu(G)$  to denote the matching of  $G$ . We use  $V(G)$  and  $E(G)$  to denote the set of vertices and edges of a graph  $G$ , respectively. For a subgraph  $S \subseteq G$ , we use  $\overline{V(S)}$  to denote the vertices in  $G$  that are not in  $V(S)$ , i.e.,  $\overline{V(S)} = V(G) \setminus V(S)$ . In case that  $G = (A, B, E)$  is a bipartite graph, we use  $\overline{A(S)}$  (respectively,  $\overline{B(S)}$ ) to denote the vertices in  $A$  (resp.,  $B$ ) that are not in  $V(S)$ . When  $G = (A, B, E)$  is a bipartite graph, for any edge  $e = (a, b) \in E$ , we always assume that  $a \in A$  and  $b \in B$ .

For a graph  $G$ , let  $m = |E(G)|$  be the number of edges in  $G$ . A graph stream is a sequence of distinct edges  $S = \langle e_1, \dots, e_m \rangle$  where  $(e_1, \dots, e_m)$  is a permutation of the edges. Throughout this paper, we assume that this permutation is chosen uniformly at random from all permutations. We use  $S_{[a,b]}$  to denote the stream  $\langle e_a, \dots, e_b \rangle$ . Also, we use  $S_{(a,b]}$  to denote the stream  $\langle e_{a+1}, \dots, e_b \rangle$ . A simple greedy algorithm (Algorithm 1) returns a  $1/2$ -approximation of the maximum matching when order of edges is arbitrary. However, it is easy to see that the approximation ratio of the greedy algorithm when edges arrive in a random order could be  $1/2 + o(1)$ . For example, consider a graph  $G = (A, B, E)$  where  $|A| = |B| = n$ . Let  $A = \{a_1, \dots, a_n\}$  and  $B = \{b_1, \dots, b_n\}$ . Then connect  $a_i$  to  $b_j$  if  $i = j$  or  $i \leq n/2$  and  $j > n/2$ . Although  $G$  has a perfect matching, the subgraph between the first  $n/2$  vertices of  $A$  and the last  $n/2$  vertices of  $B$  has  $\Omega(n^2)$  edges, and the algorithm is very likely to pick these edges in the matching.

---

**Algorithm 1:** Algorithm GREEDY.

---

**Data:** An arbitrary stream  $S = \langle e_1, \dots, e_m \rangle$  of edges of a graph  $G = (V, E)$ .

- 1: Initially set  $\mathcal{M} = \emptyset$ .
  - 2: **for** each  $e \in S$  **do**
  - 3:     **if**  $\mathcal{M} \cup \{e\}$  is a matching **then**
  - 4:          $\mathcal{M} \leftarrow \mathcal{M} \cup \{e\}$ .
  - 5: **return**  $\mathcal{M}$ .
- 

A result by Chitnis et al. [3] shows that the maximum matching can be found in the streaming model with a memory of size  $\tilde{O}(k^2)$  where  $k$  is the cardinality of the maximum matching. Therefore, if  $\mu(G) \leq \sqrt{n}$ , using their algorithm we can find the exact solution using a memory of size  $\tilde{O}(n)$ . In fact, we can always run their algorithm in parallel, and set a limitation for the memory that it

uses, and if  $\mu(G)$  is small, this algorithm finds the exact solution. Therefore, throughout this paper we can assume that  $\mu(G) \geq \sqrt{n}$ .

In our analyses, we often use the Chernoff bound to prove concentration bounds (We refer the reader to [20] for the an introduction to negative association and Chernoff bound).

**Theorem 2.1** (Chernoff bound). *Given  $n$  negatively associated random variables  $X_1, \dots, X_n$  taking values in  $\{0, 1\}$ , let  $X$  denote their sum. Then, for  $0 \leq \delta \leq 1$  we have*

$$\Pr [X \leq (1 - \delta)\mathbb{E}[X]] \leq \exp\left(-\frac{\delta^2\mathbb{E}[X]}{2}\right).$$

And,

$$\Pr [X \geq (1 + \delta)\mathbb{E}[X]] \leq \exp\left(-\frac{\delta^2\mathbb{E}[X]}{3}\right).$$

### 3 Semi-Streaming Matching in Bipartite Graphs

---

**Algorithm 2:** Algorithm BIPARTITE-MATCHING for finding an approximate matching in bipartite graphs.

---

**Data:** A random order stream  $S = \langle e_1, \dots, e_m \rangle$  of edges of a bipartite graph  $G = (A, B, E)$ .

1: Run GREEDY on the edges  $S_{[1, m/\log n]}$  to find a matching  $\mathcal{M}_0$ .

2: Run the following subroutines in parallel.

3:     **Subroutine 1:**

4:         Run BIPARTITE-AUG-PATH with the input of matching  $\mathcal{M}_0$  and the stream of edges  $e \in S_{(m/\log n, m]}$  such that exactly one of the endpoints of  $e$  is in  $V(\mathcal{M}_0)$ . Let  $T$  be the set of edges returned by BIPARTITE-AUG-PATH.

5:     **Subroutine 2:**

6:         Store all edges  $e \in S_{(m/\log n, m]}$  such that  $e \cap V(\mathcal{M}_0) = \emptyset$ . Let  $R$  be the set of these edges.

7: **return**  $\mu(\mathcal{M}_0 \cup T \cup R)$ .

---

In this section, we present our algorithm for finding an approximate matching in a random stream of edges. In our algorithm (formalized as in Algorithm 2), we first pick a small fraction of edges and find an approximate matching of these edges using GREEDY algorithm. Specifically, we pick the first  $1/\log n$  fraction of the edges in the stream and we use GREEDY algorithm to find a matching  $\mathcal{M}_0$ . Let  $R$  be the set of edges in the remaining stream that do not interfere with any of edges in  $\mathcal{M}_0$ . In the following lemma which is first proved in [9], we show that w.h.p. the number of edges in  $R$  is  $\tilde{O}(n)$ . Therefore, we can store all of them in the memory.

**Lemma 3.1.** *Let  $G = (V, E)$  be a graph with  $n$  vertices and  $m$  edges,  $H$  be a subgraph in  $G$ , and  $S = \langle e_1, \dots, e_m \rangle$  be a random stream of edges of  $G$ . Let  $\mathcal{M}_0$  be the output of GREEDY on edges of  $S_{[1, \gamma m]}$  that are in  $H$ . Assume that  $\frac{1}{\log^{100} n} \leq \gamma \leq 1$ , then w.h.p.,*

$$|E(H \setminus V(\mathcal{M}_0))| = \tilde{O}\left(\frac{n}{\gamma}\right).$$

**Proof.** Let  $m' = |E(H)|$  be the number of edges in  $H$ . If  $m' \leq n$ , then the lemma clearly holds. Otherwise assume that  $m' > n$ . We know that the expected number of edges of  $E(H)$  that are in  $S_{[1, \gamma \cdot m]}$  is  $\gamma \cdot m'$ . By Chernoff bound, the probability that less than  $\gamma \cdot m'/2$  of these edges are in  $S_{[1, \gamma \cdot m]}$ , is at most

$$\exp\left(-\frac{\gamma \cdot m'}{8}\right) \leq \exp\left(-\frac{n}{100 \log^{100} n}\right)$$

Which is less than  $n^{-15}$  for a large enough  $n$ . Therefore, with a probability of at least  $1 - n^{-15}$ , the greedy algorithm sees  $\gamma \cdot m'/2$  edges of  $H$ .

For a vertex  $v$ , Let  $p_i$  be the probability of the event that  $v$  is not matched using the first  $i$  edges in  $H$  and  $v$  has at least  $30 \log n/\gamma$  neighbors in  $H$  among unmatched vertices. We show that  $p_{\gamma \cdot m'/2} \leq n^{-15}$ . Therefore, by the union bound, the probability that at least one unmatched vertex has  $30 \log n/\gamma$  neighbors in  $H \setminus V(\mathcal{M}_0)$  is  $n^{-14}$ , and we can assume w.h.p. no such vertex exists, and it proves the lemma.

Consider the event that  $v$  is not matched by the first  $i - 1$  edges and it has  $d_i \geq 30 \log n/\gamma$  unmatched neighbors. In each step of GREEDY algorithm we pick a random edge, and we add it to our matching if it do not interfere with the matching edges. The probability that  $v$  remains unmatched in the step  $i$  is at most  $1 - d_i/(m' - i + 1)$ . Since  $d_i \geq 30 \log n/\gamma$ , we have

$$p_i \leq p_{i-1} \left(1 - \frac{30 \log n}{\gamma(m' - i + 1)}\right) \leq p_{i-1} \left(1 - \frac{30 \log n}{\gamma m'}\right).$$

Therefore,

$$p_{\gamma \cdot m'/2} \leq \prod_{i=1}^{\gamma \cdot m'/2} \left(1 - \frac{30 \log n}{\gamma \cdot m'}\right) = \left(1 - \frac{30 \log n}{\gamma \cdot m'}\right)^{\gamma \cdot m'/2} \leq e^{-15 \log n} = n^{-15}.$$

□

According to the lemma above, the number of edges that do not interfere with any edge of  $\mathcal{M}_0$  is small and w.h.p. we can store all of them in a memory of  $\tilde{O}(n)$ . Let  $R$  be the set of these residual edges. Since edges in  $R$  are disjoint from edges in  $\mathcal{M}_0$ , we can compute the maximum matching in  $R$  using any offline algorithm, and report  $\mathcal{M}_0 \cup \mu(R)$  as a matching. In the following lemma we show when size of  $\mathcal{M}_0$  is small, this approach gives us a good approximation of the maximum matching.

**Lemma 3.2.** *Given a graph  $G = (V, E)$ , and a matching  $\mathcal{M}$ , let  $R = E(G \setminus V(\mathcal{M}))$  be the set of edges that do not interfere edges in  $\mathcal{M}$ . If  $|\mathcal{M}| = \alpha \mu(G)$ , then  $\mu(R) \geq (1 - 2\alpha)\mu(G)$  and  $|\mathcal{M} \cup \mu(R)| \geq (1 - \alpha)\mu(G)$ .*

**Proof.** Let  $\mathcal{M}^*$  be an arbitrary maximum matching in  $G$ . Let  $\mathcal{M}_1^*$  be the set of the edges in  $\mathcal{M}^*$  that do not interfere with the edges in  $\mathcal{M}$ , i.e.,  $\mathcal{M}_1^* = \mathcal{M}^* \cap R$ , and let  $\mathcal{M}_2^* = \mathcal{M}^* \setminus \mathcal{M}_1^*$  be the set of edges in  $\mathcal{M}^*$  that are incident to  $V(\mathcal{M})$ . Since edges in  $\mathcal{M}_2^*$  are a matching and they each are incident to a vertex in  $V(\mathcal{M})$ , we have

$$|\mathcal{M}_2^*| \leq |V(\mathcal{M})|. \tag{1}$$

Therefore,

$$|V(\mathcal{M})| \leq 2|\mathcal{M}|.$$

This along with (1) implies that

$$|\mathcal{M}_2^*| \leq 2|\mathcal{M}| = 2\alpha\mu(G).$$

Therefore,

$$|\mathcal{M}_1^*| = |\mathcal{M}^*| - |\mathcal{M}_2^*| = \mu(G) - |\mathcal{M}_2^*| \geq \mu(G)(1 - 2\alpha).$$

Note that  $\mu(R)$  is the maximum matching in  $G \setminus V(M)$ , and all of the edges in  $\mathcal{M}_1^*$  are in  $G \setminus V(M)$ . Therefore,

$$\mu(R) \geq |\mathcal{M}_1^*| \geq \mu(G)(1 - 2\alpha). \quad (2)$$

It implies,

$$\begin{aligned} |\mathcal{M} \cup \mu(R)| &= |\mathcal{M}| + |\mu(R)| \\ &\geq |\mathcal{M}| + \mu(G)(1 - 2\alpha) && \text{By (2).} \\ &= \alpha\mu(G) + \mu(G)(1 - 2\alpha) && \text{Since } |\mathcal{M}| = \alpha\mu(G). \\ &= \mu(G)(1 - \alpha), \end{aligned}$$

which proves the lemma.  $\square$

The mentioned lemma shows that if in our algorithm the size of the matching  $\mathcal{M}_0$  is very small, then we can find a good approximate of the maximum matching by finding  $\mu(R)$ . Also, if size of the matching  $\mathcal{M}_0$  is very large, we already have a large matching. Therefore, the hard case is when size of the  $\mathcal{M}_0$  is neither too small nor too large. In this case the algorithm calls BIPARTITE-AUG-PATH. Suppose that  $E'$  is the set of edges in the remaining stream. The algorithm finds a set 3-augmenting paths and returns a set of edges  $T$  that are candidate edges for 3-augmenting paths. The algorithm BIPARTITE-AUG-PATH guarantees to return a set of edges  $T$  such that  $\mu(T \cup R) \geq (\frac{4}{7} - o(1))\mu(E')$ . We claim that w.h.p.  $\mu(E') \geq (1 - o(1))\mu(G)$ , and this immediately implies that  $\mu(T \cup R) \geq (\frac{4}{7} - o(1))\mu(G)$ . We later improve the bound to  $(\frac{3}{5} - o(1))\mu(G)$  in Subsection 3.2.

To show that  $\mu(E') \geq (1 - o(1))\mu(G)$ , fix an arbitrary maximum matching  $\mathcal{M}^*$  in  $G$ . Each edge  $e \in \mathcal{M}^*$  is in the last  $(1 - 1/\log n)$  fraction of the edges with the probability of  $(1 - 1/\log n)$ . Thus,  $e$  is in  $E'$  with the probability of  $(1 - 1/\log n)$ . Therefore, the expected number of edges from  $\mathcal{M}^*$  that are in  $E'$  is  $\mu(G)(1 - 1/\log n)$ . By Chernoff bound the probability that more than  $2/\log n$  fraction of these edges are in the first  $1/\log n$  fraction of the stream is bounded by

$$\exp\left(-\frac{|\mathcal{M}^*|}{3 \cdot \log n}\right) \leq \exp\left(-\frac{\sqrt{n}}{3 \log n}\right). \quad \text{Since } \mu(G) \geq \sqrt{n}.$$

For a large enough  $n$ , we have  $\sqrt{n}/(3 \log n) \geq 10 \log n$ . Therefore, the probability above is bounded by  $\exp(-10 \log n) = n^{-10}$ . Therefore, w.h.p. the size of the maximum matching in the remaining stream is at least  $\mu(G)(1 - 2/\log n) = \mu(G)(1 - o(1))$ .

### 3.1 Finding 3-augmenting paths

In this section we explain our algorithm for finding 3-augmenting paths. Suppose that we are given a graph  $G = (A, B, E)$  and a matching  $\mathcal{M}_0$  in  $G$ . Let  $R$  be the set of edges that do not interfere with any of edges in  $\mathcal{M}_0$ . Algorithm BIPARTITE-AUG-PATH finds a set of 3-augmenting paths and returns a set  $T$  of the candidate edges for augmenting  $\mathcal{M}_0$  such that w.h.p.

---

**Algorithm 3:** BIPARTITE-AUG-PATH Algorithm for augmenting a matching using 3-augmenting paths.

---

**Data:** A random order stream  $S = \langle e_1, \dots, e_m \rangle$  of edges of a bipartite graph  $G = (A, B, E)$ , and a maximal matching  $\mathcal{M}_0$ .

- 1: Set  $\tau = \frac{1}{100 \log^3 n}$ .
- 2: Run the following subroutines in parallel.
- 3:     **Subroutine 1:**
- 4:         Run GREEDY to find a matching  $P_1$  on the set of edges  $(a, b) \in S_{[1, m \cdot \tau]}$  such that  $a \in A(\mathcal{M}_0)$  and  $b \notin B(\mathcal{M}_0)$ .
- 5:         Let  $\mathcal{M}_P = \{e \in \mathcal{M}_0 \mid e \cap V(P_1) \neq \emptyset\}$  be the set of edges in  $\mathcal{M}_0$  that are incident to the edges in  $P_1$ .
- 6:         Run GREEDY to find a matching  $P_2$  on the set of edges  $(a, b) \in S_{(m \cdot \tau, 2m \cdot \tau]}$  such that  $a \notin A(\mathcal{M}_0)$  and  $b \in B(\mathcal{M}_P)$ .
- 7:     **Subroutine 2:**
- 8:         Run GREEDY to find a matching  $Q_1$  on the set of edges  $(a, b) \in S_{[1, m \cdot \tau]}$  such that  $a \notin A(\mathcal{M}_0)$  and  $b \in B(\mathcal{M}_0)$ .
- 9:         Let  $\mathcal{M}_Q = \{e \in \mathcal{M}_0 \mid e \cap V(Q_1) \neq \emptyset\}$  be the set of edges in  $\mathcal{M}_0$  that are incident to the edges in  $Q_1$ .
- 10:         Run GREEDY to find a matching  $Q_2$  on the set of edges  $(a, b) \in S_{(m \cdot \tau, 2m \cdot \tau]}$  such that  $a \in A(\mathcal{M}_Q)$  and  $b \notin B(\mathcal{M}_0)$ .
- 11:
- 12: **if**  $|P_2| \geq |\mathcal{M}_0| / \log n$  or  $|Q_2| \geq |\mathcal{M}_0| / \log n$  ;                     // W.l.o.g., assume  $|P_2| \geq |\mathcal{M}_0| / \log n$ .
- 13:     **1 then**
- 14:         Let  $U$  be maximum 3-augmenting paths for  $\mathcal{M}_0$  using the edges  $P_1$  and  $P_2$ .
- 15:         Let  $T$  be the output of BIPARTITE-AUG-PATH with the input of matching  $\mathcal{M}_0 \Delta U$ , and the stream of edges  $S_{(2m \cdot \tau, m]}$ .
- 16:         **return**  $T \cup \mathcal{M}_0$ .
- 17:     **else**
- 18:         Run the following subroutines in parallel.
- 19:         **Subroutine 3:**
- 20:             Store all edges  $e = (a, b) \in S_{(2m \cdot \tau, m]}$  such that  $a \notin A(\mathcal{M}_0)$  and  $b \in B(\mathcal{M}_P)$  and  $e \cap V(P_2) = \emptyset$ . Let  $R_1$  be the set of these edges.
- 21:         **Subroutine 4:**
- 22:             Store all edges  $e = (a, b) \in S_{(2m \cdot \tau, m]}$  such that  $a \in A(\mathcal{M}_Q)$  and  $b \notin B(\mathcal{M}_0)$  and  $e \cap V(Q_2) = \emptyset$ . Let  $R_2$  be the set of these edges.
- 23:         **Subroutine 5:**
- 24:             Store all edges  $e = (a, b) \in S_{(2m \cdot \tau, m]}$  such that  $a \in A(\mathcal{M}_0)$  and  $b \notin B(\mathcal{M}_0)$  and  $e \cap V(P_1) = \emptyset$ . Let  $R_3$  be the set of these edges.
- 25:         **Subroutine 6:**
- 26:             Store all edges  $e = (a, b) \in S_{(2m \cdot \tau, m]}$  such that  $a \notin A(\mathcal{M}_0)$  and  $b \in B(\mathcal{M}_0)$  and  $e \cap V(Q_1) = \emptyset$ . Let  $R_4$  be the set of these edges.
- 27:         Let  $T$  be  $\mathcal{M}_0 \cup P_1 \cup P_2 \cup Q_1 \cup Q_2 \cup R_1 \cup \dots \cup R_4$ .
- 28:         **return**  $T$ .

---

$\mu(T \cup R) \geq (\frac{4}{7} - o(1))\mu(G)$ . From the algorithm BIPARTITE-MATCHING we know that w.h.p. the number of edges in  $R$  is  $\tilde{O}(n)$  and we can store all of them in the memory. Therefore, we can find the matching  $\mu(T \cup R)$ , and in fact, get a  $4/7 - o(1)$  approximation of the maximum matching.

In order to demonstrate the intuition behind our algorithm, we first describe the algorithm when  $\mathcal{M}_0$  is a maximal matching. Since  $\mathcal{M}_0$  is maximal, we have  $|\mathcal{M}_0| \geq \mu(G)/2$ . For the sake of simplicity, we only consider the case that seems to be the worst case and we assume that the size of  $|\mathcal{M}_0|$  is exactly  $\mu(G)/2$ , i.e.,  $|\mathcal{M}_0| = \mu(G)/2$ . We show that Algorithm BIPARTITE-AUG-PATH finds a set of 3-augmenting paths such that by swapping the edges along these augmenting paths, we can get w.h.p. a matching  $\mathcal{M}_1$  with the size of at least  $(\frac{2}{3} - o(1))\mu(G)$ . Since  $\mathcal{M}_0$  is maximal and its size is  $\mu(G)/2$ , each edge in  $\mathcal{M}_0$  is part of a disjoint 3-augmenting path. Therefore there exist a perfect matching between the matched vertices of part  $A$  (resp.,  $B$ ) and unmatched vertices in  $B$  (resp.,  $A$ ). Let use  $G_U$  to denote the induced subgraph between the matched vertices in  $A$  and unmatched vertices in  $B$ . Then, we know that the size of the maximum matching in  $G_U$  is  $|\mathcal{M}_0|$ , i.e.,  $\mu(G_U) = |\mathcal{M}_0|$ . Similarly, we define  $G_L$  to be the induced subgraph graph between matched vertices in  $B$  and unmatched vertices in  $A$ . Using a similar argument we have  $\mu(G_L) = |\mathcal{M}_0|$ .

Simply finding a matching in  $G_U$  and  $G_L$  is not enough for finding 3-augmenting paths. For example, suppose we use GREEDY algorithm to find a matching in  $G_U$  and  $G_L$ . Since the approximation ratio of the greedy algorithm is  $1/2$ , it is possible that the greedy algorithm returns a matching of size  $|\mathcal{M}_0|/2$  in both  $G_U$  and  $G_L$ . The worst case scenario is when for each edge  $e \in \mathcal{M}_0$ , at most one of its endpoints is covered by these two matchings. In this case, we cannot find any 3-augmenting path for the edges in  $\mathcal{M}_0$ . This example shows that finding 3-augmenting paths is by nature harder than finding a matching between matched and unmatched vertices.

Suppose that we want to find a 3-augmenting path for some edge  $e \in \mathcal{M}_0$ . Then, we should find a path  $(e_U, e, e_L)$  in the graph where  $e_U \in E(G_U)$  and  $e_L \in E(G_L)$ . We refer to  $e_U$  as a *upper wing* and we refer to  $e_L$  as a *lower wing*. Let  $\tau = \frac{1}{100 \log^3 n}$  be a very small number. Algorithm 3 first finds a matching  $P_1$  (respectively,  $Q_1$ ) in  $G_U$  (resp.,  $G_L$ ) using the  $\tau$  fraction of the edges which is only a small portion of all edges. Consider an edge  $e = (a, b) \in \mathcal{M}_0$  such that  $a$  is matched by  $P_1$ . Then, the matching edge in  $P_1$  is an upper wing for  $e$ . If we match the other end of  $e$ , i.e., vertex  $b$ , to one of the vertices in  $A(G_L)$ , then we have a lower wing and we can find a 3-augmenting path for  $e$  and increase the size of the matching. Let  $\mathcal{M}_P = \{e \in \mathcal{M}_0 | e \cap P_1 \neq \emptyset\}$  be the matching edges in  $\mathcal{M}_0$  that are incident to the edges in  $P_1$ . In order to find 3-augmenting paths, we use another  $\tau$  fraction of the edges to match the vertices in  $B(\mathcal{M}_P)$  to the vertices in  $A(G_L)$ . Let  $P_2$  be the result of GREEDY algorithm to find this matching. We can use edges in  $P_1$  and  $P_2$  to find  $\min\{|P_1|, |P_2|\} = |P_2|$ , 3-augmenting paths for  $\mathcal{M}_0$ .

Consider the case that  $|P_2| \geq |\mathcal{M}_0|/\log n$ . In this case we can find  $|\mathcal{M}_0|/\log n$  disjoint 3-augmenting paths. Therefore, we can increase the size of  $\mathcal{M}_0$  by a multiplicative factor of  $(1 + 1/\log n)$  by swapping the edges along these augmenting paths. Therefore, our algorithm can increase the size of  $\mathcal{M}_0$  by a factor of  $(1 + 1/\log n)$  using only  $2\tau \leq \frac{1}{50 \log^3 n}$  fraction of the edges which is a very small portion of all edges. In this case we increase the size of  $\mathcal{M}_0$  and recursively call the algorithm to increase the size of the new matching. Since the size of  $\mathcal{M}_0$  will increase by a factor of  $(1 + 1/\log n)$  by each recursive call and initially the size of our matching is at least  $\mu(G)/2$ , the maximum number of recursive calls is bounded by  $\log_{1+1/\log n} 2 \leq \log n$ . Eventually, our algorithm reaches a state such that  $|P_2| < |\mathcal{M}_0|/\log n$  and at this point the algorithm has seen at the at most  $2\tau \cdot \log n = \frac{1}{50 \log^2 n}$  fraction of the edges which is still very small. Therefore, we can assume that the size of  $P_2$  (similarly,  $Q_2$ ) is at most  $|\mathcal{M}_0|/\log n = o(|\mathcal{M}_0|)$ .

Consider the case that the size of  $P_1$  is large. In this case, the greedy algorithm had returned a matching of size  $|P_2| = o(|\mathcal{M}_0|) = o(|P_1|)$  between the vertices  $B(\mathcal{M}_P)$  and  $A(G_L)$  which is very small. Roughly speaking, almost all of the vertices in  $B(\mathcal{M}_P)$  and  $A(G_L)$  are remained unmatched

using the greedy algorithm. Lemma 3.1 implies that w.h.p. the number of edges between the vertices  $B(\mathcal{M}_P)$  and  $A(G_L)$  that do not interfere with the edges in  $P_2$  is bounded by  $\tilde{O}(n)$  and we can store all of them in the memory. Recall that there exists a perfect matching in  $G_L$ . Therefore, the size of the maximum matching between  $B(\mathcal{M}_P)$  and  $A(G_L)$  is  $|P_1|$  and at most  $2|P_2|$  edges of them interfere with the vertices in  $V(P_2)$ . Since we can store all edges between  $B(\mathcal{M}_P)$  and  $A(G_L)$  that are not incident to the vertices in  $V(P_2)$ , we can find a matching of size

$$|P_1| - 2|P_2| = |P_1|(1 - o(1)),$$

between  $B(\mathcal{M}_P)$  and  $A(G_L)$ . Thus, we can find  $|P_1|(1 - o(1))$  augmenting paths for  $|\mathcal{M}_0|$  and increase the size of the matching. If  $|P_1| \geq |\mathcal{M}_0|/3$ , then the size of the matching returned by our algorithm is at least

$$\begin{aligned} |\mathcal{M}_0| + |P_1|(1 - o(1)) &\geq |\mathcal{M}_0|(1 + 1/3 - o(1)) \\ &\geq \frac{\mu(G)}{2}(1 + 1/3 - o(1)) && \text{Since } M_0 \text{ is a maximal matching.} \\ &= (2/3 - o(1))\mu(G), \end{aligned}$$

which is the desired approximation ratio.

The remaining case is when  $|P_1| < |\mathcal{M}_0|/3$ . Recall that  $P_1$  is the result of the running the greedy algorithm in order to find a matching in  $G_U$ . Lemma 3.1 implies that w.h.p. the number of edges in  $G_U \setminus V(P_1)$  is  $\tilde{O}(n)$  and we can store all of them in the memory. Let  $R'$  be the set of these edges. By Lemma 3.2 we know that

$$\begin{aligned} |P_1 \cup \mu(R')| &\geq \mu(G_U) - |P_1| \\ &= |\mathcal{M}_0| - |P_1| \\ &\geq \frac{2|\mathcal{M}_0|}{3}. && \text{Since } |P_1| < |\mathcal{M}_0|/3. \end{aligned}$$

Therefore, the algorithm can find a matching of size  $2|\mathcal{M}_0|/3$  in  $G_U$ . Using a similar argument about  $Q_1$ , we can say that the algorithm finds a matching of size at least  $2|\mathcal{M}_0|/3$  in  $G_L$ . Therefore, the number of edges in  $\mathcal{M}_0$  that are covered by both of these matchings is at least

$$\frac{2|\mathcal{M}_0|}{3} + \frac{2|\mathcal{M}_0|}{3} - |\mathcal{M}_0| = \frac{|\mathcal{M}_0|}{3}.$$

As the result, the algorithm can find  $|\mathcal{M}_0|/3$  augmenting paths. By applying these augmenting paths, we get a matching of size at least

$$\begin{aligned} |\mathcal{M}_0|(1 + 1/3) &\geq \frac{\mu(G)}{2}(1 + 1/3) && \text{Since } M_0 \text{ is a maximal matching.} \\ &= (2/3)\mu(G), \end{aligned}$$

which is the desired approximation ratio. Roughly speaking, this shows that our algorithm has an acceptable approximation ratio when  $\mathcal{M}_0$  is maximal. In the following theorem, we demonstrate the performance of our algorithm giving any matching  $\mathcal{M}_0$ .

**Theorem 3.3.** *Given a bipartite graph  $G = (A, B, E)$  with  $n$  vertices and a matching  $\mathcal{M}_0$ . Algorithm BIPARTITE-AUG-PATH returns a set of edges  $T$  such that  $\mathcal{M}_0 \subseteq T$  and w.h.p. in  $n$ , we have  $|T| = \tilde{O}(n)$  and  $\mu(T \cup R) \geq (\frac{4}{7} - o(1))\mu(G)$  where  $R = E(G \setminus V(\mathcal{M}_0))$  is the set of edges in  $G$  that do not interfere with the edges in  $\mathcal{M}_0$ .*

**Proof.** Consider an arbitrary maximum matching  $\mathcal{M}^*$  in  $G$ . Let  $\mathcal{M}_1^*$  be the set of edges in  $\mathcal{M}^*$  that are incident to the edges in  $\mathcal{M}_0$ , and  $\mathcal{M}_2^* = \mathcal{M}^* \setminus \mathcal{M}_1^*$  be the set of other edges.  $\mathcal{M}_1^*$  and  $\mathcal{M}_2^*$  partition the edges in  $\mathcal{M}^*$ . Therefore, we have

$$|\mathcal{M}_1^*| + |\mathcal{M}_2^*| = |\mathcal{M}^*| = \mu(G). \quad (3)$$

For every edge in  $\mathcal{M}_1^*$ , at least one of its endpoints is in  $V(\mathcal{M}_0)$ . Therefore,

$$|\mathcal{M}_1^*| \leq 2|\mathcal{M}_0|. \quad (4)$$

We claim that  $|\mathcal{M}_0| \leq |\mathcal{M}_1^*|$ . Since edges in  $\mathcal{M}_2^*$  do not interfere with the edges in  $\mathcal{M}_0$ , edges in  $\mathcal{M}_0 \cup \mathcal{M}_2^*$  are a matching in the graph. Recall that  $\mathcal{M}^*$  is a maximum matching in  $G$ . Since  $\mathcal{M}_0 \cup \mathcal{M}_2^*$  is a matching we have

$$\begin{aligned} |\mathcal{M}_0| + |\mathcal{M}_2^*| &\leq |\mathcal{M}^*| \\ \Rightarrow |\mathcal{M}_0| + |\mathcal{M}_2^*| &\leq |\mathcal{M}_1^*| + |\mathcal{M}_2^*| && \text{By (3).} \\ \Rightarrow |\mathcal{M}_0| &\leq |\mathcal{M}_1^*|. \end{aligned}$$

This along with (4) implies that  $|\mathcal{M}_0| = (1/2 + \delta)|\mathcal{M}_1^*|$  where  $0 \leq \delta \leq 1/2$ . Suppose that  $|\mathcal{M}_1^*| = \alpha\mu(G)$ . Considering the matching  $\mathcal{M}_0 \cup \mu(R)$ , we have

$$\begin{aligned} |\mathcal{M}_0 \cup \mu(R)| &\geq |\mathcal{M}_0| + \mu(R) \\ &\geq |\mathcal{M}_0| + |\mathcal{M}_2^*| && \text{Since every edge in } \mathcal{M}_2^* \text{ is in } R. \\ &= |\mathcal{M}_0| + (1 - \alpha)\mu(G) && \text{Since } |\mathcal{M}_1^*| + |\mathcal{M}_2^*| = \mu(G). \\ &= (1/2 + \delta)|\mathcal{M}_1^*| + (1 - \alpha)\mu(G) \\ &= (1/2 + \delta)\alpha\mu(G) + (1 - \alpha)\mu(G) && \text{Since } |\mathcal{M}_1^*| = \alpha\mu(G). \\ &= \mu(G)(1 - \alpha/2 + \alpha\delta). \end{aligned}$$

Therefore, if  $\alpha/2 - \alpha\delta \leq \frac{3}{7}$ , we have

$$|\mathcal{M}_0 \cup \mu(R)| \geq \mu(G)(1 - \alpha/2 + \alpha\delta) \geq \frac{4}{7}\mu(G).$$

Since  $\mathcal{M}_0 \subseteq T$ , in the set  $T$  returned by the algorithm we have  $\mu(T \cup R) \geq \frac{4}{7}\mu(G)$  which is the desired bound. Therefore, for the rest of the proof we can assume that

$$\alpha/2 - \alpha\delta \geq \frac{3}{7}. \quad (5)$$

This also implies that  $\alpha/2 + \alpha\delta \leq \frac{4}{7}$ , since

$$\begin{aligned} \alpha/2 + \alpha\delta &= \alpha - (\alpha/2 - \alpha\delta) \\ &\leq 1 - (\alpha/2 - \alpha\delta) && \text{Since } \alpha \leq 1. \\ &\leq 4/7. && \text{By (5).} \end{aligned} \quad (6)$$

We use  $G_U$  to denote the induced subgraph between the matched vertices in  $A$  and unmatched vertices in  $B$ , i.e.,  $G_U$  is the induced subgraph between the vertices  $A(\mathcal{M}_0)$  and  $\overline{B(\mathcal{M}_0)}$ . We also use  $G_L$  to denote the induced subgraph between the matched vertices in  $B$  and unmatched vertices in  $A$ . In order to find 3-augmenting paths for the edges in  $\mathcal{M}_0$ , the algorithm should find upper wings from  $G_U$  and lower wings from  $G_L$ . Let  $\tau = \frac{1}{100 \log^3 n}$ , the algorithm uses  $\tau$  fraction of the edges and runs GREEDY algorithm to find a matching  $P_1$  (respectively,  $Q_1$ ) in  $G_U$  (resp.,  $G_L$ ). Let  $E' \subseteq E$  be the set of edges in  $E$  that are incident to the vertices in  $V(\mathcal{M}_0)$ .  $\mathcal{M}_1^*$  is a matching using the in  $E'$ . Therefore,  $\mu(E') \geq |\mathcal{M}_1^*|$ . The following claim shows that there exists a matching of size at least  $(1/2 - \delta)|\mathcal{M}_1^*|$  in both  $G_U$  and  $G_L$ . The proof is deferred to Appendix A.

**Claim 3.4.** *Given a bipartite graph  $G = (A, B, E)$  and a maximal matching  $\mathcal{M}$ , suppose that  $|\mathcal{M}| = (1/2 + \delta)\mu(G)$  where  $0 \leq \delta \leq 1/2$ , then there exists a matching of size at least  $(1/2 - \delta)\mu(G)$  between vertices  $A(\mathcal{M})$  and  $\overline{B(\mathcal{M})}$  and also between vertices  $B(\mathcal{M})$  and  $\overline{A(\mathcal{M})}$ .*

The algorithm uses a small fraction of edges to find matchings  $P_1$  and  $Q_1$ . Let  $M_P$  be the edges in  $M_0$  that are incident to the edges in  $P_1$ . Specifically,  $M_P = \{e \in \mathcal{M}_0 | e \cap V(P_1) \neq \emptyset\}$ . Each edge in  $M_P$  is matched by the matching  $P_1$ . Therefore, we have an upper wing for every edge in  $M_P$ . If we can find lower wings for these edges, then we have found 3-augmenting paths for these edges, and we can increase the size of the matching. In order to find lower wings, the algorithm in (Subroutine 1) uses another  $\tau$  fraction of the edges and calls the greedy algorithm to match vertices in  $B(M_P)$  to the vertices in  $A(G_L)$ . Let  $P_2$  be the result of this matching. We can use edges in  $P_1$  and  $P_2$  to find  $\min\{|P_1|, |P_2|\} = |P_2|$ , 3-augmenting paths for  $\mathcal{M}_0$ . If  $|P_2| \geq (1/\log n)|\mathcal{M}_0|$ , then by switching the edges along the augmenting paths, we can increase the size of  $\mathcal{M}_0$  by a factor of  $(1 + 1/\log n)$ . In this case, the algorithm increases the size of  $\mathcal{M}_0$  and recursively calls itself with the new matching. Let  $t$  be the number of recursive calls. Since the size of matching  $\mathcal{M}_0$  is always bounded by  $\mu(G)$ , we have

$$\left(1 + \frac{1}{\log n}\right)^t \leq \mu(G) \leq n.$$

Thus,  $t$  is at most  $\log_{1+1/\log n} n$  which is at most  $2\log^2 n$  for a large enough  $n$ . By the end of all recursive calls the algorithm has only used  $t \cdot 2\tau \cdot m \leq m/(25 \log n)$  edges which is only a small fraction of all edges. After at most  $t$  recursive calls, the algorithm reaches a state where  $|P_2| < |\mathcal{M}_0|/\log n$ . At this point the algorithm has used at most  $1/(25 \log n)$  fraction of the edges. We argue that ignoring this small fraction of the edges does not affect the size of the maximum matching. Since edges arrive in a uniformly random order, for each edge in  $\mathcal{M}^*$  the probability that this edge is within the first  $1/(25 \log n)$  fraction of the edges is  $1/(25 \log n)$ . Hence, the expected number of edges in  $\mathcal{M}^*$  that are in the first  $1/(25 \log n)$  fraction of the edges is  $|\mathcal{M}^*|/(25 \log n)$ . By Chernoff bound the probability that more than  $2|\mathcal{M}^*|/(25 \log n)$  of these edges are in the first  $1/(25 \log n)$  fraction of the edges is bounded by

$$\begin{aligned} & \exp\left(-\frac{|\mathcal{M}^*|}{3 \cdot 25 \log n}\right) \\ & \leq \exp\left(-\frac{\sqrt{n}}{75 \log n}\right). \end{aligned} \quad \text{Since } \mu(G) \geq \sqrt{n}.$$

For a large enough  $n$ , we have  $\sqrt{n}/(75 \log n) \geq 10 \log n$ . Therefore, the probability above is bounded by  $\exp(-10 \log n) = n^{-10}$  for a large enough  $n$ . Therefore, w.h.p. the size of the maximum matching in the remaining stream is at least

$$|\mathcal{M}^*| \left(1 - \frac{2}{25 \log n}\right) = \mu(G) \left(1 - \frac{2}{25 \log n}\right) = \mu(G)(1 - o(1)).$$

Thus, the recursive calls of the algorithm does not change the size of the matching, and we can assume  $|P_2| < |\mathcal{M}_0|/\log n$ . Also, using a similar argument we can assume  $|Q_2| < |\mathcal{M}_0|/\log n$ .

By a slight abuse of notion, we suppose that  $\mathcal{M}^*$  is a maximum matching in the remaining stream. We also suppose that  $\mathcal{M}_1^*$  is the set of edges in  $\mathcal{M}^*$  that are incident to the edges in  $\mathcal{M}_0$ , and  $\mathcal{M}_2^* = \mathcal{M}^* \setminus \mathcal{M}_1^*$  is the set of other edges in  $\mathcal{M}^*$ . Therefore,

$$|\mathcal{M}^*| = |\mathcal{M}_1^*| + |\mathcal{M}_2^*| = \mu(G)(1 - o(1)).$$

Let  $R_P$  be the set of edges in  $\mathcal{M}_2^*$  that are incident to the edges in  $P_1$ . In other words,

$$R_P = \{e \in \mathcal{M}_2^* \mid e \cap V(P_1) \neq \emptyset\}.$$

Similarly, we define  $R_Q$  to be

$$R_Q = \{e \in \mathcal{M}_2^* \mid e \cap V(Q_1) \neq \emptyset\}.$$

In order to prove the theorem, we consider two different cases on sizes of  $P_1, Q_1, R_P$  and  $R_Q$ . The cases are as follows.

- **Case 1:**  $|P_1| - |R_P| \geq \frac{1}{7}\mu(G)$ , or  $|Q_1| - |R_Q| \geq \frac{1}{7}\mu(G)$ .

Without loss of generality, we assume that  $|P_1| - |R_P| \geq \mu(G)/7$ . Recall that  $M_P$  is the set of edges in  $\mathcal{M}_0$  that are incident to the edges in  $P_1$ . Since  $P_1$  is a matching in  $G_U$ , each edge in  $M_P$  has an upper wing, and if we find lower wings for them, we have a set of 3-augmenting path. The Algorithm in (Subroutine 1) picks a small fraction of edges and finds a greedy matching  $P_2$  between the vertices in  $B(\mathcal{M}_P)$  and  $A(G_L)$ . As we discussed before, we can assume that  $|P_2| \leq |\mathcal{M}_0|/\log n$ . Therefore, we have

$$|P_2| \leq \frac{|\mathcal{M}_0|}{\log n} \leq \frac{\mu(G)}{\log n} = o(\mu(G)).$$

Thus, the greedy algorithm has matched only a small fraction of the vertices in the induced graph between  $B(\mathcal{M}_P)$  and  $A(G_L)$ . Let  $G'$  be the induced subgraph between vertices  $B(\mathcal{M}_P)$  and  $A(G_L)$ . By Lemma 3.1 we have that w.h.p. the number of edges in the graph  $G' \setminus V(P_2)$  is bounded by  $\tilde{O}(\frac{n}{\tau}) = \tilde{O}(n)$  and we can store all of them in the memory bound of our semi-streaming algorithm. The algorithm saves all of these edges in (Subroutine 3).

In the following claim we show that there exists a matching in  $G'$  with the size of at least  $(|P_1| - \mu(G) \cdot 2\delta\alpha)$  such that its edges do not interfere with the edges in  $\mathcal{M}_2^*$ . The proof can be found in Appendix A.

**Claim 3.5.** *Let  $\mathcal{M}'$  be the set of edges in  $\mathcal{M}_1^*$  that are in  $G'$ . Then,  $|\mathcal{M}'|$  is at least  $|P_1| - \mu(G) \cdot 2\delta\alpha$ .*

We show that how this claim implies that the approximation ratio of our algorithm is at least  $(4/7 - o(1))$ . Suppose that  $\mu(G')$  is at least  $(|P_1| - \mu(G) \cdot 2\delta\alpha)$ . Let  $G''$  be the graph resulting from removing the vertices of  $P_2$  from  $G'$ , i.e.,  $G'' = G' \setminus V(P_2)$ . Since size of  $P_2$  is very small and it is at most  $o(\mu(G))$ , we have that

$$\mu(G'') \geq \mu(G') - 2|P_2| = |P_1| - \mu(G) \cdot 2\delta\alpha - o(\mu(G)) = |P_1| - \mu(G)(2\delta\alpha + o(1)).$$

Since we store all of the edges in  $G''$  in our memory, we can find a matching  $\mathcal{M}'$  of size at least  $|P_1| - \mu(G)(2\delta\alpha + o(1))$  in this subgraph such that the edges of this matching do not interfere with the edges in  $\mathcal{M}_2^*$ . Each matching edge in this subgraph will be a lower wing for an edge in  $M_P$ . Therefore, the algorithm will find  $|P_1| - \mu(G)(2\delta\alpha + o(1))$  augmenting paths. These augmenting paths interfere with at most  $|R_P|$  edges of  $\mathcal{M}_2^*$ , since none of the edges in  $\mathcal{M}'$  interfere with the edges in  $\mathcal{M}_2^*$ . Let  $T$  the union of the edges in  $\mathcal{M}_0$  and the augmenting

paths that we have found. According to what we discussed, we have

$$\begin{aligned}
& \mu(T \cup R) \\
& \geq |M_0| + (|P_1| - \mu(G)(2\delta\alpha + o(1))) + |\mathcal{M}_2^*| - |R_P| \\
& = (1/2 + \delta)|\mathcal{M}_1^*| + |P_1| - \mu(G)(2\delta\alpha + o(1)) + |\mathcal{M}_2^*| - |R_P| \\
& = (1/2 + \delta)\alpha\mu(G) + |P_1| - \mu(G)(2\delta\alpha + o(1)) + |\mathcal{M}_2^*| - |R_P| && \text{Since } |\mathcal{M}_1^*| = \alpha|\mathcal{M}^*|. \\
& = (1/2 - \delta)\alpha\mu(G) + |P_1| - o(\mu(G)) + |\mathcal{M}_2^*| - |R_P| \\
& = (1/2 - \delta)\alpha\mu(G) + |P_1| + (1 - \alpha)\mu(G) - |R_P| - o(\mu(G)) && \text{Since } |\mathcal{M}_2^*| = |\mathcal{M}^*| - |\mathcal{M}_1^*|. \\
& = (1 - \alpha/2 - \delta\alpha - o(1))\mu(G) + |P_1| - |R_P| \\
& \geq (1 - \alpha/2 - \delta\alpha - o(1))\mu(G) + \mu(G)/7 && \text{Since } |P_1| - |R_P| \geq \mu(G)/7. \\
& = (8/7 - \alpha/2 - \delta\alpha - o(1))\mu(G).
\end{aligned}$$

By inequality above and (6), we have,

$$\begin{aligned}
\mu(T \cup R) & \geq (8/7 - \alpha/2 - \delta\alpha - o(1))\mu(G) \\
& \geq (4/7 - o(1))\mu(G), && \text{By (6).}
\end{aligned}$$

which is our desired bound. We can use a similar argument to show that when  $|Q_1| - |R_Q| \geq \frac{1}{7}\mu(G)$ , the approximation ratio of algorithm is at least  $(\frac{4}{7} - o(1))$ .

- **Case 2:**  $|P_1| - |R_P| < \frac{1}{7}\mu(G)$ , and  $|Q_1| - |R_Q| < \frac{1}{7}\mu(G)$ .

The remaining case is when  $|P_1| - |R_P| < \frac{1}{7}\mu(G)$  and  $|Q_1| - |R_Q| < \frac{1}{7}\mu(G)$ . Recall that  $P_1$  and  $Q_1$  are the result of the greedy matching algorithm using  $\tau$  fraction of the edges in  $G_U$  and  $G_L$ . Therefore, by Lemma 3.1, we know that with w.h.p. we can store all of the edges in  $G_U \setminus V(P_1)$  and also all the edges in  $G_L \setminus V(Q_1)$  within a memory of size  $\tilde{O}(n/\tau) = \tilde{O}(n)$  which is within the memory bound of the semi-streaming algorithms. The algorithm stores all the edges in  $G_U \setminus V(P_1)$  and also all the edges in  $G_L \setminus V(Q_1)$  in (Subroutine 5) and (Subroutine 6) respectively. Let  $R_3$  and  $R_4$  be the set of these edges respectively, and  $T$  be  $P_1 \cup Q_1 \cup R_3 \cup R_4$ . We show that  $\mu(T) \geq (\frac{4}{7} - o(1))\mu(G)$ .

We claim that using the edges in  $R_3$  and  $P_1$  we can find a matching of size of  $(\frac{2}{7} - o(1))\mu(G)$  in  $G_U$ . Consider the matching  $\mathcal{M}_1^*$ . By Claim 3.4 we know there exists a matching of size at least  $(1/2 - \delta)|\mathcal{M}_1^*|$  in  $G_U$ . Let  $\mathcal{M}$  be this matching in  $G_U$ , and  $\mathcal{M}'$  be the set of the edges in  $\mathcal{M}$  that interfere with the edges in  $P_1$ , i.e.,  $\mathcal{M}' = \{e \in \mathcal{M} \mid e \cap V(P_1) \neq \emptyset\}$ . We show that the size of  $\mathcal{M}'$  is at most  $2|P_1| - |R_P|$ . The total number of vertices covered by  $P_1$  is exactly  $2|P_1|$ . However, we know that  $|R_P|$  edges from  $P_1$  cover at least one vertex from  $\mathcal{M}_2^*$ . Since vertices of  $\mathcal{M}_1^*$  and  $\mathcal{M}_2^*$  are disjoint, it implies that the number of edges in  $\mathcal{M}_1^*$  that are covered by  $P_1$  is at most  $2|P_1| - |R_P|$ . Thus,  $|\mathcal{M}'| \leq 2|P_1| - |R_P|$ . Note that edges in  $\mathcal{M} \setminus \mathcal{M}'$  do not share a vertex with the edges in  $P_1$ . Therefore, the size of maximum matching in  $R_3 \cup P_1$  is at least,

$$\begin{aligned}
|P_1| + |\mathcal{M}| - |\mathcal{M}'| & \geq |P_1| + |\mathcal{M}| - (2|P_1| - |R_P|) && \text{Since } |\mathcal{M}'| \leq 2|P_1| - |R_P|. \\
& = |R_P| - |P_1| + |\mathcal{M}| \\
& > |\mathcal{M}| - \mu(G)/7. && \text{By the assumption of Case 2.}
\end{aligned}$$

Since the size  $\mathcal{M}$  is at least  $(1/2 - \delta)|\mathcal{M}_1^*|$ , we have

$$\begin{aligned}
\mu(R_3 \cup P_1) &\geq |\mathcal{M}| - \mu(G)/7 \\
&\geq (1/2 - \delta)|\mathcal{M}_1^*| - \mu(G)/7 \\
&= (1/2 - \delta)\alpha|\mathcal{M}^*| - \mu(G)/7 && \text{Since } |\mathcal{M}_1^*| = \alpha|\mathcal{M}^*|. \\
&= (1/2 - \delta - o(1))\alpha\mu(G) - \mu(G)/7 \\
&= \mu(G)(\alpha/2 - \alpha\delta - 1/7 - o(1)) \\
&\geq \mu(G)(3/7 - 1/7 - o(1)) && \text{By (5).} \\
&= \mu(G)(2/7 - o(1)).
\end{aligned}$$

Using a similar argument for  $R_4 \cup Q_1$ , we can say that  $\mu(R_4 \cup Q_1) \geq (\frac{2}{7} - o(1))\mu(G)$ . Therefore,

$$\mu(T) = \mu(P_1 \cup Q_1 \cup R_3 \cup R_4) = \mu(P_1 \cup R_3) + \mu(Q_1 \cup R_4) \geq \mu(G)(4/7 - o(1)).$$

Which is our desired bound. This completes the analysis for the only remaining case and completes the proof of the theorem. □

### 3.2 Finding 5-augmenting paths

In this section, we slightly modify BIPARTITE-AUG-PATH algorithm to find a small set of 5-augmenting paths. The algorithm receives an initial matching  $\mathcal{M}_0$ , and finds a set of augmenting paths for  $\mathcal{M}_0$ . We use  $G_U$  to denote the induced subgraph between the matched vertices in  $A$  and unmatched vertices in  $B$ , i.e.,  $G_U$  is the induced subgraph between the vertices  $A(\mathcal{M}_0)$  and  $\overline{B(\mathcal{M}_0)}$ . We also use  $G_L$  to denote the induced subgraph between the matched vertices in  $B$  and unmatched vertices in  $A$ . In order to find 3-augmenting paths for the edges in  $\mathcal{M}_0$ , the algorithm should find upper wings from  $G_U$  and lower wings from  $G_L$ . Similar to our algorithm in the previous section, the algorithm uses a very small fraction of the edges to find a matching  $P_1$  in  $G_U$  and  $Q_1$  in  $G_L$ . Let  $\mathcal{M}_P$  and  $\mathcal{M}_Q$  be the matching edges in  $\mathcal{M}_0$  that are incident to the edges in  $P_1$  and  $Q_1$  respectively.

Consider an edge  $e \in \mathcal{M}_P \cap \mathcal{M}_Q$ , we can find an upper wing for this edge using the edges in  $P_1$ , and we can also find a lower wing for this edge using the edges in  $Q_1$ . Therefore we can find  $|\mathcal{M}_P \cap \mathcal{M}_Q|$  disjoint 3-augmenting paths. If the size of  $\mathcal{M}_P \cap \mathcal{M}_Q$  is large, e.g., larger than  $(1/\log n)|\mathcal{M}_0|$ , our algorithm has found a large set of 3-augmenting paths by only using  $2\tau$  fraction of the edges which is very small. We can then switch the edges along these augmenting paths, and recursively find augmenting paths for the new matching. Therefore we can assume the size of  $\mathcal{M}_P \cap \mathcal{M}_Q$  is at most  $o(|\mathcal{M}_0|)$ . In other words we can assume the sets  $\mathcal{M}_P$  and  $\mathcal{M}_Q$  are almost disjoint.

Our algorithm (formally as Algorithm 4) then finds a matching between  $B(\mathcal{M}_P)$  and  $A(\mathcal{M}_Q)$ . Consider an edge  $e$  between  $B(\mathcal{M}_P)$  and  $A(\mathcal{M}_Q)$ , this edge connects two matching edges  $e_0 \in \mathcal{M}_P$  and  $e_1 \in \mathcal{M}_Q$  to each other. Recall using the edges in  $P_1$ , we have a upper wing for  $e_0$ . Let  $e_U$  be the upper wing. Also using the edges in  $Q_1$ , we have a lower wing for  $e_1$ . Let  $e_L$  be the lower wing. Then  $(e_U, e_0, e, e_1, e_L)$  is a 5-augmenting path in our graph. Our algorithm uses a very small fraction of the edges to find a matching between  $B(\mathcal{M}_P)$  and  $A(\mathcal{M}_Q)$ . If it finds a matching with a large size, then we have a large number of 5-augmenting paths. We can exchange the edges along these augmenting paths and recursively solve the problem for the new matching. Otherwise, we can assume that the size of the matching found between  $B(\mathcal{M}_P)$  and  $A(\mathcal{M}_Q)$  is very small. In that case Lemma 3.1 roughly implies that we can store all edges between  $B(\mathcal{M}_P)$  and  $A(\mathcal{M}_Q)$  using a very small memory. We show that our algorithm finds  $(\frac{3}{5} - o(1))$  approximation of the maximum matching.

---

**Algorithm 4:** BIPARTITE-5AUG-PATH Algorithm for augmenting a matching using 3 or 5 augmenting paths.

---

**Data:** A random order stream  $S = \langle e_1, \dots, e_m \rangle$  of edges of a bipartite graph

$G = (A, B, E)$ , and a maximal matching  $\mathcal{M}_0$ .

- 1: Set  $\tau = \frac{1}{100 \log^3 n}$ .
  - 2: Run the following subroutines in parallel.
  - 3:   **Subroutine 1:**
  - 4:     Run GREEDY to find a matching  $P_1$  on the set of edges  $(a, b) \in S_{[1, m \cdot \tau]}$  such that  $a \in A(\mathcal{M}_0)$  and  $b \notin B(\mathcal{M}_0)$ .
  - 5:     Let  $\mathcal{M}_P = \{e \in \mathcal{M}_0 \mid e \cap V(P_1) \neq \emptyset\}$  be the set of edges in  $\mathcal{M}_0$  that are incident to the edges in  $P_1$ .
  - 6:     Run GREEDY to find a matching  $P_2$  on the set of edges  $(a, b) \in S_{(m \cdot \tau, 2m \cdot \tau]}$  such that  $a \notin A(\mathcal{M}_0)$  and  $b \in B(\mathcal{M}_P)$ .
  - 7:   **Subroutine 2:**
  - 8:     Run GREEDY to find a matching  $Q_1$  on the set of edges  $(a, b) \in S_{[1, m \cdot \tau]}$  such that  $a \notin A(\mathcal{M}_0)$  and  $b \in B(\mathcal{M}_0)$ .
  - 9:     Let  $\mathcal{M}_Q = \{e \in \mathcal{M}_0 \mid e \cap V(Q_1) \neq \emptyset\}$  be the set of edges in  $\mathcal{M}_0$  that are incident to the edges in  $Q_1$ .
  - 10:     Run GREEDY to find a matching  $Q_2$  on the set of edges  $(a, b) \in S_{(m \cdot \tau, 2m \cdot \tau]}$  such that  $a \in A(\mathcal{M}_Q)$  and  $b \notin B(\mathcal{M}_0)$ .
  - 11:
  - 12: Run GREEDY to find a matching  $C$  on the set of edges  $(a, b) \in S_{(2m \cdot \tau, 3m \cdot \tau]}$  such that  $a \in A(\mathcal{M}_Q)$  and  $b \in B(\mathcal{M}_P)$ .
  - 13: Let  $T' = P_1 \cup P_2 \cup Q_1 \cup Q_2 \cup C$ .
  - 14: **if** there are at least  $|\mathcal{M}_0| / \log n$  augmenting paths using the edges in  $T'$  **then**
  - 15:   Let  $U$  be maximum augmenting paths for  $\mathcal{M}_0$  using the edges  $T'$ .
  - 16:   Let  $T$  be the output of BIPARTITE-5AUG-PATH with the input of matching  $\mathcal{M}_0 \Delta U$ , and the stream of edges  $S_{(3m \cdot \tau, m]}$ .
  - 17:   **return**  $T \cup \mathcal{M}_0$ .
  - 18: **else**
  - 19:   Run the following subroutines in parallel.
  - 20:   **Subroutine 3:**
  - 21:     Store all edges  $e = (a, b) \in S_{(3m \cdot \tau, m]}$  such that  $a \notin A(\mathcal{M}_0)$  and  $b \in B(\mathcal{M}_P)$  and  $e \cap V(P_2) = \emptyset$ . Let  $R_1$  be the set of these edges.
  - 22:   **Subroutine 4:**
  - 23:     Store all edges  $e = (a, b) \in S_{(3m \cdot \tau, m]}$  such that  $a \in A(\mathcal{M}_Q)$  and  $b \notin B(\mathcal{M}_0)$  and  $e \cap V(Q_2) = \emptyset$ . Let  $R_2$  be the set of these edges.
  - 24:   **Subroutine 5:**
  - 25:     Store all edges  $e = (a, b) \in S_{(3m \cdot \tau, m]}$  such that  $a \in A(\mathcal{M}_0)$  and  $b \notin B(\mathcal{M}_0)$  and  $e \cap V(P_1) = \emptyset$ . Let  $R_3$  be the set of these edges.
  - 26:   **Subroutine 6:**
  - 27:     Store all edges  $e = (a, b) \in S_{(3m \cdot \tau, m]}$  such that  $a \notin A(\mathcal{M}_0)$  and  $b \in B(\mathcal{M}_0)$  and  $e \cap V(Q_1) = \emptyset$ . Let  $R_4$  be the set of these edges.
  - 28:   **Subroutine 7:**
  - 29:     Store all edges  $e = (a, b) \in S_{(3m \cdot \tau, m]}$  such that  $a \in A(\mathcal{M}_Q)$  and  $b \in B(\mathcal{M}_P)$  and  $e \cap V(C) = \emptyset$ . Let  $R_5$  be the set of these edges.
  - 30:
  - 31: Let  $T$  be  $\mathcal{M}_0 \cup P_1 \cup P_2 \cup Q_1 \cup Q_2 \cup C \cup R_1 \cup \dots \cup R_5$ .
  - 32: **return**  $T$ .
-

**Theorem 3.6.** *Given a bipartite graph  $G = (A, B, E)$  with  $n$  vertices and a matching  $\mathcal{M}_0$ . Algorithm BIPARTITE-5AUG-PATH returns a set of edges  $T$  such that  $\mathcal{M}_0 \subseteq T$  and w.h.p. in  $n$ , we have  $|T| = \tilde{O}(n)$  and  $\mu(T \cup R) \geq (\frac{3}{5} - o(1))\mu(G)$  where  $R = E(G \setminus V(\mathcal{M}_0))$  is the set of edges in  $G$  that do not interfere with the edges in  $\mathcal{M}_0$ .*

**Proof.** Consider an arbitrary maximum matching  $\mathcal{M}^*$  in  $G$ . Let  $\mathcal{M}_1^*$  be the set of edges in  $\mathcal{M}^*$  that are incident to the edges in  $\mathcal{M}_0$ , and  $\mathcal{M}_2^* = \mathcal{M}^* \setminus \mathcal{M}_1^*$  be the set of other edges.  $\mathcal{M}_1^*$  and  $\mathcal{M}_2^*$  partition the edges in  $\mathcal{M}^*$ . Therefore, we have  $|\mathcal{M}_1^*| + |\mathcal{M}_2^*| = |\mathcal{M}^*| = \mu(G)$ . For every edge in  $\mathcal{M}_1^*$ , at least one of its endpoints is in  $V(\mathcal{M}_0)$ . Therefore,

$$|\mathcal{M}_0| \leq |\mathcal{M}_1^*| \leq 2|\mathcal{M}_0|.$$

Let's suppose that  $|\mathcal{M}_0| = (1/2 + \delta)|\mathcal{M}_1^*|$  where  $0 \leq \delta \leq 1/2$ . Also, Suppose that  $|\mathcal{M}_1^*| = \alpha|\mathcal{M}^*|$ . Considering the matching  $\mathcal{M}_0 \cup \mu(R)$ , we have

$$\begin{aligned} |\mathcal{M}_0 \cup \mu(R)| &\geq |\mathcal{M}_0| + \mu(R) \\ &\geq |\mathcal{M}_0| + |\mathcal{M}_2^*| && \text{Since every edge in } \mathcal{M}_2^* \text{ is in } R. \\ &= |\mathcal{M}_0| + (1 - \alpha)\mu(G) && \text{Since } |\mathcal{M}_1^*| + |\mathcal{M}_2^*| = \mu(G). \\ &= (1/2 + \delta)|\mathcal{M}_1^*| + (1 - \alpha)\mu(G) \\ &= (1/2 + \delta)\alpha\mu(G) + (1 - \alpha)\mu(G) && \text{Since } |\mathcal{M}_1^*| = \alpha\mu(G). \\ &= \mu(G)(1 - \alpha/2 + \alpha\delta). \end{aligned}$$

Therefore, if  $\alpha/2 - \alpha\delta \leq \frac{2}{5}$ , we have

$$|\mathcal{M}_0 \cup \mu(R)| \geq \mu(G)(1 - \alpha/2 + \alpha\delta) \geq \frac{3}{5}\mu(G).$$

The algorithm returns a set  $T$  such that  $\mathcal{M}_0 \subseteq T$ . Therefore, in this case we have  $\mu(T \cup R) \geq \frac{3}{5}\mu(G)$  which is the desired bound. Therefore, for the rest of the proof we assume that

$$\alpha/2 - \alpha\delta \geq \frac{2}{5}. \tag{7}$$

We use  $G_U$  to denote the induced subgraph between the matched vertices in  $A$  and unmatched vertices in  $B$ , i.e.,  $G_U$  is the induced subgraph between the vertices  $A(\mathcal{M}_0)$  and  $\bar{B}(\mathcal{M}_0)$ . We also use  $G_L$  to denote the induced subgraph between the matched vertices in  $B$  and unmatched vertices in  $A$ . In order to find 3-augmenting paths for the edges in  $\mathcal{M}_0$ , the algorithm should find upper wings from  $G_U$  and lower wings from  $G_L$ . Let  $\tau = \frac{1}{100 \log^3 n}$ , the algorithm uses  $\tau$  fraction of the edges and runs GREEDY algorithm to find a matching  $P_1$  (respectively,  $Q_1$ ) in  $G_U$  (resp.,  $G_L$ ). Let  $E' \subseteq E$  be the set of edges in  $E$  that are incident to the vertices in  $V(\mathcal{M}_0)$ . Matching  $\mathcal{M}_0$  is maximal among the edges in  $E'$ . Also,  $\mathcal{M}_1^*$  is a matching in  $E'$ . Therefore,  $\mu(E') \geq |\mathcal{M}_1^*|$ . By Claim 3.4 there exists a matching of size  $(1/2 - \delta)|\mathcal{M}_1^*|$  in both  $G_U$  and  $G_L$ . Let  $\mathcal{M}_P$  be the edges in  $\mathcal{M}_0$  that are incident to the edges in  $P_1$ . In other words,  $\mathcal{M}_P = \{e \in \mathcal{M}_0 | e \cap V(P_1) \neq \emptyset\}$ . Similarly, we define  $\mathcal{M}_Q$  to be  $\mathcal{M}_Q = \{e \in \mathcal{M}_0 | e \cap V(Q_1) \neq \emptyset\}$ . Each edge in  $\mathcal{M}_P$  is matched by the matching  $P_1$ . Therefore, we have an upper wing for every edge in  $\mathcal{M}_P$ . If we can find lower wings for these edges, then we have found 3-augmenting paths for these edges, and we can increase the size of the matching.

In order to find lower wings, the algorithm in (Subroutine 1) uses another  $\tau$  fraction of the edges and calls the greedy algorithm to match vertices in  $B(\mathcal{M}_P)$  to the vertices in  $A(G_L)$ . Let  $P_2$  be the result of this matching. We can use edges in  $P_1$  and  $P_2$  to find  $\min\{|P_1|, |P_2|\} = |P_2|$ , 3-augmenting paths for  $\mathcal{M}_0$ . If  $|P_2| \geq (1/\log n)|\mathcal{M}_0|$ , then by switching the edges along the augmenting paths, we

can increase the size of  $\mathcal{M}_0$  by the factor of  $(1 + 1/\log n)$ . In this case, the algorithm increases the size of  $\mathcal{M}_0$  and recursively calls itself to find augmenting paths for the new matching. Eventually, the algorithm reaches a state such that it cannot increase the size of matching using the edges in  $P_1$  and  $P_2$ . We then have  $|P_2| < (1/\log n)|\mathcal{M}_0|$ . The algorithm also finds a matching  $Q_2$  between  $A(\mathcal{M}_Q)$  and  $B(G_U)$ . Using a similar argument for  $Q_1$  and  $Q_2$ , we can say that the algorithm eventually reaches a state that we have  $|Q_2| < (1/\log n)|\mathcal{M}_0|$ . Also, consider an edge in  $\mathcal{M}_P \cap \mathcal{M}_Q$ , for these edges we have both upper and lower wings. Therefore, we have augmenting paths for these edges. If the size of  $\mathcal{M}_P \cap \mathcal{M}_Q$  is large, we can increase the size of matching  $\mathcal{M}_0$  by these augmenting paths, and recursively find augmenting paths for the new matching. Thus, the algorithm eventually reaches a state such that we have  $|\mathcal{M}_P \cap \mathcal{M}_Q| < (1/\log n)|\mathcal{M}_0|$ .

Our algorithm then uses another  $\tau$  fraction of the edges to find a matching  $C$  between vertices  $B(\mathcal{M}_P)$  and  $A(\mathcal{M}_Q)$ . We claim that we can find  $|C|$  augmenting paths using these edges. Consider an edge  $e \in C$ , this edge is between  $B(\mathcal{M}_P)$  and  $A(\mathcal{M}_Q)$ , and it connects two matching edges  $e_0 \in \mathcal{M}_P$  and  $e_1 \in \mathcal{M}_Q$  to each other. If  $e_0 = e_1$ , then we already have both upper and lower wings for  $e_0$ . Therefore, we have an augmenting path for  $e_0$ . Consider the case that  $e_0 \neq e_1$ . Recall that using the edges in  $P_1$ , we have an upper wing for  $e_0$ . Let  $e_U$  be the upper wing. Also using the edges in  $Q_1$ , we have a lower wing for  $e_1$ . Let  $e_L$  be the lower wing. Then  $(e_U, e_0, e, e_1, e_L)$  is a 5-augmenting path in our graph. Therefore we have  $|C|$  augmenting paths. If the size of  $C$  is large, we can increase the size of matching  $\mathcal{M}_0$  by these augmenting paths, and recursively find augmenting paths for the new matching. Thus, the algorithm eventually reaches a state such that we have  $|C| < (1/\log n)|\mathcal{M}_0|$ .

Let  $t$  be the number of recursive calls in our algorithm. In each recursive call the algorithm increases the size of  $\mathcal{M}_0$  by a multiplicative factor of  $(1 + 1/\log n)$ . Therefore, we have  $t \leq \log_{1+1/\log n} n$  which is at most  $2 \log^2 n$  for a large enough  $n$ . Before each recursive call, the algorithm uses at most  $3\tau$  fraction of the edges. Therefore, by the end of all recursive calls the algorithm has only used  $t \cdot 3\tau \cdot m \leq m/(16 \log n)$  edges which is only a small fraction of all edges. Since edges arrive in a uniformly random order, for each edge in  $\mathcal{M}^*$  the probability that this edge is within the first  $1/(16 \log n)$  fraction of the edges is  $1/(16 \log n)$ . Hence, the expected number of edges in  $\mathcal{M}^*$  that are in the first  $1/(16 \log n)$  fraction of the edges is  $|\mathcal{M}^*|/(16 \log n)$ . Using Chernoff bound we can show that the probability that more than  $|\mathcal{M}^*|/(8 \log n)$  of these edges are in the first  $1/(16 \log n)$  fraction of the edges is bounded by  $n^{-10}$  for a large enough  $n$ . Therefore, w.h.p. the size of the maximum matching in the remaining stream is at least

$$|\mathcal{M}^*| \left(1 - \frac{1}{8 \log n}\right) = \mu(G) \left(1 - \frac{1}{8 \log n}\right) = \mu(G)(1 - o(1)).$$

Recall that by the end of all recursive calls the size of each of  $P_2, Q_2, (\mathcal{M}_P \cap \mathcal{M}_Q)$  and  $C$  is bounded by  $|\mathcal{M}_0|/\log n$ .

By a slight abuse of notion, we suppose that  $\mathcal{M}^*$  is a maximum matching in the remaining stream. We also suppose that  $\mathcal{M}_1^*$  is the set of edges in  $\mathcal{M}^*$  that are incident to the edges in  $\mathcal{M}_0$ , and  $\mathcal{M}_2^* = \mathcal{M}^* \setminus \mathcal{M}_1^*$  is the set of other edges in  $\mathcal{M}^*$ . Therefore,

$$|\mathcal{M}^*| = |\mathcal{M}_1^*| + |\mathcal{M}_2^*| = \mu(G)(1 - o(1)).$$

Let  $R_P$  be the set of edges in  $\mathcal{M}_2^*$  that are incident to the edges in  $P_1$ . In other words,

$$R_P = \{e \in \mathcal{M}_2^* \mid e \cap V(P_1) \neq \emptyset\}.$$

Similarly, we define  $R_Q$  to be

$$R_Q = \{e \in \mathcal{M}_2^* \mid e \cap V(Q_1) \neq \emptyset\}.$$

Also, let  $\mathcal{M}_C^*$  to be the edges in  $\mathcal{M}^*$  that are between vertices  $A(\mathcal{M}_Q)$  and  $B(\mathcal{M}_P)$ , i.e.,

$$\mathcal{M}_C^* = \{(a, b) \in \mathcal{M}^* \mid a \in A(\mathcal{M}_Q), b \in B(\mathcal{M}_P)\}.$$

Then, all edges in  $\mathcal{M}_C^*$  are also in  $\mathcal{M}_1^*$ .

In order to prove the theorem, we consider the following cases.

- **Case 1:**  $|\mathcal{M}_C^*| \geq (1/5 - 2\delta\alpha)\mu(G)$ .

Every edge in  $\mathcal{M}_C^*$  is between vertices  $A(\mathcal{M}_Q)$  and  $B(\mathcal{M}_P)$ . As we previously discussed, every edge in  $\mathcal{M}_C^*$  forms an augmenting path. We show that our algorithm finds a large number of these edges. The algorithm first uses a small fraction of the edges to find a matching  $C$  between vertices  $A(\mathcal{M}_Q)$  and  $B(\mathcal{M}_P)$ . According to what we discussed, we can assume  $|C| \leq |\mathcal{M}_0|/\log n$ . By Lemma 3.1, we know that with w.h.p. we can store all of the edges between vertices  $A(\mathcal{M}_Q)$  and  $B(\mathcal{M}_P)$  that are not incident to the edges in  $C$  within a memory of size  $\tilde{O}(n/\tau) = \tilde{O}(n)$ . The algorithm stores all these edges in (Subroutine 7). Let  $R_5$  be the set of these edges, and  $T$  be  $P_1 \cup Q_1 \cup R_5$ . We claim that  $\mu(T \cup \mathcal{M}_0) \geq (\frac{3}{5} - o(1))\mu(G)$ . We first show that  $\mu(R_5) \geq |\mathcal{M}_C^*| - o(\mu(G))$ . Set  $R_5$  contains all edges between vertices  $A(\mathcal{M}_Q)$  and  $B(\mathcal{M}_P)$  that are not incident to the edges in  $C$ . Since there are at most  $2|C|$  edges in  $\mathcal{M}_C^*$  that are incident to a vertex in  $V(C)$ , we have

$$\begin{aligned} \mu(R_5) &\geq |\mathcal{M}_C^*| - 2|C| \\ &\geq |\mathcal{M}_C^*| - 2|\mathcal{M}_0|/\log n \\ &\geq |\mathcal{M}_C^*| - 2\mu(G)/\log n \\ &= |\mathcal{M}_C^*| - o(\mu(G)). \end{aligned} \tag{8}$$

Each edge in  $R_5$  is between vertices  $A(\mathcal{M}_Q)$  and  $B(\mathcal{M}_P)$ . According to what we previously discussed, each edge in  $R_5$  along with the edges in  $P_1$  and  $Q_1$  forms 3 or 5 augmenting paths for  $\mathcal{M}_0$ . Therefore, we can find  $\mu(R_5)$  disjoint augmenting paths. It follows that the size of the matching in  $T \cup \mathcal{M}_0$  is at least

$$\begin{aligned} &\mu(T \cup \mathcal{M}_0) \\ &\geq |\mathcal{M}_0| + \mu(R_5) \\ &\geq |\mathcal{M}_0| + |\mathcal{M}_C^*| - o(\mu(G)) && \text{By (8).} \\ &\geq |\mathcal{M}_0| + (1/5 - 2\delta\alpha - o(1))\mu(G) && \text{By the assumption of Case 1.} \\ &= (1/2 + \delta)|\mathcal{M}_1^*| + (1/5 - 2\delta\alpha - o(1))\mu(G) && \text{Since } |\mathcal{M}_0| = (1/2 + \delta)|\mathcal{M}_1^*|. \\ &= (1/2 + \delta - o(1))\alpha\mu(G) + (1/5 - 2\delta\alpha - o(1))\mu(G) && \text{Since } |\mathcal{M}_1^*| = \alpha|\mathcal{M}^*|. \\ &= (1/5 + \alpha/2 - \delta\alpha - o(1))\mu(G) \\ &\geq (3/5 - o(1))\mu(G), && \text{By (7).} \end{aligned}$$

which is the desired bound, and it completes the proof for this case.

- **Case 2:**  $|P_1| - |R_P| + |Q_1| - |Q_P| \leq (\alpha - 2\alpha\delta - 3/5)\mu(G)$ .

Recall that  $P_1$  and  $Q_1$  are the result of the greedy matching algorithm using  $\tau$  fraction of the edges in  $G_U$  and  $G_L$ . Therefore, by Lemma 3.1, we know that with w.h.p. we can store all of the edges in  $G_U \setminus V(P_1)$  and also all the edges in  $G_L \setminus V(Q_1)$  within a memory of size  $\tilde{O}(n/\tau) = \tilde{O}(n)$  which is within the memory bound of the semi-streaming algorithms.

The algorithm stores all the edges in  $G_U \setminus V(P_1)$  and also all the edges in  $G_L \setminus V(Q_1)$  in (Subroutine 5) and (Subroutine 6) respectively. Let  $R_3$  and  $R_4$  be the set of these edges respectively, and  $T$  be  $P_1 \cup Q_1 \cup R_3 \cup R_4$ . We show that  $\mu(T) \geq (\frac{3}{5} - o(1))\mu(G)$ .

By Claim 3.4 we know there exists a matching of size at least  $(1/2 - \delta)|\mathcal{M}_1^*|$  in both  $G_U$  and  $G_L$  using the edges in  $\mathcal{M}_1^*$ . Therefore, there exists a matching of size at least  $(1 - 2\delta)|\mathcal{M}_1^*|$  in  $G_U \cup G_L$ . Let  $\mathcal{M}$  be this matching, and  $\mathcal{M}'$  be the set of the edges in  $\mathcal{M}$  that interfere with the edges in  $P_1$  or  $Q_1$ , i.e.,  $\mathcal{M}' = \{e \in \mathcal{M} \mid e \cap (V(P_1) \cup V(Q_1)) \neq \emptyset\}$ . We show that the size of  $\mathcal{M}'$  is at most  $2|P_1| + 2|Q_1| - |R_P| - |R_Q|$ . The total number of vertices covered by  $P_1$  and  $Q_1$  is exactly  $2|P_1| + 2|Q_1|$ . However, we know that  $|R_P|$  edges from  $P_1$  cover at least one vertex from  $\mathcal{M}_2^*$ . Since vertices of  $\mathcal{M}_1^*$  and  $\mathcal{M}_2^*$  are disjoint, it implies that the number of edges from  $\mathcal{M}_1^*$  that are covered by  $P_1$  is at most  $2|P_1| - |R_P|$ . Using a similar argument, the number of edges from  $\mathcal{M}_1^*$  that are covered by  $Q_1$  is at most  $2|Q_1| - |R_Q|$ . Thus,

$$|\mathcal{M}'| \leq 2|P_1| + 2|Q_1| - |R_P| - |R_Q|. \quad (9)$$

Note that edges in  $\mathcal{M} \setminus \mathcal{M}'$  do not share a vertex with the edges in  $P_1$  or  $Q_1$ . Therefore, the size of maximum matching in  $P_1 \cup Q_1 \cup R_3 \cup R_4$  is at least

$$\begin{aligned} & |P_1| + |Q_1| + |\mathcal{M}| - |\mathcal{M}'| \\ & \geq |P_1| + |Q_1| + |\mathcal{M}| - (2|P_1| + 2|Q_1| - |R_P| - |R_Q|) && \text{By (9).} \\ & = |R_P| + |R_Q| - |P_1| - |Q_1| + |\mathcal{M}| \\ & \geq |\mathcal{M}| - (\alpha - 2\alpha\delta - 3/5)\mu(G). && \text{By the assumption of Case 2.} \end{aligned}$$

Since the size  $\mathcal{M}$  is at least  $(1 - 2\delta)|\mathcal{M}_1^*|$ , we have

$$\begin{aligned} \mu(T) & \geq |\mathcal{M}| - (\alpha - 2\alpha\delta - 3/5)\mu(G) \\ & \geq (1 - 2\delta)|\mathcal{M}_1^*| - (\alpha - 2\alpha\delta - 3/5)\mu(G) \\ & = (1 - 2\delta)\alpha|\mathcal{M}_1^*| - (\alpha - 2\alpha\delta - 3/5)\mu(G) \\ & = (1 - 2\delta - o(1))\alpha\mu(G) - (\alpha - 2\alpha\delta - 3/5)\mu(G) \\ & = \mu(G)(3/5 - o(1)). \end{aligned}$$

Which is our desired bound. This completes the analysis for this case.

- **Case 3:**  $|\mathcal{M}_C^*| < (1/5 - 2\delta\alpha)\mu(G)$  and  $|P_1| - |R_P| + |Q_1| - |Q_P| > (\alpha - 2\alpha\delta - 3/5)\mu(G)$ .

The only remaining case is when  $|\mathcal{M}_C^*| < (1/5 - 2\delta\alpha)\mu(G)$  and  $|P_1| - |R_P| + |Q_1| - |Q_P| > (\alpha - 2\alpha\delta - 3/5)\mu(G)$ . Recall that  $\mathcal{M}_P$  is the set of edges in  $\mathcal{M}_0$  that are incident to the edges in  $P_1$ . Since  $P_1$  is a matching in  $G_U$ , each edge in  $\mathcal{M}_P$  has an upper wing, and if we find lower wings for them, we have a set of 3-augmenting path. The Algorithm in (Subroutine 1) picks a small fraction of edges and finds a greedy matching  $P_2$  between the vertices in  $B(\mathcal{M}_P)$  and  $A(G_L)$ . As we discussed before, we can assume that  $|P_2| \leq |\mathcal{M}_0|/\log n$ . Thus, we have  $|P_2| = o(\mu(G))$ . Therefore, the greedy algorithm has matched only a small fraction of the vertices in the induced graph between  $B(\mathcal{M}_P)$  and  $A(G_L)$ . Let  $G_P$  be the induced subgraph between vertices  $B(\mathcal{M}_P)$  and  $A(G_L)$ . By Lemma 3.1 we have that w.h.p. the number of edges in the graph  $G_P \setminus V(P_2)$  is bounded by  $\tilde{O}(\frac{n}{\tau}) = \tilde{O}(n)$  and we can store all of them in the memory bound of our semi-streaming algorithm. The algorithm saves all of these edges in (Subroutine 3). Let  $R_1$  be the set of these edges. Similarly, let  $G_Q$  be the induced subgraph between vertices  $A(\mathcal{M}_Q)$  and  $B(G_U)$ . By Lemma 3.1 we have that w.h.p. the number of

edges in the graph  $G_Q \setminus V(Q_2)$  is bounded by  $\tilde{O}(\frac{n}{\tau}) = \tilde{O}(n)$  and we can store all of them in the memory bound of our semi-streaming algorithm. Let  $R_2$  be the set of these edges.

We claim that  $\mu(G_P) + \mu(G_Q)$  is at least  $|P_1| + |Q_1| - \mu(G)(1/5 + o(1))$ . Every edge of  $\mathcal{M}_1^*$  is covered by the vertices of  $\mathcal{M}_0$ . Recall that set  $\mathcal{M}_C^*$  is the set of edges in  $\mathcal{M}_1^*$  that are between vertices  $A(\mathcal{M}_Q)$  and  $B(\mathcal{M}_P)$ . Let  $V' = V(\mathcal{M}_0) \setminus (A(\mathcal{M}_Q) \cup B(\mathcal{M}_P))$ . It follows that the number of edges in  $\mathcal{M}_1^*$  that are not covered by the vertices in  $V'$  is bounded by

$$\mu(G_P) + \mu(G_Q) + |\mathcal{M}_C^*|.$$

Therefore we should have

$$\begin{aligned} |\mathcal{M}_1^*| &\leq |V'| + \mu(G_P) + \mu(G_Q) + |\mathcal{M}_C^*| \\ &= 2|\mathcal{M}_0| - |P_1| - |Q_1| + \mu(G_P) + \mu(G_Q) + |\mathcal{M}_C^*|. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mu(G_P) + \mu(G_Q) \\ &\geq |P_1| + |Q_1| + |\mathcal{M}_1^*| - 2|\mathcal{M}_0| - |\mathcal{M}_C^*| \\ &= |P_1| + |Q_1| + (\alpha - o(1))\mu(G) - 2|\mathcal{M}_0| - |\mathcal{M}_C^*| && \text{Since } |\mathcal{M}_1^*| = \alpha|\mathcal{M}^*|. \\ &= |P_1| + |Q_1| - (2\delta\alpha + o(1))\mu(G) - |\mathcal{M}_C^*| && \text{Since } |\mathcal{M}_0| = (1/2 + \delta)|\mathcal{M}_1^*|. \\ &\geq |P_1| + |Q_1| - \mu(G)(1/5 + o(1)). && \text{Since } |\mathcal{M}_C^*| < (1/5 - 2\delta\alpha)\mu(G). \end{aligned} \quad (10)$$

The algorithm stores in sets  $R_1$  and  $R_2$  all edges in  $G_P$  and  $G_Q$  that do not interfere with the edges in  $P_2$  and  $Q_2$ . Since the size of  $P_2$  and  $Q_2$  is at most  $o(\mu(G))$ , we have

$$\begin{aligned} \mu(R_1) + \mu(R_2) &\geq \mu(G_P) + \mu(G_Q) - 2|P_2| - 2|Q_2| \\ &= \mu(G_P) + \mu(G_Q) - o(\mu(G)) \\ &\geq |P_1| + |Q_1| - \mu(G)(1/5 + o(1)). && \text{By (10).} \end{aligned} \quad (11)$$

Since we store all of the edges in  $R_1$  in our memory, we can find a matching of size  $\mu(R_1)$  in  $G_P$  such that the edges of this matching do not interfere with the edges in  $\mathcal{M}_2^*$ . Each matching edge in this subgraph will be a lower wing for an edge in  $\mathcal{M}_P$ . Therefore, the algorithm finds  $\mu(R_1)$  augmenting paths. These augmenting paths interfere with at most  $|R_P|$  edges of  $\mathcal{M}_2^*$ . Using these augmenting paths, we can find a matching of size at least

$$|\mathcal{M}_0| + \mu(R_1) - |R_P| + |\mathcal{M}_2^*|. \quad (12)$$

Similarly, using the edges in  $R_2$ , we can get a matching of size at least

$$|\mathcal{M}_0| + \mu(R_2) - |R_Q| + |\mathcal{M}_2^*|. \quad (13)$$

We show that the summation of sizes of these matchings is at least  $\mu(G)(6/5 - o(1))$ , therefore the size of at least one of them is  $\mu(G)(3/5 - o(1))$ . By taking the summation of (12) and

(13) we get

$$\begin{aligned}
& 2|\mathcal{M}_0| + \mu(R_1) + \mu(R_2) - |R_P| - |R_Q| + 2|\mathcal{M}_2^*| \\
& \geq 2|\mathcal{M}_0| + 2|\mathcal{M}_2^*| - \mu(G)(1/5 + o(1)) \\
& \quad + |P_1| + |Q_1| - |R_P| - |R_Q| \qquad \qquad \qquad \text{By (11).} \\
& \geq 2|\mathcal{M}_0| + 2|\mathcal{M}_2^*| + \mu(G)(\alpha - 2\alpha\delta - 4/5 - o(1)) \qquad \text{By the assumption of Case 3.} \\
& = (1 + 2\delta)|\mathcal{M}_1^*| + 2|\mathcal{M}_2^*| + \mu(G)(\alpha - 2\alpha\delta - 4/5 - o(1)) \qquad \text{Since } |\mathcal{M}_0| = (1/2 + \delta)|\mathcal{M}_1^*|. \\
& = 2|\mathcal{M}_2^*| + \mu(G)(2\alpha - 4/5 - o(1)) \qquad \qquad \qquad \text{Since } |\mathcal{M}_1^*| = \alpha|\mathcal{M}^*|. \\
& = \mu(G)(2 - 4/5 - o(1)) \qquad \qquad \qquad \text{Since } |\mathcal{M}_2^*| = (1 - \alpha)|\mathcal{M}^*|. \\
& = \mu(G)(6/5 - o(1)).
\end{aligned}$$

Therefore, the size of the matching found by the algorithm is at least  $\mu(G)(3/5 - o(1))$  which is the desired bound. This completes the analysis for the last remaining case and completes the proof of the theorem. □

## 4 Reduction from General Graphs to Bipartite Graphs

---

**Algorithm 5:** Algorithm MATCHING-VIA-BIPARTITE for finding an approximate matching in general graphs.

---

- Data:** A random order stream  $S = \langle e_1, \dots, e_m \rangle$  of edges of a bipartite graph  $G = (A, B, E)$ , and a semi-streaming algorithm ALG for finding a matching in bipartite graphs.
- 1: Run GREEDY on the edges  $S_{[1, m/\log n]}$  to find a matching  $\mathcal{M}_0$ .
  - 2: Run the following subroutines in parallel.
  - 3:   **Subroutine 1:**
  - 4:     Let  $G' = (V(\mathcal{M}_0), \overline{V(\mathcal{M}_0)})$  be a bipartite graph such that an each  $e$  is in  $G'$  if exactly one of its endpoints is in  $V(\mathcal{M}_0)$ .
  - 5:     Run ALG to find a matching  $\mathcal{M}_1$  using the stream of edges  $e \in S_{(m/\log n, m]}$  such that  $e \in E(G')$ .
  - 6:   **Subroutine 2:**
  - 7:     Store all edges  $e \in S_{(m/\log n, m]}$  such that  $e \cap V(\mathcal{M}_0) = \emptyset$ . Let  $R$  be the set of these edges.
  - 8: **return**  $\mu(\mathcal{M}_0 \cup R \cup \mathcal{M}_1)$ .
- 

In this section, we present a black-box reduction from the finding a maximum matching in general graphs to that of the finding a maximum matching in bipartite graphs when edges arrive in a uniformly random order. The algorithm is formalized in Algorithm 5. Algorithm 5 gets a graph  $G = (V, E)$ , and an arbitrary semi-streaming algorithm for finding a maximum matching in bipartite graphs. Assuming that the approximation ratio of the bipartite algorithm is  $p$  when edges arrives in a random order, Algorithm 5 is guaranteed to have an approximation ratio of  $\frac{2p}{2p+1}$  w.h.p. in  $n$  where  $n$  is the number of vertices in  $G$ . Using the fact that there exists a single-pass semi-streaming algorithm that finds  $3/5 - o(1)$  approximation of maximum matching in bipartite graphs w.h.p. in  $n$ , this immediately implies the following theorem.

**Theorem 1.3.** *There exists a deterministic single-pass semi-streaming algorithm that finds w.h.p. a  $(\frac{6}{11} - o(1)) (\approx 0.545)$  approximation of the maximum matching in general graphs when edges arrive in a random order.*

Now we show that the approximation ratio of our reduction method is  $\frac{2p}{2p+1}$ .

**Theorem 1.2.** *Given any single-pass semi-streaming algorithm with the approximation ratio of  $p$  for finding a maximum matching in bipartite graphs in random arrival of edges, there exists a semi-streaming algorithm that finds w.h.p. a  $(\frac{2p}{2p+1} - o(1))$  approximation of the maximum matching when edges arrive in a random order.*

**Proof.** The algorithm picks the first  $1/\log n$  fraction of the edges and runs the greedy algorithm to find a matching  $\mathcal{M}_0$ . Let  $R$  be the set of edges in the graph that do not interfere with the edges in  $\mathcal{M}_0$ , i.e.,  $R = E(G \setminus V(\mathcal{M}_0))$ . By Lemma 3.1, we have that w.h.p., the number of edges in  $R$  is bounded by  $\tilde{O}(n)$  and all of them can be stored within the memory bound of semi-streaming algorithms. Consider an arbitrary maximum matching  $\mathcal{M}^*$  among the remaining  $(1 - 1/\log n)$  fraction of the edges. Each edge  $e \in \mu(G)$  is in the last  $(1 - 1/\log n)$  fraction of the edges with the probability of  $(1 - 1/\log n)$ . Therefore,  $\mathbb{E}[|\mathcal{M}^*|] \geq (1 - o(1))\mu(G)$ . Recall that we can assume  $\mu(G) \geq \sqrt{n}$ . In that case, with the standard application of Chernoff bound it can be shown that w.h.p. in  $n$ , we have  $|\mathcal{M}^*| \geq (1 - o(1))\mu(G)$ . Let  $\mathcal{M}_1^*$  be the set of edges in  $\mathcal{M}^*$  that are incident to the edges in  $\mathcal{M}_0$  and  $\mathcal{M}_2^*$  be the set of all other edges. Let  $|\mathcal{M}_1^*| = \alpha\mu(G)$ . It implies that

$$|\mathcal{M}_2^*| = |\mathcal{M}^*| - |\mathcal{M}_1^*| \geq (1 - \alpha - o(1))\mu(G). \quad (14)$$

Since  $\mathcal{M}_0$  is a maximal matching in the graph  $\mathcal{M}_1^*$ , we have  $|\mathcal{M}_0| = (1/2 + \delta)|\mathcal{M}_1^*|$  where  $0 \leq \delta \leq 1/2$ . Since our semi-streaming algorithm can store all of the edges in  $R$ , it can output  $\mu(\mathcal{M}_0 \cup R)$  as a matching. The size of this matching is at least,

$$\begin{aligned} \mu(\mathcal{M}_0 \cup R) &= |\mathcal{M}_0| + \mu(R) \\ &\geq |\mathcal{M}_0| + |\mathcal{M}_2^*| \\ &\geq |\mathcal{M}_0| + (1 - \alpha - o(1))\mu(G) && \text{By (14).} \\ &= (1/2 + \delta)|\mathcal{M}_1^*| + (1 - \alpha - o(1))\mu(G) \\ &= (1/2 + \delta)\alpha\mu(G) + (1 - \alpha - o(1))\mu(G) && \text{Since } |\mathcal{M}_1^*| = \alpha\mu(G). \\ &= \mu(G)(1 - \alpha/2 + \alpha\delta - o(1)). \end{aligned}$$

Consider the case that  $\frac{\alpha}{2} - \alpha\delta \leq \frac{1}{1+2p}$ , in that case we have

$$\begin{aligned} \mu(\mathcal{M}_0 \cup R) &\geq \mu(G)(1 - \alpha/2 + \alpha\delta - o(1)) \\ &\geq \mu(G)(1 - \frac{1}{1+2p} - o(1)) \\ &= \mu(G)(\frac{2p}{1+2p} - o(1)). \end{aligned}$$

Which is our desired bound.

The remaining case is when  $\frac{\alpha}{2} - \alpha\delta > \frac{1}{1+2p}$ . In this case, the algorithm creates a bipartite graph between the matched vertices and unmatched vertices by  $\mathcal{M}_0$ . Specifically, let  $G' = (V(\mathcal{M}_0), \overline{V(\mathcal{M}_0)})$  be a bipartite graph that has the matched vertices on one of its side and unmatched vertices on the other side. We add an edge  $e$  to  $G'$  if it is between matched and unmatched vertices. We claim that  $\mu(G') \geq 2|\mathcal{M}_1^*| - 2|\mathcal{M}_0|$ . To show that our claim holds, note that  $\mathcal{M}_0$  is a maximal matching among the edges in  $\mathcal{M}_1^*$ . There are  $2|\mathcal{M}_0|$  vertices that are covered by  $\mathcal{M}_0$  and  $2|\mathcal{M}_1^*|$  vertices that are covered by  $\mathcal{M}_1^*$ . Consider  $2|\mathcal{M}_1^*| - 2|\mathcal{M}_0|$  vertices that are

covered by  $\mathcal{M}_1^*$  but are not covered by  $\mathcal{M}_0$ . Since  $\mathcal{M}_0$  is maximal, each edge in  $\mathcal{M}_1^*$  that covers these vertices is between  $V(\mathcal{M}_0)$  and  $\overline{V(\mathcal{M}_0)}$ . Therefore,  $\mu(G') \geq 2|\mathcal{M}_1^*| - 2|\mathcal{M}_0|$ . Let ALG be the algorithm for the bipartite matching. Since the approximation ratio of ALG is  $p$ , using this algorithm we can find  $p$  approximation of the maximum matching in  $G'$ . Let  $\mathcal{M}_1$  be the matching returned by ALG in  $G'$ . Then, we have

$$\begin{aligned}
|\mathcal{M}_1| &\geq p \cdot \mu(G') \\
&\geq p(2|\mathcal{M}_1^*| - 2|\mathcal{M}_0|) \\
&= 2p(|\mathcal{M}_1^*| - |\mathcal{M}_0|) \\
&= 2p(\alpha\mu(G) - |\mathcal{M}_0|) && \text{Since } |\mathcal{M}_1^*| = \alpha\mu(G). \\
&= 2p(\alpha\mu(G) - (1/2 + \delta)\alpha\mu(G)) && \text{Since } |\mathcal{M}_0| = (1/2 + \delta)|\mathcal{M}_1^*|. \\
&= 2p \cdot \mu(G)(\alpha/2 - \alpha\delta) \\
&> 2p \cdot \mu(G)\left(\frac{1}{1 + 2p}\right) && \text{Since } \alpha/2 - \alpha\delta > \frac{1}{1 + 2p}. \\
&= \frac{2p}{1 + 2p}\mu(G),
\end{aligned}$$

which is the desired bound and completes the proof for the reduction.  $\square$

## References

- [1] Assadi, S., Bateni, M., Bernstein, A., Mirrokni, V. S., and Stein, C. (2019). Coresets meet EDCS: algorithms for matching and vertex cover on massive graphs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 1616–1635.
- [2] Bernstein, A. and Stein, C. (2015). Fully dynamic matching in bipartite graphs. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 167–179.
- [3] Chitnis, R., Cormode, G., Esfandiari, H., Hajiaghayi, M., McGregor, A., Monemizadeh, M., and Vorotnikova, S. (2016). Kernelization via sampling with applications to finding matchings and related problems in dynamic graph streams. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1326–1344.
- [4] Crouch, M. and Stubbs, D. S. (2014). Improved streaming algorithms for weighted matching, via unweighted matching. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014, September 4-6, 2014, Barcelona, Spain*, pages 96–104.
- [5] Epstein, L., Levin, A., Mestre, J., and Segev, D. (2011). Improved approximation guarantees for weighted matching in the semi-streaming model. *SIAM J. Discrete Math.*, 25(3):1251–1265.
- [6] Esfandiari, H., Hajiaghayi, M., and Monemizadeh, M. (2016). Finding large matchings in semi-streaming. In *IEEE International Conference on Data Mining Workshops, ICDM Workshops 2016, December 12-15, 2016, Barcelona, Spain*, pages 608–614.

- [7] Feigenbaum, J., Kannan, S., McGregor, A., Suri, S., and Zhang, J. (2005a). Graph distances in the streaming model: the value of space. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005, Vancouver, British Columbia, Canada, January 23-25, 2005*, pages 745–754.
- [8] Feigenbaum, J., Kannan, S., McGregor, A., Suri, S., and Zhang, J. (2005b). On graph problems in a semi-streaming model. *Theor. Comput. Sci.*, 348(2-3):207–216.
- [9] Gamlath, B., Kale, S., Mitrovic, S., and Svensson, O. (2018). Weighted matchings via unweighted augmentations. *CoRR*, abs/1811.02760.
- [10] Ghaffari, M. and Wajc, D. (2019). Simplified and space-optimal semi-streaming  $(2+\epsilon)$ -approximate matching. In *2nd Symposium on Simplicity in Algorithms, SOSA@SODA 2019, January 8-9, 2019 - San Diego, CA, USA*, pages 13:1–13:8.
- [11] Goel, A., Kapralov, M., and Khanna, S. (2012). On the communication and streaming complexity of maximum bipartite matching. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 468–485.
- [12] Kale, S. and Tirodkar, S. (2017). Maximum matching in two, three, and a few more passes over graph streams. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2017, August 16-18, 2017, Berkeley, CA, USA*, pages 15:1–15:21.
- [13] Kapralov, M. (2013). Better bounds for matchings in the streaming model. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 1679–1697.
- [14] Kapralov, M., Khanna, S., and Sudan, M. (2014). Approximating matching size from random streams. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 734–751.
- [15] Konrad, C. (2018). A simple augmentation method for matchings with applications to streaming algorithms. In *43rd International Symposium on Mathematical Foundations of Computer Science, MFCS 2018, August 27-31, 2018, Liverpool, UK*, pages 74:1–74:16.
- [16] Konrad, C., Magniez, F., and Mathieu, C. (2012). Maximum matching in semi-streaming with few passes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 15th International Workshop, APPROX 2012, and 16th International Workshop, RANDOM 2012, Cambridge, MA, USA, August 15-17, 2012. Proceedings*, pages 231–242.
- [17] McGregor, A. (2005). Finding graph matchings in data streams. In *Approximation, Randomization and Combinatorial Optimization, Algorithms and Techniques, 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th International Workshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22-24, 2005, Proceedings*, pages 170–181.
- [18] Paz, A. and Schwartzman, G. (2017). A  $(2 + \epsilon)$ -approximation for maximum weight matching in the semi-streaming model. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 2153–2161.

- [19] Plummer, M. D. and Lovász, L. (1986). *Matching theory*, volume 29. Elsevier.
- [20] Wajc, D. (2017). Negative association: definition, properties, and applications.
- [21] Zelke, M. (2012). Weighted matching in the semi-streaming model. *Algorithmica*, 62(1-2):1–20.

## A Omitted proofs

**Claim 3.4.** *Given a bipartite graph  $G = (A, B, E)$  and a maximal matching  $\mathcal{M}$ , suppose that  $|\mathcal{M}| = (1/2 + \delta)\mu(G)$  where  $0 \leq \delta \leq 1/2$ , then there exists a matching of size at least  $(1/2 - \delta)\mu(G)$  between vertices  $A(\mathcal{M})$  and  $\overline{B(\mathcal{M})}$  and also between vertices  $B(\mathcal{M})$  and  $\overline{A(\mathcal{M})}$ .*

**Proof.** Pick any arbitrary maximum matching  $\mathcal{M}^*$  in  $G$ . This matching covers exactly  $\mu(G)$  vertices of  $A$  and exactly  $\mu(G)$  vertices of  $B$ . Similarly, the matching  $\mathcal{M}$  covers exactly  $|\mathcal{M}|$  vertices in both  $A$  and  $B$ . Therefore, there are at least  $\mu(G) - |\mathcal{M}|$  vertices in  $A$  that are covered by  $\mathcal{M}^*$  but are not covered by  $\mathcal{M}$ . Since  $\mathcal{M}$  is maximal, the edges in  $\mathcal{M}^*$  that cover these vertices are incident to the edges in  $\mathcal{M}$ . Therefore, these edges are between  $B(\mathcal{M})$  and  $\overline{A(\mathcal{M})}$ . Thus, the size of a maximum matching between  $B(\mathcal{M})$  and  $\overline{A(\mathcal{M})}$  is at least

$$\mu(G) - |\mathcal{M}| = (1/2 - \delta)\mu(G).$$

Using a similar argument it can also be shown that the size of maximum matching between  $A(\mathcal{M})$  and  $\overline{B(\mathcal{M})}$  is also at least  $(1/2 - \delta)\mu(G)$ .  $\square$

**Claim 3.5.** *Let  $\mathcal{M}'$  be the set of edges in  $\mathcal{M}_1^*$  that are in  $G'$ . Then,  $|\mathcal{M}'|$  is at least  $|P_1| - \mu(G) \cdot 2\delta\alpha$ .*

**Proof.** We know that there exists a matching of size  $(1/2 - \delta)|\mathcal{M}_1^*|$  in  $G_L$ . Therefore the number of vertices in  $B(\mathcal{M}_P)$  that are not matched to a vertex in  $A(G_L)$  is bounded by

$$\begin{aligned} |\mathcal{M}_0| - (1/2 - \delta)|\mathcal{M}_1^*| &= (1/2 + \delta)|\mathcal{M}_1^*| - (1/2 - \delta)|\mathcal{M}_1^*| \\ &= 2\delta|\mathcal{M}_1^*| = \mu(G) \cdot 2\delta\alpha. \end{aligned}$$

Therefore, there are at most  $\mu(G) \cdot 2\delta\alpha$  vertices in  $B(\mathcal{M}_P)$  that cannot be matched which proves the claim.  $\square$