

The Open University of Israel
Department of Mathematics and and Computer Science

Maximum Matching sans Maximal Matching: A New Approach for Finding Maximum Matchings in the Data Stream Model

Thesis submitted as partial fulfillment of the requirements
towards an M.Sc. degree in Computer Science
The Open University of Israel
Department of Mathematics and Computer Science

By
Ariel Szarf

Prepared under the supervision of
Prof. Moran Feldman and Prof. Zeev Nutov

October 2021

Table of Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 1.1 | Related Work | 5 |
| 2 | Preliminaries | 6 |
| 3 | Two-Pass Non-MMF Algorithm | 7 |
| 4 | Three-Pass Algorithm for Triangle-Free Graphs | 13 |
| 5 | Three-Pass Algorithm for General Graphs | 18 |
| 6 | Conclusion and Future Work | 22 |
| A | Three-Pass Non-MMF Algorithm | 24 |
| B | Two-Pass Algorithm for Triangle-Free Graphs | 29 |

List of Tables and Figures

| | | |
|----------|--|----|
| Table 1 | Known and new approximation ratios using two or three passes | 4 |
| Figure 1 | Types of components and M^* edges that can incident them | 11 |

List of Algorithms

| | | |
|---|--|----|
| 1 | MAXIMUM MATCHING VIA GREEDY TRIANGLES - TWO PASSES | 8 |
| 2 | TRIANGLEFREEALG – FIRST PASS | 13 |
| 3 | TRIANGLEFREEALG – SECOND PASS | 14 |
| 4 | TRIANGLEFREEALG – THIRD PASS | 15 |
| 5 | MAXIMUM MATCHING VIA AUGMENTING PATHS – GENERAL GRAPHS | 19 |
| 6 | Construction of W'_2 | 20 |
| 7 | Construction of \mathcal{P}' | 21 |
| 8 | MAXIMUM MATCHING VIA GREEDY TRIANGLES - 3 PASSES | 25 |
| 9 | TRIANGLEFREEALG – FIRST AND SECOND PASSES | 30 |

Abstract

The problem of finding a maximum size matching in a graph (known as the *maximum matching* problem) is one of the most classical problems in computer science. Despite a significant body of work dedicated to the study of this problem in the data stream model, the state-of-the-art single-pass semi-streaming algorithm for it is still a simple greedy algorithm that computes a maximal matching, and this way obtains $1/2$ -approximation. Some previous works described two/three-pass algorithms that improve over this approximation ratio by using their second and third passes to improve the above mentioned maximal matching. One contribution of this thesis continues this line of work by presenting new three-pass semi-streaming algorithms that work along these lines and obtain improved approximation ratios of 0.6111 and 0.5694 for triangle-free and general graphs, respectively.

Unfortunately, a recent work [20] shows that the strategy of constructing a maximal matching in the first pass and then improving it in further passes has limitations. Additionally, this technique is unlikely to get us closer to single-pass semi-streaming algorithms obtaining a better than $1/2$ -approximation. Therefore, it is interesting to come up with algorithms that do something else with their first pass (we term such algorithms non-maximal-matching-first algorithms). No such algorithms are currently known (to the best of our knowledge), and the main contribution of this thesis is describing such algorithms that obtain approximation ratios of 0.5384 and 0.5555 in two and three passes, respectively, for general graphs (the result for three passes improves over the previous state-of-the-art, but is worse than the result of this thesis mentioned in the previous paragraph for general graphs). The improvements obtained by these results are, unfortunately, numerically not very impressive, but the main importance (in our opinion) of these results is in demonstrating the potential of non-maximal-matching-first algorithms.

Keywords: Maximum matching, semi-streaming algorithms, multi-pass algorithms

1 Introduction

The problem of finding a maximum size matching in a graph (known as the *maximum matching* problem) is one of the most classical problems in computer science, and many polynomial time algorithms have been designed for it over the years (see, *e.g.*, [4, 8, 14]). Due to its central role, the maximum matching problem is often one of the first problems considered when new computational models are suggested. One such model is the data stream model, which is motivated by Big-Data applications, and has been the subject of an enormous amount of research over the last couple of decades.

In the data stream model, the algorithm receives the input in the form of a stream which it can read sequentially, but due to memory restrictions, the algorithm can store only a small part of this stream. This means that the algorithm has to process (in some sense) the input stream while reading it, and never gets an opportunity to see all the parts of the input at the same time. Traditional algorithms for this model, known as *streaming algorithms*, are allowed only memory that is poly-logarithmic in the natural parameters of the problem. Obtaining a streaming algorithm for a problem is very desirable, but is often not possible. In particular, many graph problems provably do not admit streaming algorithms, and the maximum matching problem is among these problems if one would like an algorithm for the problem to output an (approximately) maximum matching because such a matching might be of linear size in the number of vertices. Nevertheless, non-trivial streaming algorithms have been designed for the maximum matching problem when only the (approximate) size of a maximum matching is desired (see Section 1.1 for details).

The resistance of many graph problems to streaming algorithms has motivated Feigenbaum et al. [11] to suggest semi-streaming algorithms, which are algorithms for the data stream model that are allowed a space complexity of $O(n \log^c n)$ for some constant $c \geq 0$, where n is the number of vertices in the graph. Such algorithms turn out to be a sweet-spot that on the one hand allows many results of interest, and on the other hand, does not lead to triviality because $O(n \log^c n)$ is less than the space necessary for storing the input graph (unless this graph is very sparse). In particular, Feigenbaum et al. [11] observed that one can obtain $1/2$ -approximation for the maximum matching problem using a simple semi-streaming algorithm that greedily constructs a maximal matching.¹

The above $1/2$ -approximation semi-streaming algorithm for the maximum matching problem also has the desirable property that it reads the input stream only once (*i.e.*, it makes a single pass over it). Surprisingly, no single-pass semi-streaming algorithm improving over the approximation ratio of this simple algorithm was suggested in the decade and a half that has already passed since the work of [11] (in contrast, Kapralov [17] showed that no such algorithm can have an approximation ratio better than $1/(1 + \ln 2) \approx 0.59$, improving over previous inapproximability results due to [13, 16]). Given this lack of progress, interest arose in obtaining improved approximation ratios for relaxed versions of the above problem. Perhaps, the simplest such relaxation is to allow the algorithm to make multiple (sequential) passes over the input stream. Some works tried to understand the approximation ratio that can be obtained as the number of passes grows (but remains constant)—see Section 1.1 for more detail. Another line of work is interested in studying semi-streaming algorithms with very few passes (usually two or three).

The state-of-the-art results for the last line of work are summarized in Table 1. We note that beside the state-of-the-art results for general input graphs, Table 1 also gives improved results for bipartite and triangle-free graphs. All the known results in this line of work (to the best of our knowledge) start by greedily constructing a maximal matching during the first pass over the input stream, and then augmenting this matching in the subsequent passes. Recently, Konrad

¹A maximal matching is a matching that is inclusion-wise maximal, and it is well-known that the size of any maximal matching is a $1/2$ -approximation for the size of a maximum matching.

Table 1: The state-of-the-art approximation ratios for semi-streaming algorithms using two or three passes, and our improvements over these ratios (the number to the right of each improvement is the number of the theorem formally stating it).

| Number of Passes | Type of Graphs | State-of-the-Art | This Thesis |
|------------------|----------------|--|---|
| Two-Pass | Bipartite | $2 - \sqrt{2} \approx \frac{1}{2} + \frac{1}{11.66} \approx 0.5857$ [19] | - |
| | Triangle-Free | $\frac{1}{2} + \frac{1}{16} = 0.5625$ [15] | - |
| | General | $\frac{1}{2} + \frac{1}{32} = 0.53125$ [15] | $\frac{1}{2} + \frac{1}{26} \approx 0.5385$ (1.1) |
| Three-Pass | Bipartite | $0.6067 \approx \frac{1}{2} + \frac{1}{9.37}$ [19] | $\frac{1}{2} + \frac{1}{9} \approx 0.6111$ (1.3) |
| | Triangle-Free | $\frac{1}{2} + \frac{1}{10} = 0.6$ [15] | $\frac{1}{2} + \frac{1}{9} \approx 0.6111$ (1.3) |
| | General | $\frac{1}{2} + \frac{81}{1600} \approx \frac{1}{2} + \frac{1}{19.753} \approx 0.5506$ [15] | $\frac{1}{2} + \frac{1}{14.4} \approx 0.5694$ (1.4) |

and Naidu [20] showed that this technique has limitations (specifically, even for bipartite graphs, a two-pass semi-streaming algorithm based on this technique cannot obtain a better than $2/3$ -approximation). Additionally, and arguably more importantly, multi-pass algorithms that use their first pass for constructing a maximal matching are unlikely to be a step towards a single-pass semi-streaming algorithm with a better than $1/2$ -approximation guarantee.

Given the above observations, it is natural to believe that the future of the study of semi-streaming algorithms for the maximum matching problem lies in algorithms that use their first pass in a more sophisticated way than simply constructing the traditional maximal matching. We term such algorithms *non-maximal-matching-first algorithms* (or non-MMF algorithms for short). In this thesis, we present the first non-MMF algorithms, which leads to improvements over the state-of-the-art both for two and three passes. Admittedly, the improvements we obtain are numerically not very impressive, but their main importance (in our opinion) is in demonstrating the potential of non-MMF algorithms.

To intuitively understand our non-MMF algorithms, one should note that greedily constructing a maximal matching is equivalent to greedily constructing a graph whose connected components are of size at most 2 (where the size of a connected component is defined as the number of vertices in it). Therefore, a natural generalization is to greedily construct in the first pass a graph whose connected components are of size at most 3. There are two intuitive advantages for doing that compared to constructing a maximal matching.

- If many connected components end up to be of size 2 rather than 3, then it is not possible for many of the edges of a maximum matching to intersect only a single connected component of the constructed graph; and therefore, the constructed graph must have many connected components compared to the size of a maximum matching.
- A connected component of size 3 can contribute two edges to the output matching if it is “augmented” during in the next passes with a single additional edge. In contrast, doing the same with a connected component of size 2 requires “augmenting” it with two additional edges. It is important to note that there is a significant conceptual difference between an augmentation of a connected component with one or two edges. Augmenting a connected component with two edges requires finding pairs of edges that augment the *same* connected component, while augmenting with a single edge does not require such a synchronization.

Using the above ideas, we prove in Section 3 and Appendix A the following two theorems, respectively.

Theorem 1.1. *There exists a non-MMF 2-pass ($7/13 = 1/2 + 1/26$)-approximation semi-streaming algorithm for finding a maximum size matching in a general graph.*

Theorem 1.2. *There exists a non-MMF 3-pass ($5/9 = 1/2 + 1/18$)-approximation semi-streaming algorithm for finding a maximum size matching in a general graph.*

As mentioned above, both Theorems 1.1 and 1.2 represent an improvement over the state-of-the-art. However, it turns out that we can further improve over Theorem 1.2 using new MMF algorithms (*i.e.*, algorithms that construct a maximal matching in their first pass). This leads to the following theorems whose proofs appear in Sections 4 and 5, respectively.

Theorem 1.3. *There exists a 3-pass ($11/18 = 1/2 + 1/9$)-approximation semi-streaming algorithm for finding a maximum size matching in a triangle-free graph.²*

Theorem 1.4. *There exists a 3-pass ($1/2 + 1/14.4$)-approximation semi-streaming algorithm for finding a maximum size matching in a general graph.*

The algorithms used to prove Theorems 1.3 and 1.4 are strongly based on the algorithms suggested by Kale and Tirodkar [15]. For example, the first two passes of the algorithm suggested by Theorem 1.3 are identical to a two-pass algorithm presented by [15], and the third pass of this algorithm is very similar to the third pass of the three-pass algorithm of [15]. Our novelty, however, is in our ability to analyze the algorithm obtained by putting these two components together.

In Section B, we analyze the approximation ratio produced by the first and second passes of our algorithm for triangle-free graph (without the third pass). This does not lead to an improvement over the state-of-the-art approximation ratio for two passes, however, our analysis is stronger than the analysis of Kale and Tirodkar in [15] for the same passes. Specifically, Kale and Tirodkar showed that these passes guarantee $(1/2 + 1/20)$ -approximation in triangle-free graphs, and we show that the same algorithm in fact guarantees $(1/2 + 1/18)$ -approximation for triangle-free graphs and $(1/2 + 1/14)$ -approximation for bipartite graphs.

1.1 Related Work

As mentioned in Section 1, streaming algorithms are not appropriate for the maximum matching problem when the algorithm is required to output an (approximately) maximum matching. However, some non-trivial streaming algorithms are known for this problem when the algorithm is only required to estimate the size of the maximum matching. Kapralov et al. [18] designed a poly-log approximation streaming algorithm for this problem under the assumption that the edges in the input stream are ordered in a uniformly random order. A different line of work [6, 10, 22] considered graphs of bounded arboricity α , cumulating with the work of McGregor and Vorotnikova [23], who designed $(\alpha + 2)(1 + \varepsilon)$ -approximation streaming algorithm for this problem requiring only $O(\varepsilon^{-2} \log n)$ space.

Recall that, to date, the best single-pass semi-streaming algorithm for the maximum matching problem is still the natural greedy algorithm, which guarantees $1/2$ -approximation. Chitnis et al. [5] presented an exact single-pass algorithm for this problem. However, this algorithm requires

²We recall that every bipartite graph is triangle-free, and therefore, the same result is obtained also for bipartite graphs.

$\tilde{O}(k^2)$ memory, where k is an upper bound on the size of the maximum matching (which the algorithm needs to know upfront), and thus, this algorithm is a semi-streaming algorithm only when $k = \tilde{O}(\sqrt{n})$. Given the difficulty to improve over the guarantee of the greedy algorithm using single-pass semi-streaming algorithms, people started to consider relaxed versions of the maximum matching problem. One standard relaxation is to allow the algorithm to make multiple passes over the input stream. Section 1 surveys algorithms of this kind that use two or three passes. Another line of work considers algorithms that assume a constant (but possibly large) number of passes. The first result of this kind was presented by Feigenbaum et al. [11] (in the same paper that also introduced the notion of semi-streaming algorithms), and guaranteed $(2/3 - \varepsilon)$ -approximation using $O(\varepsilon^{-1} \log \varepsilon^{-1})$ passes for bipartite graphs. Later [21] showed how to obtain $(1 - \varepsilon)$ -approximation for general graphs using $(\varepsilon^{-1})^{O(\varepsilon^{-1})}$ passes, and the number of passes necessary to obtain this guarantee was improved by many further works (see, *e.g.*, [1, 3, 12]). Another standard relaxation for the maximum matching problem is to assume that the edges of the input stream appear in a uniformly random order. The state-of-the-art for this relaxation is a $(2/3 + \varepsilon_0)$ -approximation single-pass semi-streaming algorithm, where $\varepsilon_0 > 0$ is some absolute constant [2] (see also the references therein for previous works on this relaxation).

The related maximum weight matching problem was also studied heavily in the context of the data stream model. Here, it is not immediately clear that one can obtain a constant approximation ratio using a single-pass semi-streaming algorithm. However, Feigenbaum et al. [11] presented the first such algorithm guaranteeing $1/6$ -approximation, and this ratio was improved in series of works [7, 9, 21, 25]. The current state-of-the-art for the problem is $(1/2 - \varepsilon)$ -approximation due to Paz and Schwartzman [24]. Since this approximation ratio is essentially identical to the state-of-the-art for the (unweighted) maximum matching problem, any further progress on the maximum weight matching problem will imply an improvement over the guarantee of the greedy algorithm for the (unweighted) maximum matching problem.

2 Preliminaries

In this section we present the problem that we study more formally, and also introduce the notation used throughout the rest of the thesis. We are interested in semi-streaming algorithms for the problem of finding a maximum size matching in a graph $G = (V, E)$ of n vertices. A semi-streaming algorithm for this problem is an algorithm with a space complexity of $O(n \log^c n)$ (for some constant $c \geq 0$) that initially has no knowledge about the edges of E . Instead, the edges of E appear sequentially in an “input stream”, and the algorithm may make one or more passes over this input stream. In each pass the algorithm sees the edges one by one, and may do arbitrary calculations after viewing each edge. It is important to note that the space complexity allowed for the algorithm does not suffice for storing all the edges of the graph (unless the graph is very sparse), and this is the reason that the algorithm might benefit from doing multiple passes over the input stream. It is standard to assume that the vertices of V are known upfront, and that each vertex of V can be stored using $O(\log n)$ bits (which implies that every edge of E can also be stored using this asymptotic number of bits).

Throughout the thesis, we consider only unweighted graphs and matchings. We also denote by M^* an arbitrary maximum matching of G (*i.e.*, an arbitrary optimal solution for our problem). Notation-wise, we treat M^* (and any other matching considered in the thesis) as a set of the edges included in it. Similarly, when considering a connected component C of a graph, we treat it as a set of the vertices in it, which in particular, implies that $|C|$ is the number of such vertices.

Given a set of edges S or a path P in a graph, we denote by $V(S)$ and $V(P)$ the set of vertices

intersecting any edge of S or P , respectively. Similarly, the set of edges included in the path P is denoted by $E(P)$. Often we need to consider collections of paths (or triangles) in a given graph. For clarity, such collections are always denoted using calligraphic letters, and we extend the above notation to such collections. In other words, if \mathcal{P} is a collection of paths, then $V(\mathcal{P})$ and $E(\mathcal{P})$ is the set of vertices and edges, respectively, that are included in these paths. Finally, given a set S of edges and a vertex v , we use $\deg_S(v)$ to denote the degree of the vertex v in the subgraph (V, S) .

3 Two-Pass Non-MMF Algorithm

In this section we prove Theorem 1.1, which we repeat below for convenience.

Theorem 1.1. *There exists a non-MMF 2-pass $(7/13 = 1/2 + 1/26)$ -approximation semi-streaming algorithm for finding a maximum size matching in a general graph.*

The algorithm whose existence is guaranteed by Theorem 1.1 appears as Algorithm 1. In its first pass, this algorithm greedily grows a set P of edges that form either triangles or partial triangles (*i.e.*, isolated edges or paths of length 2). For simplicity, we refer below to the connected components of (V, P) that are not isolated vertices as partial triangles although, technically, they can also be full triangles. In the second pass of Algorithm 1, the algorithm tries to convert the partial triangles of P into more involved structures in one of two ways. To understand these ways, we need to define some terms. First, we designate some of the vertices of every partial triangle as “connection vertices”. Specifically, all the vertices of a triangle are considered connection vertices; in a path of length 2 only the two end points are considered to be connection vertices; and finally, in an isolated edge there are no connection vertices. We refer to a partial triangle that was not converted yet into a more involved structure as a “naïve” partial triangle. The first way in which Algorithm 1 tries to convert the partial triangles of P into more involved structures is by greedily adding edges that connect a connection vertex of a naïve partial triangle with an isolated vertex. The set A_1 in the algorithm includes the edges that were added in this way. In parallel, the algorithm also tries a second way to convert the partial triangles of P into more involved structures, which is to greedily add edges that connect a connection vertex of a naïve partial triangle either to a connection vertex of another naïve partial triangle or to an isolated vertex. The set A_2 in the algorithm includes the edges that were added in this way. Upon termination, Algorithm 1 outputs a maximum matching in the set of all the edges that it kept. We recall that given a connected component C of a graph, the notation $|C|$ represents the number of vertices in C .

We begin the analysis of Algorithm 1 by showing that it is indeed a semi-streaming algorithm.

Observation 3.1. *Algorithm 1 is a semi-streaming algorithm.*

Proof. Since every connected component of the graph (V, P) is of size at most 3, the set P contains at most n edges. Furthermore, each connected component of (V, P) intersects at most a single edge of the set A_1 and at most a single edge of the set A_2 , and therefore, each one of these sets can include at most $n/2$ edges. Hence, in total, Algorithm 1 keeps only $O(n)$ edges. \square

In the rest of this section we analyze the approximation ratio of Algorithm 1. Recall that we use M^* to denote some maximum matching of G . Our first objective in the analysis of the approximation ratio of Algorithm 1 is to lower bound the number of edges of M^* that can potentially be added either to A_1 or to A_2 . Towards this goal, we define a charging scheme π . Under the charging scheme π , every edge $(u, v) \in M^*$ charges the connected components of u and v in (V, P) . Each one of these connected components is charged one unit by (u, v) , unless it is an isolated edge

Algorithm 1: MAXIMUM MATCHING VIA GREEDY TRIANGLES - TWO PASSES

```
// First Pass
1 Let  $P \leftarrow \emptyset$ .
2 for every edge  $e$  that arrives do
3   if every connected component of the graph  $(V, P \cup \{e\})$  is either a path of length at
4     most 2 or a triangle (cycle of size 3) then
5        $\lfloor$  Add  $e$  to  $P$ .

// Second Pass
6 Let  $A_1 \leftarrow \emptyset$  and  $A_2 \leftarrow \emptyset$ .
7 for every edge  $(u, v) \notin P$  that arrives do
8   Let  $C_u$  and  $C_v$  be the connected components of  $u$  and  $v$ , respectively, in  $(V, P)$ . We
9     assume without loss of generality that  $|C_u| > 1$ , otherwise we swap the roles of  $u$ 
10    and  $v$ . // Note that we cannot have  $|C_u| = |C_v| = 1$  because the edge  $(u, v)$ 
11    was not added to  $P$  in the first pass.
12  if no edge of  $A_1$  intersects  $C_u$  and  $C_v$ ,  $|C_v| = 1$  and  $u$  is a connection vertex of  $C_u$  then
13     $\lfloor$  Add the edge  $(u, v)$  to  $A_1$ .
14  if no edge of  $A_2$  intersects  $C_u$  and  $C_v$ ,  $|C_v| = 1$  and  $u$  is a connection vertex of  $C_u$  then
15     $\lfloor$  Add the edge  $(u, v)$  to  $A_2$ .
16  else if no edge of  $A_2$  intersects  $C_u$  and  $C_v$ , and  $u$  and  $v$  are connection vertices of  $C_u$ 
17    and  $C_v$ , respectively then
18     $\lfloor$  Add the edge  $(u, v)$  to  $A_2$ .

19 return a maximum matching in the graph  $(V, P \cup A_1 \cup A_2)$ .
```

or an isolated vertex, in which case it is charged only half a unit or nothing by (u, v) , respectively. We note that when u and v belong to the same connected component of (V, P) , then this connected component is charged twice by (u, v) .³

The following observation provides an upper bound on the total charged by all the edges of M^* together. Let $(\#single)$ be the number of isolated edges in P , $(\#double)$ be the number of connected components in (V, P) that are paths of length 2 and $(\#triangle)$ be the number of triangles in P .

Observation 3.2. *The total charge according to π is at most $(\#single)+3(\#double)+3(\#triangle)$.*

Proof. Every positive amount charged by π is charged to some connected component of (V, P) which is not an isolated vertex. Therefore, to prove the observation we only need to show that every isolated edge of (V, P) is charged at most one unit, and every connected component of (V, P) that is either a path of length 2 or a triangle is charged at most 3 units. Below we argue that this is indeed the case.

Each connected component C of (V, P) can be charged at most once for every one of its vertices since the fact that M^* is a matching implies that every vertex of C can appear in at most a single edge of M^* . For isolated edges of (V, P) , this implies that they can be charged at most twice, and

³Intuitively, the charge assigned to the connected components of u and v is proportional to the “blame” that can be assigned to them if (u, v) ends up to be outside P . For example, an isolated edge could not alone prevent (u, v) from being added to P , but two such edges (one intersecting u and the other intersecting v) could, together, prevent (u, v) from being added to P . Therefore, we assign a charge of $1/2$ to isolated edges. Observation 3.3 is based on this intuition.

therefore, they are charged at most one unit because they are charged half a unit in each charge. Similarly, connected components of (V, P) that are either paths of length 2 or triangles contain 3 vertices, and therefore, can be charged at most three times. Since every one of these charges is of a single unit, the total charge to each connected component of these kinds is at most 3. \square

To complement the last observation, let us now describe a simple lower bound on the total charging done by all the edges of M^* according to π . Let $(\# \text{component-free})$ be the number of edges of M^* that connect a connection vertex of a connected component of (V, P) to an isolated vertex of (V, P) , $(\# \text{component-component})$ be the number of edges of M^* that connect connection vertices of two different connected components of (V, P) , $(\# \text{single-single})$ be the number of edges of M^* whose two end points belong to (not necessary distinct) isolated edges of (V, P) , $(\# \text{single-component})$ be the number of edges of M^* that connect a vertex of an isolated edge of (V, P) with a connection vertex of some (other) connected component of (V, P) and $(\# \text{middle})$ be the number of edges that either intersect the middle vertex of a length 2 path connected component of (V, P) or are included within a triangle connected component of (V, P) .

Observation 3.3. *The total charge of all the edges of M^* according to the charging scheme π is at least $(\# \text{component-free}) + 2(\# \text{component-component}) + (\# \text{single-single}) + 1.5(\# \text{single-component}) + (\# \text{middle})$.*

Proof. Since the edges of M^* counted by $(\# \text{component-free})$ intersect a connection vertex, they must intersect a connected component of (V, P) which is not an isolated vertex or an isolated edge, and therefore, they charge this connected component one unit. Hence, the total charge by all the edges counted by $(\# \text{component-free})$ is at least $(\# \text{component-free})$. Similar logic shows that the total charge by all the edges counted by $(\# \text{component-component})$, $(\# \text{single-single})$, $(\# \text{single-component})$ and $(\# \text{middle})$ are at least $2(\# \text{component-component})$, $(\# \text{single-single})$, $1.5(\# \text{single-component})$ and $(\# \text{middle})$, respectively. The observation now follows since the edges of M^* counted by $(\# \text{component-free})$, $(\# \text{component-component})$, $(\# \text{single-single})$, $(\# \text{middle})$ and $(\# \text{single-component})$ are distinct. \square

Combining Observations 3.2 and 3.3, we get the following inequality.

$$\begin{aligned} (\# \text{component-free}) + 2(\# \text{component-component}) + (\# \text{single-single}) & \quad (1) \\ + 1.5(\# \text{single-component}) + (\# \text{middle}) & \leq (\# \text{single}) + 3(\# \text{double}) + 3(\# \text{triangle}) . \end{aligned}$$

In its current form, Inequality (1) is not very useful. We later derive from it a more convenient inequality, but before doing this we need to prove a few other inequalities. Let $(\# \text{non-}M^*\text{-triangles})$ denote the number of triangle connected components of (V, P) that do not include any edge of M^* within them.

Lemma 3.4. *The following inequalities hold*

$$\begin{aligned} (\# \text{component-free}) + (\# \text{component-component}) + (\# \text{single-single}) & \quad (2) \\ + (\# \text{middle}) + (\# \text{single-component}) & \geq |M^*| , \end{aligned}$$

$$(\# \text{double}) + (\# \text{triangle}) - (\# \text{non-}M^*\text{-triangles}) \geq (\# \text{middle}) , \quad (3)$$

$$(\# \text{single-component}) \leq 2(\# \text{single-single}) + (\# \text{single-component}) \leq 2(\# \text{single}) , \quad (4)$$

and they imply together

$$\begin{aligned} (\# \text{component-free}) + (\# \text{component-component}) + 2(\# \text{single}) & \\ + (\# \text{double}) + (\# \text{triangle}) - (\# \text{non-}M^*\text{-triangles}) & \geq |M^*| . \end{aligned}$$

Proof. Since every edge that is included in a connected component of (V, P) which is a path of length 2 must include the middle vertex of this path, every edge $e \in M^*$ that is not counted by either $(\#component-free)$, $(\#component-component)$, $(\#single-single)$, $(\#single-component)$ or $(\#middle)$ must either connect a vertex of an isolated edge of (V, P) to an isolated vertex or connect two isolated vertices of (V, P) . However, such edges cannot exist. Specifically, assume towards a contradiction that (u, v) is an edge of M^* such that u is an isolated vertex of (V, P) and v is either another isolated vertex of (V, P) or belongs to an isolated edge of this graph. Then, the edge (u, v) should have been added by Algorithm 1 to P upon arrival, which contradicts the fact that its end point u ended up as an isolated vertex of (V, P) . Hence, every edge $e \in M^*$ is counted by either $(\#component-free)$, $(\#component-component)$, $(\#single-single)$, $(\#single-component)$ or $(\#middle)$, which implies Inequality (2).

Recall that every edge counted by $(\#middle)$ must either be included in a triangle connected component of (V, P) or intersect the middle vertex of a path of length 2 connected component of (V, P) . Since M^* is a matching, only one edge of M^* can intersect the middle vertex of a given length 2 path or be included in a given triangle, and therefore, every edge counted by $(\#middle)$ can be associated with a distinct path of length 2 or triangle component of (V, P) that is not counted by $(\#non-M^*-triangles)$, which implies Inequality (3).

Every edge counted by $(\#single-single)$ touches two end-points of isolated edges of (V, P) . Similarly, every edge counted by $(\#single-component)$ intersects an end-point of an isolated edge of (V, P) . Since every end-point of an isolated edge of (V, P) can be touched by at most a single edge of M^* because M^* is a matching, this implies that the number of end points of the isolated edges of (V, P) is at least $2(\#single-single) + (\#single-component)$. However, this number is also equal to $2(\#single)$, which implies Inequality (4). \square

The last inequality in the previous lemma provides a lower bound on $(\#component-free) + (\#component-component)$, and one can view $(\#component-free) + (\#component-component)$ as a count of edges of M^* that have potential to be added to A_2 in Algorithm 1. The next lemma is the promised derivative of Inequality (1), and it provides a lower bound on $(\#component-free)$. Observe that $(\#component-free)$ is a count of edges of M^* that have the potential to be added to A_1 .

Lemma 3.5. $2|M^*| \leq (\#component-free) - (\#non-M^*-triangles) + 2(\#single) + 4(\#double) + 4(\#triangle)$.

Proof. Adding twice Inequality (2) to Inequality (1), we get

$$\begin{aligned} & 2|M^*| - (\#component-free) - (\#single-single) - 0.5(\#single-component) - (\#middle) \\ & \leq (\#single) + 3(\#double) + 3(\#triangle) . \end{aligned}$$

The lemma now follows by adding Inequality (3) and half of Inequality (4) to the last inequality. \square

So far we have shown lower bounds on the size of the sets of edges that have a potential to be added to A_1 or A_2 by Algorithm 1. Our next step is to lower bound the size of the sets A_1 and A_2 that Algorithm 1 ends up constructing using this potential.

Lemma 3.6. $3|A_1| \geq (\#component-free) - (\#non-M^*-triangles)$.

Proof. We say that an edge e of M^* counted by $(\#component-free)$ is excluded by an edge $f \in A_1$ if e and f intersect the same connected component of (V, P) . One can observe that every edge e counted by $(\#component-free)$ is excluded by some edge of A_1 (possibly itself) when Algorithm 1

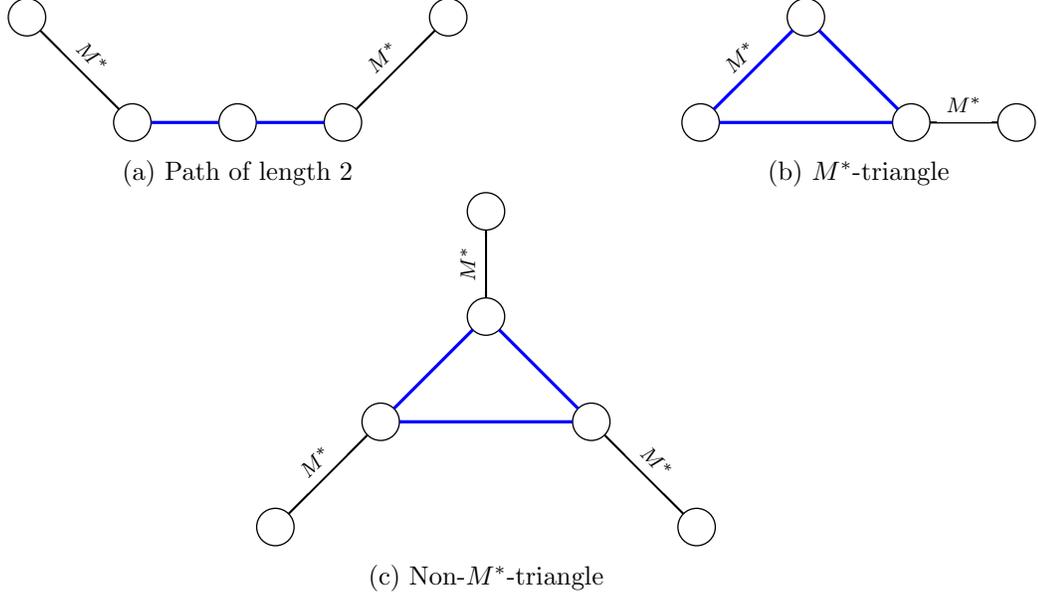


Figure 1: A graphical study of the maximum number of M^* edges counted by $(\# \text{component-free})$ that can intersect connection vertices of various types of partial triangles. Sub-figures (a) and (b) show that at most two such edges can intersect the connection vertices of a path of length 2 and an M^* -triangle (*i.e.*, a triangle that includes an edge of M^*). Sub-figure (c) shows that the connection vertices of a non- M^* -triangle can intersect up to 3 edges of M^* .

terminates because otherwise Algorithm 1 would have added e to A_1 , which would have resulted in e excluding itself. Therefore, we can upper bound $(\# \text{component-free})$ by counting the number of edges excluded by the edges of A_1 .

Let (u, v) be an edge of A_1 , and assume without loss of generality that v is the end point of this edge which is an isolated vertex of (V, P) . This implies that u is a connection vertex of a connected component C_u of (V, P) which is either a path of length 2 or a triangle. If C_u is a path of length 2, then the edge (u, v) can exclude only edges counted by $(\# \text{component-free})$ that intersect either v or a connection vertex of C_u , and there can be only 3 such edges because M^* is a matching (see Figure 1a). Next, consider the case in which C_u is a triangle which is not counted by $(\# \text{non-}M^*\text{-triangles})$. In this case there can be at most 2 edges of OPT intersecting C_u (see Figure 1b), and therefore, even though (u, v) can exclude any edge of M^* intersecting C_u or v , there can be only 3 such edges. It remains to consider the case in which C_u is a triangle counted by $(\# \text{non-}M^*\text{-triangles})$. In this case, (u, v) can again exclude every edge of M^* that intersects C_u or v , and this time there can be at most 4 such edges (see Figure 1c). Combining all the above, we get that the number of edges excluded by all the edges of A_1 is at most

$$3|A_1| + |\{e \in A_1 \mid e \text{ intersects a triangle counted by } (\# \text{non-}M^*\text{-triangles})\}| .$$

As explained above, this expression is an upper bound on $(\# \text{component-free})$. Furthermore, since A_1 includes at most a single edge intersecting every connected component of (V, P) , the second term in this expression is upper bounded by $(\# \text{non-}M^*\text{-triangles})$. Therefore, we get

$$(\# \text{component-free}) \leq 3|A_1| + (\# \text{non-}M^*\text{-triangles}) .$$

The lemma now follows by rearranging this inequality. □

The next corollary now follows by combining Lemmata 3.5 and 3.6.

Corollary 3.7. $2|M^*| \leq 3|A_1| + 2(\#single) + 4(\#double) + 4(\#triangle)$.

Lemma 3.8. $4|A_2| \geq (\#component-component) + (\#component-free) - (\#non-M^*-triangles)$.

Proof. The proof of Lemma 3.8 is very similar to the proof of Lemma 3.6, and therefore, we only sketch it. We first define that an edge $e \in A_2$ excludes an edge f of M^* counted by either $(\#component-component)$ or $(\#component-free)$ if they both intersect the same connected component of (V, P) . Like in the proof of Lemma 3.6, it can be argued that $(\#component-component) + (\#component-free)$ is upper bounded by the total number of edges of M^* excluded by the edges of A_2 , and on the other hand, every edge e of A_2 excludes up to $4 + T(e)$ edges, where $T(e)$ is the number of triangles counted by $(\#non-M^*-triangles)$ that intersect e . Therefore,

$$(\#component-component) + (\#component-free) \leq \sum_{e \in A_2} [4 + T(e)] \leq 4|A_2| + (\#non-M^*-triangles) ,$$

where the second inequality holds since every connected component of (V, P) intersects only a single edge of A_2 . The lemma now follows by rearranging the last inequality. \square

The next corollary follows by combining Lemma 3.8 and the final inequality in Lemma 3.4.

Corollary 3.9. $|M^*| \leq 4|A_2| + 2(\#single) + (\#double) + (\#triangle)$.

Let us now denote $L = (\#single) + (\#double) + (\#triangle) + \max\{|A_1|, |A_2|\}$. We argue below that L is a lower bound on the size of the solution produced by Algorithm 1. However, before proving this, let us show first that L is large.

Lemma 3.10. $L \geq 7/13|M^*|$.

Proof. Plugging the definition of L into Corollaries 3.7 and 3.9 yields the inequalities

$$2|M^*| \leq 3L - (\#single) + (\#double) + (\#triangle)$$

and

$$|M^*| \leq 4L - 2(\#single) - 3(\#double) - 3(\#triangle) .$$

Adding the first of these inequalities three times to the second one gives

$$7|M^*| \leq 13L - 5(\#single) \leq 13L ,$$

where the second inequality holds since $(\#single)$ is non-negative by definition. The lemma now follows by rearranging the above inequality. \square

As promised, we now argue that the size of the matching produced by Algorithm 1 is at least L .

Lemma 3.11. *Algorithm 1 outputs a matching of size at least L .*

Proof. Since Algorithm 1 outputs a maximum matching in $(V, P \cup A_1 \cup A_2)$, to prove the lemma it suffices to show that the graph $(V, P \cup A_1)$ includes a matching of size $(\#single) + (\#double) + (\#triangle) + |A_1|$ and the graph $(V, P \cup A_2)$ includes a matching of size $(\#single) + (\#double) + (\#triangle) + |A_2|$. We prove below only the claim regarding $(V, P \cup A_2)$. The claim regarding $(V, P \cup A_1)$ can be proved analogously.

Let H be the number of edges in A_2 that connect two non-isolated vertices of (V, P) . Then, we classify the connected components of $(V, P \cup A_2)$ as follows, and show how to build a large matching M based on this classification.

- $(V, P \cup A_2)$ includes $(\#single) + (\#double) + (\#triangle) - |A_2| - H$ connected components that are (i) not an isolated node, and (ii) appear also in (V, P) . Each one of these connected components contains at least one edge, and therefore, can contribute some edge to M .
- $(V, P \cup A_2)$ includes $|A_2| - H$ connected components that consist of a connected component C of (V, P) that has connection vertices and an edge e connecting a connection vertex of C to an isolated vertex of (V, P) . One can observe that the combination of C and e must be either a path of length 3 or a triangle and an edge attached to one of its vertices, and in both cases this combined connected component contains two vertex disjoint edges which it can contribute to the matching M .
- $(V, P \cup A_2)$ includes H connected components that consist of two connected components C_1, C_2 of (V, P) that have connection vertices and an edge e connecting a connecting vertex of C_1 with a connecting vertex of C_2 . There are three shapes that the connected component obtained in this way can take: a path of length 5, a triangle with a path of length 3 attached to one of its vertices or two triangles and an edge connecting them. However, one can observe that all these shapes include three vertex disjoint edges that can be contributed to the matching M .

By collecting from every connected component of $(V, P \cup A_2)$ the edges that it can contribute to M according to the above analysis, we get a matching in $(V, P \cup A_2)$ of size at least

$$\begin{aligned} & [(\#single) + (\#double) + (\#triangle) - |A_2| - H] + 2[|A_2| - H] + 3H \\ & = (\#single) + (\#double) + (\#triangle) + |A_2| . \end{aligned} \quad \square$$

Lemmata 3.10 and 3.11 imply together the following corollary. Together with Observation 3.1, this corollary implies Theorem 1.1.

Corollary 3.12. *Algorithm 1 is a $7/13$ -approximation algorithm.*

4 Three-Pass Algorithm for Triangle-Free Graphs

In this section we prove Theorem 1.3, which we repeat here for convenience.

Theorem 1.3. *There exists a 3-pass ($11/18 = 1/2 + 1/9$)-approximation semi-streaming algorithm for finding a maximum size matching in a triangle-free graph.*

We refer to the algorithm whose existence is guaranteed by Theorem 1.3 as `TRIANGLEFREEALG`. In its first pass, `TRIANGLEFREEALG` constructs a maximal matching M_0 of G . Formally, the pseudocode for this pass appears as Algorithm 2.

Algorithm 2: `TRIANGLEFREEALG` – FIRST PASS

- 1 Let $M_0 \leftarrow \emptyset$.
 - 2 **for** every edge e that arrives **do**
 - 3 \lfloor Add e to M_0 if it does not intersect any edge that already belongs to M_0 .
-

We say that an edge $e \in E$ is a *wing* if e includes exactly one vertex of $V(M_0)$. Intuitively, the reason we are interested in wings is that one can obtain an augmenting path⁴ for M_0 by combining

⁴A path P is an augmenting path for a matching M if $M \oplus E(P)$ is a valid matching of size $|M| + 1$.

an edge $(u, v) \in M_0$ with two wings: one wing that intersects u and one wing that intersects v . The second pass of TRIANGLEFREEALG grows a set W of wings. Since we hope to construct multiple augmenting paths using these wings, the algorithm makes sure to limit the number of wings in W that intersect any given vertex u (specifically, the algorithm allows only a single wing in W to intersect u if $u \in V(M_0)$, and otherwise it allows up to two wings of W to intersect u). The pseudocode of this second pass appears as Algorithm 3.

Algorithm 3 also includes a post-processing step in which a set \mathcal{P}_1 of augmenting paths (with respect to M_0) is constructed using W . This is done by constructing an auxiliary multi-graph G_A over the vertices of $V \setminus V(M_0)$ in which there is an edge between two nodes $u, v \in V \setminus V(M_0)$ for every path $P_{u,v}$ of length 3 in $W \cup M_0$ between them. One can note that every such path $P_{u,v}$ must be an augmenting path consisting of an edge $e \in M_0$ and two wings from W : one intersecting u and an end-point of e , and the other intersecting v and the other end-point of e . Algorithm 3 finds a maximum size matching M_A in G_A , and then sets \mathcal{P}_1 to be the collection of (augmenting) paths corresponding to the edges of M_A .

Algorithm 3: TRIANGLEFREEALG – SECOND PASS

```

1 Let  $W \leftarrow \emptyset$ .
2 for every edge  $e$  that arrives do
3   if  $e$  intersects exactly one vertex  $u \in V(M_0)$  then
4     Let  $v$  denote the other end-point of  $e$  (i.e., the end-point that is not  $u$ ).
5     if  $\deg_W(u) < 1$  and  $\deg_W(v) < 2$  then
6       Add  $e$  to  $W$ .

// Post-processing
7 Let  $G_A$  be a multi-graph over the vertices  $V \setminus V(M_0)$ . For every path  $P_{u,v}$  of length 3 in
    $W \cup M_0$  between two vertices  $u, v \in V \setminus V(M_0)$ , we add an edge  $(u, v)$  to the graph  $G_A$ .
   // This is a multi-graph because there might be multiple such paths
   between a pair of vertices of  $V \setminus V(M_0)$ .
8 Find a maximum size matching  $M_A$  in  $G_A$ .
9 Let  $\mathcal{P}_1 \leftarrow \{P_{u,v} \mid (u, v) \in M_A\}$ .
```

Consider now an edge $e \in M_0$ that does not appear in any path of \mathcal{P}_1 and is connected by some wing $w \in W$ to some vertex $u \notin V(M_0) \cup V(\mathcal{P}_1)$. The pair e, w can be extended into an augmenting path if one can find another wing w' connecting the other end of e (the end that does not intersect w) to a vertex $v \notin V(M_0) \cup V(\mathcal{P}_1)$ that is not u . The third pass of TRIANGLEFREEALG greedily constructs a collection \mathcal{P}_2 of augmenting paths in this way. A pseudocode of this pass appears as Algorithm 4. After completing the pass, Algorithm 4 returns the matching obtained by augmenting M_0 with the augmenting paths of \mathcal{P}_1 and \mathcal{P}_2 .

We begin the analysis of TRIANGLEFREEALG with the following lemma, which shows that this algorithm returns a matching, and also gives a basic lower bound on the size of this matching.

Lemma 4.1. *The paths in \mathcal{P}_1 and \mathcal{P}_2 are vertex disjoint, and therefore, the output of TRIANGLEFREEALG is a matching of size $|M_0| + |\mathcal{P}_1| + |\mathcal{P}_2|$.*

Proof. Given the above discussion, it is clear that all the paths in $\mathcal{P}_1 \cup \mathcal{P}_2$ are augmentation paths with respect to M_0 , which implies that the first part of the lemma indeed implies the second part. Furthermore, one can observe that the condition in Line 3 of Algorithm 4 guarantees that the paths

Algorithm 4: TRIANGLEFREEALG – THIRD PASS

```
1 Let  $\mathcal{P}_2 \leftarrow \emptyset$ .
2 for every edge  $w'$  that arrives do
3   if there exist 4 vertices  $u, a, b, v \in V \setminus (V(\mathcal{P}_1) \cup V(\mathcal{P}_2))$  such that: (i)  $u \notin V(M_0)$ , (ii)
    $w' = (u, a)$ , (iii)  $(a, b) \in M_0$  and (iv)  $(b, v) \in W$  then
4     ┌ Add the path  $u, a, b, v$  to  $\mathcal{P}_2$ . // Note that  $u \neq v$  because otherwise  $u, a, b, v$ 
     └ would have been a triangle.
5 return  $M_0 \oplus (\bigcup_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} E(P))$ .
```

in \mathcal{P}_2 are vertex disjoint from each other and from the paths of \mathcal{P}_1 . Thus, to complete the proof of the lemma, it remains to argue that the paths in \mathcal{P}_1 are also vertex disjoint.

Recall that the end-points of every path in \mathcal{P}_1 belong to $V \setminus V(M_0)$ and the internal points of these paths belong to $V(M_0)$. Therefore, to show that the paths in \mathcal{P}_1 are vertex disjoint, it suffices to argue this separately for their end-points and their internal nodes. Every path $P_{u,v} \in \mathcal{P}_1$ corresponds to an edge (u, v) in the matching M_A . Since the end-points of the path $P_{u,v}$ are also the end-points of this edge, we get that the paths in \mathcal{P}_1 must have disjoint end-points because M_A is a matching. Consider now some path $P_{u,v} \in \mathcal{P}_1$, and let us denote the internal nodes of this path by a and b . Since a and b appear only in the edge (a, b) of M_0 (because M_0 is a matching), we get that if one of them belongs to a path of \mathcal{P}_1 , then the other belongs to this path as well. Furthermore, by Line 8 of Algorithm 3, $\deg_W(a) = \deg_W(b) = 1$, which implies that any path of \mathcal{P}_1 that includes the nodes a and b as internal nodes must in fact be identical to $P_{u,v}$ itself. Hence, no two paths in \mathcal{P}_1 share internal nodes. \square

Using the last lemma we can also bound the space complexity of Algorithm 4.

Corollary 4.2. TRIANGLEFREEALG is a semi-streaming algorithm.

Proof. Aside from a constant number of other vertices and edges, TRIANGLEFREEALG has to store only the edges of $M_0 \cup W$ and the paths of $\mathcal{P}_1 \cup \mathcal{P}_2$. As these paths are of constant length (specifically, a length of 3), to prove the corollary we only need to argue that M_0 , W , \mathcal{P}_1 and \mathcal{P}_2 are all of size $O(n)$. Below we argue that this is indeed the case.

- M_0 is a matching in the graph G , and therefore, its size is at most $n/2$.
- Every edge of W is a wing, and thus, has one end point in M_0 . Since Line 8 of Algorithm 3 guarantees that $\deg_W(u) \leq 1$ for every vertex $u \in M_0$, this implies $|W| \leq 2|M_0| \leq n$.
- Since the paths in $\mathcal{P}_1 \cup \mathcal{P}_2$ are vertex disjoint by Lemma 4.1, and each path contains 4 vertices, the number of paths in both sets together cannot exceed $n/4$. \square

It remains to analyze the approximation ratio of TRIANGLEFREEALG. Our analysis roughly follows the flow of the algorithm, and thus, we begin by observing that the matching M_0 constructed in the first pass of this algorithm is of size at least $|M^*|/2$ (recall that M^* is a maximum size matching of G) because M_0 is a maximal matching of G by construction.

In its second pass, TRIANGLEFREEALG constructs the set W of wings. Our next objective is to lower bound the size of W . Towards this goal, we need to define W_M to be the set of all edges of M^* that are wings (we recall that an edge e is a wing if exactly one of its end points appear in $V(M_0)$).

Observation 4.3. $|W_M| \geq 2(|M^*| - |M_0|)$.

Proof. Since M_0 is a maximal matching, every edge of M^* intersects at least one edge of M_0 . Hence, every edge of W_M includes a single end-point of an edge of M_0 , and every edge of $M^* \setminus W_M$ includes two end-points of edges of M_0 (the two end-points might belong to different edges or to the same edge), which implies $|M_0| \geq (|W_M| + 2|M^* \setminus W_M|)/2 = |M^*| - |W_M|/2$. Rearranging this inequality completes the proof of the observation. \square

Lemma 4.4. $|W| \geq \frac{2}{3}|W_M| \geq \frac{4}{3}(|M^*| - |M_0|)$.

Proof. Let $I = V(M_0) \cap V(W_M)$, and let I_F be the set of vertices in I that do not appear in any edge of W . Every vertex $a \in I_F \subseteq I$ must belong to some wing $w(a) \in W_M$ by the definition of I . However, this wing was not added to W (because $a \in I_F$), which implies that the condition in Line 8 of Algorithm 3 evaluated to FALSE when $w(a)$ arrived. Since a is not covered by any edge of W (i.e., $\deg_W(v) = 0$), the fact that this condition evaluated to FALSE implies that the end point of $w(a)$ that does not belong to $V(M_0)$ must have a degree of 2 under W . Formally, if we denote by $u(a)$ the end point of $w(a)$ that does not belong to $V(M_0)$, then we must have $\deg_W(u(a)) = 2$.

We now observe that (i) every wing in W_M contains a disjoint vertex of $V \setminus V(M_0)$ because W_M is a subset of the optimal matching M^* , and (ii) every wing in W contains only one vertex of $V \setminus V(M)$ because it is a wing. These two observations imply together

$$|W| \geq \sum_{a \in I_F} \deg_W(u(a)) = 2|I_F| . \quad (5)$$

In contrast, since (i) every wing in W contains a single vertex of $V(M_0)$, and (ii) all the vertices of $I \setminus I_F \subseteq V(M_0)$ appear in some wing of W ,

$$|W| \geq |I| - |I_F| = |W_M| - |I_F| , \quad (6)$$

where the equality holds since every edge of W_M is a wing, and therefore, intersects a single vertex of $V(M_0)$. The lemma now follows by adding two copies of Inequality (6) to Inequality (5). \square

We now get to the analysis of the third pass of TRIANGLEFREEALG, and our first goal in this analysis is to identify a set of paths that have a potential (in some sense) to end up in \mathcal{P}_2 . Let \mathcal{P}' be the set of paths of length 3 in G that consist of a wing of W_M followed by an edge of M_0 and then a wing of W . We think of the paths in \mathcal{P}' as directed from their W_M to their W edge, and consider two paths that differ only in their direction to be different paths. This is important because if there is an edge $e \in M_0$ incident to two edges $w_1, w_2 \in W \cap W_M$, then the path w_1, e, w_2 fulfills the requirements to belong to \mathcal{P}' both when w_1 is considered the first edge in it and when w_2 is considered the first edge of the path. Thus, the fact that we treat the direction of the path as part of the path's definition allows both the paths w_1, e, w_2 and w_2, e, w_1 to appear in \mathcal{P}' .

Observation 4.5. $|\mathcal{P}'| \geq \frac{10}{3}|M^*| - \frac{16}{3}|M_0|$.

Proof. Since $\deg_W(a) \leq 1$ for every vertex $a \in V(M_0)$, there are $|W|$ end-points of M_0 that intersect an edge of W . Let us denote these end-points by V_W , and for every end-point $a \in V_W$, we denote by $b(a)$ the other end-point of the same edge of M_0 . Formally, $V_W = V(M_0) \cap V(W)$, and $b(a)$ is the single element of the set $\{b \mid (a, b) \in M_0\}$. One can now observe that \mathcal{P}' includes a (distinct) path for every wing of W_M that intersect $b(a)$ for some vertex $a \in V_W$. Therefore,

$$\begin{aligned} |\mathcal{P}'| &= |\{b(a) \mid a \in V_W\} \cap V(W_M)| \\ &\geq |\{b(a) \mid a \in V_W\}| + |V(W_M) \cap V(M_0)| - |V(M_0)| = |W| + |W_M| - |V(M_0)| \\ &\geq \frac{4}{3}(|M^*| - |M_0|) + 2(|M^*| - |M_0|) - |V(M_0)| = \frac{10}{3}|M^*| - \frac{16}{3}|M_0| , \end{aligned}$$

where the first equality holds since $\{b(a) \mid a \in V_W\}$ is a subset of $V(M_0)$, and the last inequality follows from Observation 4.3 and Lemma 4.4. \square

A path in \mathcal{P}' has a potential to be added to \mathcal{P}_2 only if none of its vertices appears in \mathcal{P}_1 . Let \mathcal{P}'' be the set of such paths (formally, $\mathcal{P}'' = \{P \in \mathcal{P}' \mid V(P) \cap V(\mathcal{P}_1) = \emptyset\}$). The following lemma lower bounds the size of \mathcal{P}'' .

Lemma 4.6. $|\mathcal{P}''| \geq |\mathcal{P}'| - 6|\mathcal{P}_1| \geq \frac{10}{3}|M^*| - \frac{16}{3}|M_0| - 6|\mathcal{P}_1|$.

Proof. The second inequality of the lemma follows from Observation 4.5, and therefore, we concentrate on proving the first inequality. Towards this goal, assume that $P' \in \mathcal{P}'$ is a path that intersects with a path $P_1 \in \mathcal{P}_1$ on an internal vertex. Since the middle edge of both paths is an edge of M_0 , this implies that the two paths intersect on both their internal vertices. Furthermore, since both end-edges of P_1 and one end-edge of P' belong to W , there must be an internal vertex $a \in V(M_0)$ of both paths that intersects an edge of W in both paths. However, since $\deg_W(a) \leq 1$, the edges of W intersecting a in both paths must be identical, which implies that the paths P' and P_1 intersect also on some end-point. Since P' and P_1 were chosen as general paths of \mathcal{P}' and \mathcal{P}_1 , respectively, that intersect on an internal node, this implies that the difference $|\mathcal{P}'| - |\mathcal{P}''|$ is equal to the number of paths in \mathcal{P}' that intersect a path of \mathcal{P}_1 in an *end-point*. The rest of the proof is devoted to proving that the last number is at most $6|\mathcal{P}_1|$.

Since each path of \mathcal{P}_1 has only two end points, to prove that the paths of \mathcal{P}_1 intersect at most $6|\mathcal{P}_1|$ paths of \mathcal{P}' at an end-point, it suffices to show that every vertex of $V \setminus V(M_0)$ can appear in at most 3 paths of \mathcal{P}' . To see why that is the case, consider an arbitrary vertex $u \in V \setminus V(M_0)$. If u belongs to some path $P' \in \mathcal{P}'$, then it must be in one of two roles as follows.

- If u is the last vertex of the path, then the last edge of the path is an edge $e \in W$ that includes u , and the other edges of the path P' are the single edge of M_0 intersecting e and the single edge of W_M intersecting e . Note that this means that the identity of the entire path is determined by the edge e , and therefore, the number of paths of \mathcal{P}' in which u is the last vertex can be upper bounded by $\deg_W(u) \leq 2$.
- If u is the first vertex of the path, then the first edge of the path is the single edge $e \in W_M$ that includes u , and the other edges of the path are the single edge $e' \in M_0$ that intersect e and the single edge $e'' \in W$ that intersects e' . Hence, the entire path is determined by the fact that u is its first vertex, and therefore, there can be only a single path in \mathcal{P}' in which u is the first vertex. \square

Originally, all the paths of \mathcal{P}'' can be picked in the third pass of TRIANGLEFREEALG (Algorithm 4) since they are vertex disjoint from the paths of \mathcal{P}_1 . However, as Algorithm 4 starts to add paths to \mathcal{P}_2 , it stops being possible to add some paths of \mathcal{P}'' to \mathcal{P}_2 . Still, we can lower bound the size of \mathcal{P}_2 in terms of the size of \mathcal{P}'' .

Lemma 4.7. $|\mathcal{P}_2| \geq \frac{1}{6}|\mathcal{P}''| \geq \frac{5}{9}|M^*| - \frac{8}{9}|M_0| - |\mathcal{P}_1|$.

Proof. We begin the proof by observing that no edge $e \in M_0$ is connect by two distinct wings $w_1, w_2 \in W$ to vertices of $V \setminus (V(M_0) \cup V(\mathcal{P}_1))$. Assume towards a contradiction that this is not true, then there is an edge e in G_A corresponds to the path P defined as w_1, e, w_2 . Since M_A is a maximum matching in G_A , it must include at least one edge that contains some end-point of P (otherwise, the edge corresponding to P could be added to M_A , which violates its maximality); which contradicts the definition of either w_1 or w_2 .

For every path $P'' \in \mathcal{P}''$, let us charge a cost of 1 to some path of \mathcal{P}_2 that intersects it. To see why such a path must exist, let us denote by e_M the edge of P'' that belongs to W_M (the first edge of P''). When e_M arrives, the path P'' was one candidate to be added to \mathcal{P}_2 by Algorithm 4. If this candidate was still feasible at this time (in the sense that it was vertex disjoint from \mathcal{P}_2), then Algorithm 4 must have added either P'' to \mathcal{P}_2 or another path that includes e_M . In either case, following the arrival of e_M , some path intersecting P'' (which is possibly P'' itself) appears in \mathcal{P}_2 —and can be charged.

Our next goal is to show that the total cost charged to any single path of \mathcal{P}_2 is at most 6, which implies the lemma because the total cost charged to all the paths of \mathcal{P}_2 is exactly $|\mathcal{P}''|$. We do that by making two observations.

- Since $\mathcal{P}'' \subseteq \mathcal{P}'$, we get by the proof of Lemma 4.6 that at most 3 paths of \mathcal{P}'' can include any given vertex $u \in V \setminus V(M_0)$.
- Our second observation is that, if a path $P'' \in \mathcal{P}''$ intersects a path $P_2 \in \mathcal{P}_2$, then they must intersect on an end-point of P_2 . Assume towards a contradiction that they only intersect on an internal node a . Since the middle edges of both paths are edges of M_0 that include a , both internal edges must be the same. Let us denote this internal edge by e . Furthermore, as explained above, there can be only a single edge $w \in W$ that intersects e and does not include a vertex of $V(\mathcal{P}_1)$. This edge must belong also to both paths, and therefore, the end-point of w that does not belong to $V(M_0)$ is an end-point of both P'' and P_2 .

Combining the above two observations, we get that, for every path $P_2 \in \mathcal{P}_2$, only paths of \mathcal{P}'' intersecting an end-point of P_2 can charge a cost to P_2 , and there can be at most 3 paths of \mathcal{P}'' intersecting each such end-point. Since P_2 has only two end-points, this implies that at most 6 paths of \mathcal{P}'' can charge P_2 . \square

Corollary 4.8. *The size of the output of TRIANGLEFREEALG is $|M_0| + |\mathcal{P}_1| + |\mathcal{P}_2| \geq \frac{11}{18}|M^*| = (\frac{1}{2} + \frac{1}{9})|M^*|$.*

Proof. The size of the output of TRIANGLEFREEALG is $|M_0| + |\mathcal{P}_1| + |\mathcal{P}_2|$ by Lemma 4.1, thus, we only need to lower bound this sum. To do this, note that

$$\begin{aligned} |M_0| + |\mathcal{P}_1| + |\mathcal{P}_2| &\geq |M_0| + |\mathcal{P}_1| + \left\{ \frac{5}{9}|M^*| - \frac{8}{9}|M_0| - |\mathcal{P}_1| \right\} \\ &= \frac{5}{9}|M^*| + \frac{1}{9}|M_0| \geq \frac{5}{9}|M^*| + \frac{1}{18}|M^*| = \frac{11}{18}|M^*| , \end{aligned}$$

where the first inequality follows from Lemma 4.7, and the second inequality follows from the observation made at the beginning of this section (namely, that $|M_0|$ is a $1/2$ -approximation for $|M^*|$ because M_0 is a maximal matching). \square

Theorem 1.3 now follows from Corollaries 4.2 and 4.8.

5 Three-Pass Algorithm for General Graphs

In this section we prove Theorem 1.4, which we repeat here for convenience.

Theorem 1.4. *There exists a 3-pass $(1/2 + 1/14.4)$ -approximation semi-streaming algorithm for finding a maximum size matching in a general graph.*

The algorithm that we use to prove Theorem 1.4 is given as Algorithm 5. Since this algorithm is very similar to the algorithm TRIANGLEFREEALG presented in Section 4, we use below the terminology and notation defined in the last section.

Intuitively, the reason why TRIANGLEFREEALG does not apply to general graphs is that given an edge $(a, b) \in M_0$, a wing $(u, a) \in W_M$ and a wing $(b, v) \in W$, we are not guaranteed that these three edges form an augmenting path for the matching M_0 because they might represent a triangle. To overcome this hurdle, Algorithm 5 constructs two sets of edges in its second pass: a set W_1 constructed exactly like the set W in TRIANGLEFREEALG, and a set W_2 constructed in the same way, but while excluding the edges of W_1 . Since W_1 and W_2 are disjoint, given an edge $(a, b) \in M_0$ and a wing $(u, a) \in W_M$, at most one of the sets W_1 or W_2 can contain a wing that forms a triangle together with these two edges, which intuitively allows us to bound the deterioration in the approximation guarantee resulting from the existence of such triangles.

Algorithm 5: MAXIMUM MATCHING VIA AUGMENTING PATHS – GENERAL GRAPHS

```

// First Pass
1 Let  $M_0 \leftarrow \emptyset$ .
2 for every edge  $e$  that arrives do
3   | Add  $e$  to  $M_0$  if it does not intersect any edge that already belongs to  $M_0$ .

// Second Pass
4 Let  $W_1 \leftarrow \emptyset$ ,  $W_2 \leftarrow \emptyset$ .
5 for every edge  $e$  that arrives do
6   | if  $e$  intersects exactly one vertex  $u \in V(M_0)$  then
7     | Let  $v$  denote the other end-point of  $e$  (i.e., the end-point that is not  $u$ ).
8     | if  $\deg_{W_1}(u) < 1$  and  $\deg_{W_1}(v) < 2$  then
9     |   | Add  $e$  to  $W_1$ .
10    | else if  $\deg_{W_2}(u) < 1$  and  $\deg_{W_2}(v) < 2$  then
11    |   | Add  $e$  to  $W_2$ .

// Post-processing
12 Let  $G_A$  be a multi-graph over the vertices  $V \setminus V(M_0)$ . For every path  $P_{u,v}$  of length 3 in
     $W_1 \cup W_2 \cup M_0$  between two distinct vertices  $u, v \in V \setminus V(M_0)$ , we add an edge  $(u, v)$  to
    the graph  $G_A$ . // This is a multi-graph because there might be multiple
    such paths between a pair of vertices of  $V \setminus V(M_0)$ .
13 Find a maximum size matching  $M_A$  in  $G_A$ .
14 Let  $\mathcal{P}_1 \leftarrow \{P_{u,v} \mid (u, v) \in M_A\}$ .

// Third Pass
15 Let  $\mathcal{P}_2 \leftarrow \emptyset$ .
16 for every edge  $w'$  that arrives do
17   | if there exist 4 vertices  $u, a, b, v \in V \setminus (V(\mathcal{P}_1) \cup V(\mathcal{P}_2))$  such that: (i)  $u \notin V(M_0)$ , (ii)
     $w' = (u, a)$ , (iii)  $(a, b) \in M_0$ , (iv)  $(b, v) \in W_1 \cup W_2$  and (v)  $u \neq v$  then
18   |   | Add the path  $u, a, b, v$  to  $\mathcal{P}_2$ .

19 return  $M_0 \oplus (\bigcup_{P \in \mathcal{P}_1 \cup \mathcal{P}_2} E(P))$ .
```

We note that the analysis of TRIANGLEFREEALG up to Lemma 4.4 applies to Algorithm 5 with two differences.

- The proof of Corollary 4.2 upper bounds by n the size of the set W (recall that this set is identical to the set W_1 in Algorithm 5). To make this proof apply to Algorithm 5 as well, we need to observe that the size of the set W_2 is at most n due to the same argument. In particular, this implies that Algorithm 5 is a semi-streaming algorithm.
- Lemma 4.4 provides a lower bound on the size of the set W , which translates into an identical lower bound on the size of the corresponding set W_1 in Algorithm 5.

In the rest of this section, it will be convenient to work with the set W'_2 constructed by Algorithm 6 (note that Algorithm 6 is used for analysis purposes only). Intuitively, W'_2 is constructed in the same general way in which W_1 and W_2 are constructed; however, while all the edges of the input stream are considered in the construction of W_1 , and only the edges of $E \setminus W_1$ are considered in the construction of W_2 , the construction of W'_2 takes into account the edges of $(E \setminus W_1) \cup W_M$.

Algorithm 6: Construction of W'_2

```

1 Let  $W'_2 \leftarrow W_2$ .
2 for every edge  $(u, v) \in W_1 \cap W_M$  do
3   Assume without loss of generality that  $u$  is the end point of  $(u, v)$  that belongs to
    $V(M_0)$ .
4   if  $\deg_{W'_2}(u) < 1$  and  $\deg_{W'_2}(v) < 2$  then
5     Add  $(u, v)$  to  $W'_2$ .
```

Since W'_2 is a subset of $W_1 \cup W_2$ by construction, the set $W_1 \cup W_2$ that is often referred to by Algorithm 5 is identical to the set $W_1 \cup W'_2$. Furthermore, one can observe that the lower bound proved by Lemma 4.4 for W_1 applies also to W'_2 because all the edges of W_M are considered for addition to W'_2 at some point (either during the construction of W_2 or in Algorithm 6). This implies the following observation.

Observation 5.1. $|W_1| + |W'_2| \geq \frac{4}{3}|W_M|$.

We now define a multi-set \mathcal{P}' similar to the set of the same name used in the analysis of TRIANGLEFREEALG. Specifically, \mathcal{P}' includes every triangle or path obtained by combining an edge $(u, a) \in W_M$, an edge $(a, b) \in M_0$ and an edge (b, v) of either W_1 or W'_2 . Moreover, if there are multiple options to obtain a path or triangle in this way, then the multiplicity of the path or triangle in \mathcal{P}' will be equal to the number of these options. To make this point clearer, we provide a pseudocode for constructing \mathcal{P}' as Algorithm 7 (again, Algorithm 7 is used for analysis purposes only).

Observation 5.2. $|\mathcal{P}'| \geq \frac{20}{3}|M^*| - \frac{32}{3}|M_0|$.

Proof. Repeating the proof of Observation 4.5, we get that at least $|W_1| + |W_M| - |V(M_0)|$ paths are added to \mathcal{P}' in Line 5 of Algorithm 7, and at least $|W'_2| + |W_M| - |V(M_0)|$ paths are added to \mathcal{P}' in Line 7 of Algorithm 7. Therefore,

$$\begin{aligned}
|\mathcal{P}'| &\geq |W_1| + |W'_2| + 2|W_M| - 2|V(M_0)| \geq \left(\frac{4}{3} + 2\right)|W_M| - 2|V(M_0)| \\
&\geq 2\left(\frac{4}{3} + 2\right)(|M^*| - |M_0|) - 2|V(M_0)| = \frac{20}{3}|M^*| - \frac{32}{3}|M_0|,
\end{aligned}$$

where the second inequality follows from Observation 5.1, and the last inequality follows from Observation 4.3. \square

Algorithm 7: Construction of \mathcal{P}'

```

1 Let  $\mathcal{P}' \leftarrow \emptyset$ .
2 for every edge  $(u, a) \in W_M$  do
3   for every edge  $(a, b) \in M_0$  do
4     for every edge  $(b, v) \in W_1$  do
5        $\lfloor$  Add the path/triangle  $(u, a), (a, b), (b, v)$  to  $\mathcal{P}'$ .
6     for every edge  $(b, v) \in W'_2$  do
7        $\lfloor$  Add the path/triangle  $(u, a), (a, b), (b, v)$  to  $\mathcal{P}'$ .

```

An element (path or triangle) of \mathcal{P}' has a potential to be added to \mathcal{P}_2 by Algorithm 5 only if it is a path (*i.e.*, not a triangle) and none of its vertices appears in \mathcal{P}_1 . Let \mathcal{P}'' be the multi-set of such paths. The following lemma lower bounds the size of \mathcal{P}'' .

Lemma 5.3. $|\mathcal{P}''| \geq |\mathcal{P}'| - 12|\mathcal{P}_1| - |M_0| \geq \frac{20}{3}|M^*| - \frac{35}{3}|M_0| - 12|\mathcal{P}_1|$.

Proof. The second inequality of the lemma follows from Observation 5.2, and therefore, we concentrate on proving the first inequality. Let $\tilde{\mathcal{P}}'$ be the multi-set of paths/triangles from \mathcal{P}' that do not intersect any vertex of \mathcal{P}_1 . Repeating the proof of Lemma 4.6, we get that $\tilde{\mathcal{P}}'$ contains all the paths/triangles added to \mathcal{P}' by Line 5 of Algorithm 7 except for up to $6|\mathcal{P}_1|$ paths/triangles, and the same is true for the paths/triangles added to \mathcal{P}' by Line 7 of Algorithm 7. Since every path/triangle in \mathcal{P}' was added to this multi-set by either Line 5 or Line 7 of Algorithm 7, we get

$$|\tilde{\mathcal{P}}'| \geq |\mathcal{P}'| - 12|\mathcal{P}_1| .$$

Since \mathcal{P}'' includes every path of $\tilde{\mathcal{P}}'$, to complete the proof of the lemma it remains to show that $\tilde{\mathcal{P}}'$ contains at most $|M_0|$ triangles. To see that this is indeed the case, we recall that every triangle (or path) in $\tilde{\mathcal{P}}'$ must include a single edge of M_0 , and we claim that no two triangles in $\tilde{\mathcal{P}}'$ can share this edge (and therefore, the number of triangles is upper bounded by the number of edges in M_0). Assume towards a contradiction that this claim does not hold, *i.e.*, that there exist two triangles $T_1, T_2 \in \tilde{\mathcal{P}}'$ sharing an edge $e \in M_0$. Each one of these triangles must include one edge of W_M . Let e_1 and e_2 denote the edges of W_M in T_1 and T_2 , respectively, and let e'_1 the single edge of T_1 which is not e or e_1 and e'_2 be the single edge of T_2 which is not either e or e_2 . We now need to consider two cases. The first case is when $e_1 = e_2$. In this case e'_1 and e'_2 must be also identical, and cannot belong to W_M because $e_1 = e_2$ belongs to W_M and W_M is a subset of the matching M^* . However, this leads to a contradiction because one of the edges e'_1 or e'_2 must belong to W_1 , and the other of these edges must belong to W'_2 , and the sets W_1 and W'_2 can intersect only on edges of W_M .

It remains to consider the case in which $e_1 \neq e_2$. Let u_1, u_2 be the end-points of these edges, respectively, that do not belong to the edge e of M_0 . Since $e_1 \neq e_2$ are edges of the W_M , which is a subset of the matching M^* , u_1 and u_2 must be distinct. Consider now the path e'_1, e, e'_2 . One can observe that this is indeed a path because (i) $u_1 \neq u_2$ and (ii) the fact that e_1 and e_2 are vertex disjoint implies that e'_1 and e'_2 intersect different end-points of e . Furthermore, since $T_1, T_2 \in \tilde{\mathcal{P}}'$, this path does not intersect any vertex of \mathcal{P}_1 , and thus, its existence contradicts the maximality of the matching M_A constructed by Algorithm 5 because both e'_1 and e'_2 belong to $W_1 \cup W'_2 = W_1 \cup W_2$. \square

We are now ready to lower bound the number of augmenting paths found by Algorithm 5 during its third pass.

Lemma 5.4. $|P_2| \geq |\mathcal{P}''|/12 \geq \frac{5}{9}|M^*| - \frac{35}{36}|M_0| - |\mathcal{P}_1|$.

Proof. The proof of the lemma is very similar to the proof of Lemma 4.7, except that now every path of \mathcal{P}_2 might get a charge of up to 12 because the paths of \mathcal{P}'' originally added to \mathcal{P}' by Line 5 of Algorithm 7 can contribute up to 6 to this charge, and the same goes for the paths of \mathcal{P}'' originally added to \mathcal{P}' by Line 7 of this algorithm. \square

Theorem 1.4 now follows from Corollary 4.2 and the next corollary.

Corollary 5.5. *The size of the matching produced by Algorithm 5 is at least $(\frac{1}{2} + \frac{1}{14.4})|M^*|$.*

Proof. By Lemma 4.1, the size of the matching produced by Algorithm 5 is at least

$$|M_0| + |\mathcal{P}_1| + |\mathcal{P}_2| \geq \frac{5}{9}|M^*| + \frac{1}{36}|M_0| \geq \frac{5}{9}|M^*| + \frac{1}{72}|M^*| = (\frac{1}{2} + \frac{1}{14.4})|M^*| ,$$

where the first inequality holds by Lemma 5.4, and the second inequality holds since M_0 (as a maximal matching) is of size at least $\frac{1}{2}|M^*|$. \square

6 Conclusion and Future Work

We presented in this thesis a new approach to semi-streaming algorithms for the maximum matching problem, and showed that this approach can be used to improve the state-of-the-art in two and three passes for general graphs. In this approach, we do more sophisticated logic in the first pass rather than simply building a maximal matching in a greed fashion, as is done by previous algorithms. Specifically, we greedily build in this pass connected components of size 3 (note that greedily building a maximal matching is equivalent to greedily building connected components of size 2). In addition to the above, we have used the traditional technique to improve over the state-of-the-art for triangle-free graphs and general graphs in 3 passes (for general graphs we get a more significant improvement in this way compared to the improvement obtained using our new approach).

The main open problems related to this work are to improve (hopefully significantly) the approximation ratio for two-pass and three-pass algorithms (or to prove that this is not possible), and more importantly, to break the state-of-the-art for single-pass (which is still the natural greedy algorithm). We believe that our new approach can be an important step in both of these directions.

References

- [1] Kook Jin Ahn and Sudipto Guha. Access to data and number of iterations: Dual primal algorithms for maximum matching under resource constraints. *ACM Trans. Parallel Comput.*, 4(4):17:1–17:40, 2018.
- [2] Sepehr Assadi and Soheil Behnezhad. Beating two-thirds for random-order streaming matching. In *48th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 19:1–19:13, 2021.
- [3] Sepehr Assadi, S. Cliff Liu, and Robert E. Tarjan. An auction algorithm for bipartite matching in streaming and massively parallel computation models. In *4th Symposium on Simplicity in Algorithms (SOSA)*, pages 165–171, 2021.
- [4] Michel L. Balinski and Jaime Gonzalez. Maximum matchings in bipartite graphs via strong spanning trees. *Networks*, 21(2):165–179, 1991.

- [5] Rajesh Chitnis, Graham Cormode, Hossein Esfandiari, MohammadTaghi Hajiaghayi, Andrew McGregor, Morteza Monemizadeh, and Sofya Vorotnikova. Kernelization via sampling with applications to finding matchings and related problems in dynamic graph streams. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1326–1344. SIAM, 2016.
- [6] Graham Cormode, Hossein Jowhari, Morteza Monemizadeh, and S. Muthukrishnan. The sparse awakens: Streaming algorithms for matching size estimation in sparse graphs. In *25th Annual European Symposium on Algorithms (ESA)*, pages 29:1–29:15, 2017.
- [7] Michael S. Crouch and Daniel M. Stubbs. Improved streaming algorithms for weighted matching, via unweighted matching. In Klaus Jansen, José D. P. Rolim, Nikhil R. Devanur, and Christopher Moore, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, volume 28 of *LIPICs*, pages 96–104. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2014.
- [8] Jack Edmonds. Maximum matching and a polyhedron with 0, 1-vertices. *Journal of research of the National Bureau of Standards B*, 69(125-130):55–56, 1965.
- [9] Leah Epstein, Asaf Levin, Julián Mestre, and Danny Segev. Improved approximation guarantees for weighted matching in the semi-streaming model. *SIAM J. Discret. Math.*, 25(3):1251–1265, 2011.
- [10] Hossein Esfandiari, MohammadTaghi Hajiaghayi, Vahid Liaghat, Morteza Monemizadeh, and Krzysztof Onak. Streaming algorithms for estimating the matching size in planar graphs and beyond. *ACM Trans. Algorithms*, 14(4):48:1–48:23, 2018.
- [11] Joan Feigenbaum, Sampath Kannan, Andrew McGregor, Siddharth Suri, and Jian Zhang. On graph problems in a semi-streaming model. In Josep Díaz, Juhani Karhumäki, Arto Lepistö, and Donald Sannella, editors, *Automata, Languages and Programming: 31st International Colloquium, (ICALP)*, volume 3142 of *Lecture Notes in Computer Science*, pages 531–543. Springer, 2004.
- [12] Manuela Fischer, Slobodan Mitrovic, and Jara Uitto. Deterministic $(1+\epsilon)$ -approximate maximum matching with $\text{poly}(1/\epsilon)$ passes in the semi-streaming model. *CoRR*, abs/2106.04179, 2021.
- [13] Ashish Goel, Michael Kapralov, and Sanjeev Khanna. On the communication and streaming complexity of maximum bipartite matching. In Yuval Rabani, editor, *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 468–485. SIAM, 2012.
- [14] John E. Hopcroft and Richard M. Karp. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.*, 2(4):225–231, 1973.
- [15] Sagar Kale and Sumedh Tirodkar. Maximum matching in two, three, and a few more passes over graph streams. In Klaus Jansen, José D. P. Rolim, David Williamson, and Santosh S. Vempala, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, volume 81 of *LIPICs*, pages 15:1–15:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.

- [16] Michael Kapralov. Better bounds for matchings in the streaming model. In Sanjeev Khanna, editor, *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1679–1697. SIAM, 2013.
- [17] Michael Kapralov. Space lower bounds for approximating maximum matching in the edge arrival model. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, (SODA)*, pages 1874–1893, 2021.
- [18] Michael Kapralov, Sanjeev Khanna, and Madhu Sudan. Approximating matching size from random streams. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 734–751, 2014.
- [19] Christian Konrad. A simple augmentation method for matchings with applications to streaming algorithms. In Igor Potapov, Paul G. Spirakis, and James Worrell, editors, *43rd International Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 117 of *LIPICs*, pages 74:1–74:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.
- [20] Christian Konrad and Kheeran K. Naidu. On two-pass streaming algorithms for maximum bipartite matching. *CoRR*, abs/2107.07841, 2021.
- [21] Andrew McGregor. 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) and 9th International Workshop on Randomization and Computation (RANDOM)*, pages 170–181, 2005.
- [22] Andrew McGregor and Sofya Vorotnikova. Planar matching in streams revisited. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 17:1–17:12, 2016.
- [23] Andrew McGregor and Sofya Vorotnikova. A simple, space-efficient, streaming algorithm for matchings in low arboricity graphs. In Raimund Seidel, editor, *1st Symposium on Simplicity in Algorithms (SOSA)*, volume 61 of *OASICS*, pages 14:1–14:4. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.
- [24] Ami Paz and Gregory Schwartzman. A $(2+\epsilon)$ -approximation for maximum weight matching in the semi-streaming model. In Philip N. Klein, editor, *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2153–2161. SIAM, 2017.
- [25] Mariano Zelke. Weighted matching in the semi-streaming model. *Algorithmica*, 62(1-2):1–20, 2012.

A Three-Pass Non-MMF Algorithm

In this section we prove Theorem 1.2, which we repeat below for convenience.

Theorem 1.2. *There exists a non-MMF 3-pass $(5/9 = 1/2 + 1/18)$ -approximation semi-streaming algorithm for finding a maximum size matching in a general graph.*

The algorithm used for proving Theorem 1.2 is a modified version of Algorithm 1 that appears as Algorithm 8 and manages to obtain an improved approximation ratio at the cost of making an

additional pass (*i.e.*, it makes 3 passes). The first pass of Algorithm 8 is identical to the first pass of Algorithm 1, however, the second and third passes of Algorithm 8 each consider only one of the two kinds of edges considered together in the second pass of Algorithm 1. To describe this in more details we use the terminology defined in Section 3 for describing Algorithm 1. In the second pass of Algorithm 8, we construct a set A_1 in the same way in which this is done by Algorithm 1, *i.e.*, by greedily adding to A_1 edges that connect a connection vertex of a naïve partial triangle with an isolated vertex. Then, in the third pass of Algorithm 8, we greedily collect into another set, termed A_2 , edges that connect connection vertices of two distinct naïve partial triangles. We stress that the construction of A_2 by Algorithm 8 is slightly different compared to the construction of the set carrying the same name in Algorithm 1. Upon termination of its third pass, Algorithm 8 outputs a maximum matching in the set of all the edges that it kept.

Algorithm 8: MAXIMUM MATCHING VIA GREEDY TRIANGLES - 3 PASSES

```

// First Pass
1 Let  $P \leftarrow \emptyset$ .
2 for every edge  $e$  that arrives do
3   if every connected component of the graph  $(V, P \cup \{e\})$  is either a path of length at
4     most 2 or a triangle (cycle of size 3) then
       Add  $e$  to  $P$ .

// Second Pass
5 Let  $A_1 \leftarrow \emptyset$ .
6 for every edge  $(u, v) \notin P$  that arrives do
7   Let  $C_u$  and  $C_v$  be the connected components of  $u$  and  $v$ , respectively, in  $(V, P)$ . We
       assume without loss of generality that  $|C_u| > 1$ , otherwise we swap the roles of  $u$ 
       and  $v$ . // Note that we cannot have  $|C_u| = |C_v| = 1$  because the edge  $(u, v)$ 
       was not added to  $P$  in the first pass.
8   if no edge of  $A_1$  intersects  $C_u$  and  $C_v$ ,  $|C_v| = 1$  and  $u$  is a connection vertex of  $C_u$  then
9     Add the edge  $(u, v)$  to  $A_1$ .

// Third Pass
10 Let  $A_2 \leftarrow \emptyset$ .
11 for every edge  $(u, v) \notin P \cup A_1$  that arrives do
12   Let  $C_u$  and  $C_v$  be the connected components of  $u$  and  $v$ , respectively, in  $(V, P)$ . We
       assume without loss of generality that  $|C_u| > 1$ , otherwise we swap the roles of  $u$ 
       and  $v$ . // Again, we cannot have  $|C_u| = |C_v| = 1$ .
13   if no edge of  $A_1 \cup A_2$  intersects  $C_u$  and  $C_v$ , and  $u$  and  $v$  are connection vertices of  $C_u$ 
       and  $C_v$ , respectively then
14     Add the edge  $(u, v)$  to  $A_2$ .

15 return a maximum matching in the graph  $(V, P \cup A_1 \cup A_2)$ .

```

The proof of Observation 3.1 applies to Algorithm 8 as well, and therefore, Algorithm 8 is a semi-streaming algorithm. Below we concentrate on analyzing the approximation guarantee of this algorithm. It is important to note that the analysis of the approximation ratio of Algorithm 1 up to Lemma 3.5 only depends on the behavior of the algorithm during its first pass, and therefore, applies also to Algorithm 8 since the two algorithms have identical first passes.

In principle, the proof of Lemma 3.6 applies also to Algorithm 8 since this proof is based on the method used by Algorithm 1 to construct the set A_1 , and this set is constructed in the same way by the two algorithms. However, it turns out that we need in this section a slightly stronger version of Lemma 3.6. Specifically, Lemma 3.6 includes the value ($\#non-M^*$ -triangles) in one of its terms. This value counts the number of connected components in (V, P) that are triangles and do not include within them any edge of M^* . Each such connected component is intersected by at most a single edge of A_1 or A_2 , and in this section we need to count separately the connected components of this kind that intersect edges from each one of these sets. Formally, we let ($\#non-M^*$ -triangles- A_1) be the number of connected components of (V, P) that (1) are triangles, (2) do not include any edge of M^* , and (3) intersect an edge of A_1 . Similarly, ($\#non-M^*$ -triangles- A_2) is the number of connected components of (V, P) that (1) are triangles, (2) do not include any edge of M^* , and (3) intersect an edge of A_2 . Since every partial triangle in (V, P) intersects at most a single edge of $A_1 \cup A_2$, and the sets A_1 and A_2 are disjoint, we immediately get from these definitions $(\#non-M^*$ -triangles- A_1) + ($\#non-M^*$ -triangles- A_2) \leq ($\#non-M^*$ -triangles). Furthermore, it is not difficult to verify that the proof of Lemma 3.6 in fact implies the following stronger version of the lemma.

Lemma A.1 (Stronger version of Lemma 3.6).

$$3|A_1| \geq (\#component-free) - (\#non-M^*-triangles-A_1) .$$

Lemma A.1 lower bounds the size of the set A_1 . Our next objective is to find a lower bound also for the size of A_2 . As a first step towards this goal, we upper bound the number of edges that have a potential to be added to A_2 immediately after the first pass of Algorithm 8, but are removed from this potential during the second pass of the algorithm. To formalize this notion, let us recall that ($\#component-component$) is the set of edges of M^* that connect connection vertices of two distinct partial triangles of (V, P) . Intuitively, ($\#component-component$) counts edges that have a potential to be added to A_2 ; however, for such an edge to really end up in A_2 , it is required that the two partial triangles it intersect remain naïve after the second pass. Therefore, the size of the “lost potential” is the number of edges that are counted by ($\#component-component$), but intersect at least one partial triangle of (V, P) that is also intersected by an edge of A_1 . In the following, we denote this number by ($\#lost-component-component$).

Lemma A.2.

$$(\#lost-component-component) \leq 3|A_1| - (\#component-free) + (\#non-M^*-triangles-A_1) .$$

Proof. The proof of this lemma is similar to the proof of Lemma 3.6, however, we write it fully for completeness.

We say that an edge e of M^* counted by ($\#component-free$) or ($\#component-component$) is excluded by an edge $f \in A_1$ if e and f intersect the same connected component of (V, P) . One can observe that every edge e counted by ($\#component-free$) is excluded by some edge of A_1 (possibly itself) when Algorithm 8 terminates because otherwise Algorithm 8 would have added e to A_1 , which would have resulted in e excluding itself. Therefore, the number of edges counted by ($\#component-component$) that are excluded by some edge of A_1 , which is exactly ($\#lost-component-component$), can be upper bound by the difference $|J| - (\#component-free)$, where J is the set of edges counted by either ($\#component-free$) or ($\#component-component$) that are excluded by the edges of A_1 . In other words,

$$(\#lost-component-component) \leq |J| - (\#component-free) . \tag{7}$$

Let (u, v) be an edge of A_1 , and assume without loss of generality that v is the end point of this edge which is an isolated vertex of (V, P) . This implies that u is a connection vertex of a connected component C_u of (V, P) which is either a path of length 2 or a triangle. If C_u is a path of length 2, then the edge (u, v) can exclude only edges counted by either ($\#$ component-free) or ($\#$ component-component) that intersect either v or a connection vertex of C_u , and there can be only 3 such edges because M^* is a matching. Next, consider the case in which C_u is a triangle which is not counted by ($\#$ non- M^* -triangles). In this case there can be at most 2 edges of M^* intersecting C_u , and therefore, even though (u, v) can exclude any edge of ($\#$ component-free) or ($\#$ component-component) intersecting C_u or v , there can be only 3 such edges. It remains to consider the case in which C_u is a triangle counted by ($\#$ non- M^* -triangles). In this case, (u, v) can again exclude every edge of ($\#$ component-free) or ($\#$ component-component) that intersects C_u or v , and this time there can be at most 4 such edges. Combining all the above, we get that the number $|J|$ of edges excluded by all the edges of A_1 is at most

$$\begin{aligned} & 3|A_1| + |\{e \in A_1 \mid e \text{ intersects a triangle counted by } (\# \text{non-}M^* \text{-triangles})\}| \\ &= 3|A_1| + (\# \text{non-}M^* \text{-triangles-}A_1) , \end{aligned}$$

where the equality holds because a triangle counted by ($\#$ non- M^* -triangles) is counted also by ($\#$ non- M^* -triangles- A_1) if and only if some edge of A_1 intersects it. Plugging the last upper bound on $|J|$ into Inequality (7) completes the proof of the lemma. \square

We can now prove the promised lower bound on the size of A_2 .

Lemma A.3.

$$4|A_2| \geq (\# \text{component-component}) - (\# \text{lost-component-component}) - (\# \text{non-}M^* \text{-triangles-}A_2) .$$

Proof. Recall that ($\#$ lost-component-component) counts a subset of the edges that are counted by ($\#$ component-component). Let D be the set of edges (of M^*) counted by ($\#$ component-component) but not by ($\#$ lost-component-component). We say that an edge $e \in D$ is excluded by an edge $f \in A_2$ if e and f intersect the same connected component of (V, P) . One can observe that every edge $e \in D$ is excluded by some edge of A_2 (possibly itself) when Algorithm 8 terminates because otherwise Algorithm 8 would have added e to A_2 , which would have resulted in e excluding itself. Therefore, we can upper bound the size of D by counting the number of edges excluded by the edges of A_2 .

Let (u, v) be an edge of A_2 , and let C_u and C_v be the connected components of (V, P) that include u and v respectively. Notice that since $(u, v) \in A_2$, both C_u and C_v must be either paths of length 2 or triangles. The edge (u, v) excludes every edge of D that intersects either C_u or C_v . The number of $D \subseteq M^*$ edges that intersect C_u can be at most 2, unless C_u is a triangle counted by ($\#$ non- M^* -triangles), in which case there might be 3 edges of D intersecting C_u . Since a similar claim applies to C_v , we get that the number of edges excluded by all the edges of A_2 is at most

$$4|A_2| + \sum_{e \in A_2} T(e) = 4|A_2| + (\# \text{non-}M^* \text{-triangles-}A_2) ,$$

where $T(e)$ is the number of triangles counted by ($\#$ non- M^* -triangles) that intersect e , and the equality holds since a triangle is counted by ($\#$ non- M^* -triangles- A_2) if and only if it is both counted by ($\#$ non- M^* -triangles) and intersects an edge of A_2 . As explained above, the last expression is an upper bound on the size of D . Therefore, we get

$$\begin{aligned} (\# \text{component-component}) - (\# \text{lost-component-component}) &= |D| \\ &\leq 4|A_2| + (\# \text{non-}M^* \text{-triangles-}A_2) . \end{aligned}$$

The lemma now follows by rearranging this inequality. \square

Corollary A.4.

$$12|A_1| + 12|A_2| \geq 4(\#component\text{-free}) + 3(\#component\text{-component}) \\ - 4(\#non\text{-}M^*\text{-triangles-}A_1) - 3(\#non\text{-}M^*\text{-triangles-}A_2) .$$

Proof. Plugging Lemma A.2 into Lemma A.3, we get

$$4|A_2| \geq (\#component\text{-component}) - (\#lost\text{-component-}component) - (\#non\text{-}M^*\text{-triangles-}A_2) \\ \geq (\#component\text{-component}) - (3|A_1| - (\#component\text{-free}) + (\#non\text{-}M^*\text{-triangles-}A_1)) \\ - (\#non\text{-}M^*\text{-triangles-}A_2) .$$

Rearranging the last inequality, and multiplying it by 3, yields

$$9|A_1| + 12|A_2| \geq 3(\#component\text{-component}) + 3(\#component\text{-free}) \\ - 3(\#non\text{-}M^*\text{-triangles-}A_1) - 3(\#non\text{-}M^*\text{-triangles-}A_2) .$$

The corollary now follows by adding Lemma A.1 to the last inequality. \square

Let us now define $L_2 = (\#single) + (\#double) + (\#triangle) + |A_1| + |A_2|$. The following lemma shows that one can obtain an approximation guarantee for Algorithm 8 by lower bounding L_2 . Since the proof of this lemma is very similar to the proof of Lemma 3.11, we omit it.

Lemma A.5. *Algorithm 8 outputs a matching of size at least L_2 .*

It remains now to lower bound L_2 , which we do in the next lemma. Together with Lemma A.5 and the above observation that Algorithm 8 is a semi-streaming algorithm, this lemma completes the proof of Theorem 1.2.

Lemma A.6. $L_2 \geq 5/9|M^*|$.

Proof. Observe that

$$12L_2 = 12(\#single) + 12(\#double) + 12(\#triangle) + 12|A_1| + 12|A_2| \\ \geq 12(\#single) + 12(\#double) + 12(\#triangle) + 4(\#component\text{-free}) \\ + 3(\#component\text{-component}) - 4(\#non\text{-}M^*\text{-triangles-}A_1) \\ - 3(\#non\text{-}M^*\text{-triangles-}A_2) \\ \geq 12(\#single) + 12(\#double) + 12(\#triangle) + 4(\#component\text{-free}) \\ + 3(\#component\text{-component}) - 4(\#non\text{-}M^*\text{-triangles}) ,$$

where the first inequality follows from Corollary A.4; and the second inequality holds since we already observed that $(\#non\text{-}M^*\text{-triangles}) \geq (\#non\text{-}M^*\text{-triangles-}A_1) + (\#non\text{-}M^*\text{-triangles-}A_2)$, and the value $(\#non\text{-}M^*\text{-triangles-}A_2)$ is non-negative by definition.

To further develop the last inequality, we recall that the analysis from Section 3 up until, and including, Lemma 3.4 applies to Algorithm 8 as well. Therefore,

$$\begin{aligned}
12L_2 &\geq 12(\#\text{single}) + 12(\#\text{double}) + 12(\#\text{triangle}) + 4(\#\text{component-free}) \\
&\quad + 3(\#\text{component-component}) - 4(\#\text{non-}M^*\text{-triangles}) \\
&\geq \frac{28}{3}(\#\text{single}) + 4(\#\text{double}) + 4(\#\text{triangle}) + \frac{20}{3}(\#\text{component-free}) \\
&\quad + \frac{25}{3}(\#\text{component-component}) + \frac{8}{3}(\#\text{single-single}) + 4(\#\text{single-component}) \\
&\quad + \frac{8}{3}(\#\text{middle}) - 4(\#\text{non-}M^*\text{-triangles}) \\
&\geq \frac{28}{3}(\#\text{single}) + \frac{20}{3}(\#\text{component-free}) + \frac{25}{3}(\#\text{component-component}) \\
&\quad + \frac{8}{3}(\#\text{single-single}) + 4(\#\text{single-component}) + \frac{20}{3}(\#\text{middle}) \\
&\geq \frac{20}{3}(\#\text{component-free}) + \frac{25}{3}(\#\text{component-component}) + 12(\#\text{single-single}) \\
&\quad + \frac{26}{3}(\#\text{single-component}) + \frac{20}{3}(\#\text{middle}) \\
&\geq \frac{20}{3}|M^*| + \frac{5}{3}(\#\text{component-component}) + \frac{16}{3}(\#\text{single-single}) + 2(\#\text{single-component}) ,
\end{aligned}$$

where the second Inequality holds by Inequality (1), the third inequality follows from Inequality (3) (of Lemma 3.4), the fourth inequality follows from Inequality (4) (of Lemma 3.4), and the last inequality holds by Inequality (2) (of Lemma 3.4).

The lemma now follows by rearranging the last inequality and observing that $(\#\text{single-single})$, $(\#\text{component-component})$ and $(\#\text{single-component})$ are all non-negative values by definition. \square

B Two-Pass Algorithm for Triangle-Free Graphs

In this section we analyze the approximation ratio produced by the first and second passes of TRIANGLEFREEALG. Formally, we prove the following theorem.

Theorem B.1. *There exists a 2-pass $(1/2 + 1/18)$ -approximation semi-streaming algorithm for finding a maximum size matching in a triangle-free graph, and a 2-pass $(1/2 + 1/14)$ -approximation semi-streaming algorithm for finding a maximum size matching in a bipartite graph.*

For the convenience, we state as Algorithm 9 the algorithm obtained by using only the first and second passes of TRIANGLEFREEALG. We also note that Kale and Tirodkar analyzed in [15] the same algorithm, and showed a weaker guarantee of $1/2 + 1/20$ on its approximation ratio for triangle-free graphs. Our analysis improves over their analysis; however, the approximation-ratios we achieve are still worse than the state-of-the-art for triangle-free graphs and bipartite graphs.

Before beginning our analysis of Algorithm 9, we notice that the analysis given in Section 4 up to Lemma 4.4 applies also for the algorithm we consider here. Next, we show that there is a large fractional matching in the multi-graph G_A .

Lemma B.2. *Let V_A and E_A be the sets of vertices and edges of G_A , respectively, and let \mathcal{F}_A be the fractional matching polytope of G_A defined by*

$$\begin{aligned}
\sum_{e \in \delta(u)} x_e &\leq 1 & u \in V_A \\
x_e &\geq 0 & e \in E_A .
\end{aligned}$$

Then, there exists a point x' in this polytope for which $\|x'\|_1 \geq \frac{1}{6}(4|M^| - 7|M_0|)$.*

Proof. Let M' be the set of edges of M_0 that intersect two wings of W . Since every wing of W intersects a single edge of M_0 ,

$$|M'| \geq |W| - |M_0| \geq \frac{4}{3}(|M^*| - |M_0|) - |M_0| = \frac{1}{3}(4|M^*| - 7|M_0|) ,$$

Algorithm 9: TRIANGLEFREEALG – FIRST AND SECOND PASSES

```

// First Pass
1 Let  $M_0 \leftarrow \emptyset$ .
2 for every edge  $e$  that arrives do
3    $\lfloor$  Add  $e$  to  $M_0$  if it does not intersect any edge that already belongs to  $M_0$ .
// Second Pass
4 Let  $W \leftarrow \emptyset$ .
5 for every edge  $e$  that arrives do
6   if  $e$  intersects exactly one vertex  $u \in V(M_0)$  then
7     Let  $v$  denote the other end-point of  $e$  (i.e., the end-point that is not  $u$ ).
8     if  $\deg_W(u) < 1$  and  $\deg_W(v) < 2$  then
9        $\lfloor$  Add  $e$  to  $W$ .

// Post-processing
10 Let  $G_A$  be a multi-graph over the vertices  $V \setminus V(M_0)$ . For every path  $P_{u,v}$  of length 3 in
     $W \cup M_0$  between two vertices  $u, v \in V \setminus V(M_0)$ , we add an edge  $(u, v)$  to the graph  $G_A$ .
    // This is a multi-graph because there might be multiple such paths
    between a pair of vertices of  $V \setminus V(M_0)$ .
11 Find a maximum size matching  $M_A$  in  $G_A$ .
12 Let  $\mathcal{P}_1 \leftarrow \{P_{u,v} \mid (u, v) \in M_A\}$ .
13 return  $M_0 \oplus (\bigcup_{P \in \mathcal{P}_1} E(P))$ .

```

where the second inequality follows from Lemma 4.4.

Together with the wings of W intersecting it, every edge $e \in M'$ forms a path of length 3 in $W \cup M_0$ between two vertices of $V_A = V \setminus V(M_0)$. Let us denote this path by $P(e)$. We can also observe that G_A includes an edge corresponding to the path $P(e)$, and we denote this edge below by $A(e)$. Since the paths in $\{P(e) \mid e \in M'\}$ are disjoint in the $V(M_0)$ vertices that they include (because $M' \subseteq M_0$ is a matching), they must be disjoint also in the W edges that they include. Therefore, the number of edges in $\{A(e) \mid e \in M'\}$ intersecting every vertex $u \in V_A$ is upper bounded by $\deg_W(u) \leq 2$, which implies that the vector

$$x'_e = \begin{cases} 1/2 & \text{if } e \in \{A(e) \mid e \in M'\} , \\ 0 & \text{if } e \in E_A \setminus \{A(e) \mid e \in M'\} \end{cases}$$

belongs to the polytope \mathcal{F}_A . Furthermore, $\|x'\|_1 = \frac{1}{2}|M'| \geq \frac{1}{6}(4|M^*| - 7|M_0|)$. \square

Corollary B.3. *It always holds that $|\mathcal{P}_1| = |M_A| \geq \frac{1}{9}(4|M^*| - 7|M_0|)$. Furthermore, when G is bipartite we have a stronger guarantee of $|\mathcal{P}_1| = |M_A| \geq \frac{1}{6}(4|M^*| - 7|M_0|)$.*

Proof. Consider the linear program

$$\begin{aligned} \max \quad & \|x\|_1 \\ \text{s.t.} \quad & x \in \mathcal{F}_A , \end{aligned} \tag{8}$$

and let c_A be its integrability gap. Since the size of M_A is the maximum value of an integral solution for LP (8), while x' is a feasible solution for it, we get $|M_A| \geq c_A \cdot \|x'\| \geq c_A \cdot \frac{1}{6}(4|M^*| - 7|M_0|)$. We now need to consider two cases.

- If G is bipartite, then G_A is also bipartite with the same partition of the vertices into two sides. To see why this is the case, note that every edge (u, v) of G_A corresponds to a path $P_{u,v}$ in G of length 3, and therefore, the vertices u and v appear on different sides of G . As a consequence of the bipartiteness of G_A , we get that LP (8) is integral (*i.e.*, $c_A = 1$).
- For general graphs, the integrability gap of LP (8) is known to be $2/3$. \square

Finally, to complete the proof of Theorem B.1, we consider the matching $M_0 \oplus (\bigcup_{P \in \mathcal{P}_1} E(P))$.

Proof of Theorem B.1. For bipartite graphs, the size of the matching $M_0 \oplus (\bigcup_{P \in \mathcal{P}_1} E(P))$ returned by Algorithm 9 is

$$\begin{aligned} |M_0| + |\mathcal{P}_1| &\geq |M_0| + \max\{\frac{1}{6}(4|M^*| - 7|M_0|), 0\} = \max\{\frac{1}{6}(4|M^*| - |M_0|), |M_0|\} \\ &\geq \frac{(4|M^*| - |M_0|) + |M_0|}{7} = \frac{4}{7}|M^*| = (\frac{1}{2} + \frac{1}{14}) \cdot |M^*| , \end{aligned}$$

where the first inequality follows from Corollary B.3. Similarly, for triangle-free graphs the size of the matching returned by Algorithm 9 is

$$\begin{aligned} |M_0| + |\mathcal{P}_1| &\geq |M_0| + \frac{1}{9}(4|M^*| - 7|M_0|) = \frac{1}{9}(4|M^*| + 2|M_0|) \\ &\geq \frac{1}{9}(4|M^*| + |M^*|) = \frac{5}{9}|M^*| = (\frac{1}{2} + \frac{1}{18}) \cdot |M^*| , \end{aligned}$$

where the first inequality follows again from Corollary B.3, and the second inequality follows from the fact that M_0 is a maximal matching. \square

תקציר

בעבודה זו אנו עוסקים בבעיית זיווג המקסימום בגרפים לא ממושקלים במודל זרם המידע עם שניים או שלושה מעברים על זרם הקלט. זיווג מקסימום היא בעיית יסוד במחקר האלגוריתמי במדעי המחשב. למרות שבעיה זו זכתה לתשומת לב מחקרית רבה בכלל, ובמודל זרם המידע בפרט, התוצאה הטובה ביותר הידועה כיום עבורה במודל זרם המידע במעבר יחיד מתקבלת באמצעות האלגוריתם החמדן הטבעי, שמספק יחס קירוב של $\frac{1}{2}$ בלבד.

על-מנת שבכל זאת ניתן יהיה להעמיד תוצאות מעניינות עבור הבעיה, מקובל לאפשר לאלגוריתמים לעבור על זרם הקלט יותר מפעם אחת, ואכן ישנן בספרות עבודות שהשיגו תוצאות טובות יותר לבעיה זו במספר רב של מעברים, וכן בשניים או שלושה מעברים. למיטב ידיעתנו, כל התוצאות הקיימות מסוג זה בונות במעבר הראשון זיווג מקסימלי באופן חמדן, ואז משפרות אותו במעברים הבאים. אחת התרומות של העבודה הנוכחית הינה זוג אלגוריתמים המבוססים על שיטה זו ומשפרים את יחס הקירוב הקיים עבור גרפים כלליים וגרפים חסרי משולשים (ובמשתמע גם גרפים דו-צדדיים) בשלושה מעברים. תוצאות אלה, לצד תוצאות העבר ותוצאות נוספות המוצגות בפסקה הבאה, מפורטות בטבלה המופיעה מטה.

לשיטה המתוארת לעיל, אותה נכנה מעתה "אלגוריתמי זיווג מקסימלי תחילה", יש מגבלות על יחס הקירוב שהיא מאפשרת.¹ בנוסף, באופן אינטואיטיבי, שיטה זו לא יכולה לקדם את המחקר לעבר פתרון הבעיה המרכזית – דהיינו, שיפור יחס הקירוב הקיים במעבר אחד. לכן, מעניין לחקור שיטות חלופיות המבוססות על אלגוריתמים שאינם אלגוריתמי זיווג מקסימלי תחילה, אלא מבצעים לוגיקה חכמה יותר במעבר הראשון שלהם. התוצאה המרכזית של עבודה זו היא הצגת זוג אלגוריתמים שלמיטב ידיעתנו הם הראשונים מסוג זה. אלגוריתמים אלה משפרים את יחס הקירוב הקיים עבור גרפים כלליים בשניים ושלושה מעברים. השיפור המושג ביחס הקירוב אינו רב, לצערנו, אבל אנו מאמינים כי האלגוריתמים שלנו מדגימים כי אלגוריתמים שאינם אלגוריתמי זיווג מקסימלי תחילה הם בעלי פוטנציאל מחקרי רב.

| מספר המעברים | סוג הגרף | יחס הקירוב הטוב ביותר הידוע כיום | יחס הקירוב המוצג בעבודה זו | שיטה בה התקבל השיפור |
|--------------|-------------|--|---------------------------------------|-------------------------|
| שני מעברים | דו-צדדי | $2 - \sqrt{2} \approx \frac{1}{2} + 1/11.66 \approx 0.5857$ [19] | - | - |
| | חסר-משולשים | $\frac{1}{2} + 1/16 = 0.5625$ [15] | - | - |
| | כללי | $\frac{1}{2} + 1/32 = 0.53125$ [15] | $\frac{1}{2} + 1/26 \approx 0.5385$ | ללא זיווג מקסימלי תחילה |
| שלושה מעברים | דו-צדדי | $0.6067 \approx \frac{1}{2} + 1/9.37$ [19] | $\frac{1}{2} + 1/9 \approx 0.6111$ | עם זיווג מקסימלי תחילה |
| | חסר-משולשים | $\frac{1}{2} + 1/10 = 0.6$ [15] | $\frac{1}{2} + 1/9 \approx 0.6111$ | עם זיווג מקסימלי תחילה |
| | כללי | $\frac{1}{2} + 81/1600 \approx \frac{1}{2} + 1/19.753 \approx 0.5506$ [15] | $\frac{1}{2} + 1/14.4 \approx 0.5694$ | עם זיווג מקסימלי תחילה |
| | | | $\frac{1}{2} + 1/18 \approx 0.5555$ | ללא זיווג מקסימלי תחילה |

¹למשל, Konrad and Naidu (20) פרסמו לאחרונה חסם פורמלי על יחס הקירוב שניתן להשיג באמצעות אלגוריתמים לגרפים דו צדדיים בשני מעברים המבוססים על שיטה זו.

תוכן עניינים

| | |
|----|---|
| 3 | .1 מבוא |
| 5 | .1.1 עבודות קשורות נוספות |
| 6 | .2 ידע קודם נדרש |
| 7 | .3 אלגוריתם ללא זיווג מקסימלי תחילה בשני מעברים |
| 13 | .4 אלגוריתם לגרפים חסרי-משולשים בשלושה מעברים |
| 18 | .5 אלגוריתם לגרפים כלליים בשלושה מעברים |
| 22 | .6 סיכום ועבודה עתידית |
| 24 | .A אלגוריתם ללא זיווג מקסימלי תחילה בשלושה מעברים |
| 29 | .B אלגוריתם לגרפים חסרי-משולשים בשני מעברים |

האוניברסיטה הפתוחה
המחלקה למתמטיקה ולמדעי המחשב

זיווג מקסימום ללא זיווג מקסימלי: גישה חדשה למציאת זיווג מקסימום במודל זרם המידע

עבודת תזה זו הוגשה כחלק מהדרישות לקבלת תואר
"מוסמך למדעים" M.Sc. במדעי המחשב
באוניברסיטה הפתוחה
המחלקה למתמטיקה ולמדעי המחשב

על-ידי
אריאל שארף

העבודה הוכנה בהדרכתם של
פרופ' מורן פלדמן ופרופ' זאב נוטוב

אוקטובר 2021