Planar Matching in Streams Revisited*

Andrew McGregor[1] and Sofya Vorotnikova[2]

1 **College of Information and Computer Sciences, University of Massachusetts, Amherst, USA**
mcgregor@cs.umass.edu

2 **College of Information and Computer Sciences, University of Massachusetts, Amherst, USA**
svorotni@cs.umass.edu

Abstract

We present data stream algorithms for estimating the size or weight of the maximum matching in low arboricity graphs. A large body of work has focused on improving the constant approximation factor for general graphs when the data stream algorithm is permitted $O(n \operatorname{polylog} n)$ space where n is the number of nodes. This space is necessary if the algorithm must return the matching. Recently, Esfandiari et al. (SODA 2015) showed that it was possible to estimate the maximum cardinality of a matching in a planar graph up to a factor of $24 + \epsilon$ using $O(\epsilon^{-2} n^{2/3} \operatorname{polylog} n)$ space. We first present an algorithm (with a simple analysis) that improves this to a factor $5 + \epsilon$ using the same space. We also improve upon the previous results for other graphs with bounded arboricity. We then present a factor 12.5 approximation for matching in planar graphs that can be implemented using $O(\log n)$ space in the adjacency list data stream model where the stream is a concatenation of the adjacency lists of the graph. The main idea behind our results is finding "local" fractional matchings, i.e., fractional matchings where the value of any edge e is solely determined by the edges sharing an endpoint with e. Our work also improves upon the results for the *dynamic* data stream model where the stream consists of a sequence of edges being inserted and deleted from the graph. We also extend our results to weighted graphs, improving over the bounds given by Bury and Schwiegelshohn (ESA 2015), via a reduction to the unweighted problem that increases the approximation by at most a factor of two.

1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases data streams, planar graphs, arboricity, matchings

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2016.17

1 Introduction

A large body of work has focused on finding better approximation algorithms for finding large graph matchings in the data stream model [1, 11, 24, 26, 7, 8, 13, 19, 18, 15, 21, 22, 4, 20, 17]. In this model, the edges of an input graph on n nodes arrive in an arbitrary order and the algorithm has a limited amount of memory available. For a survey of graph algorithms in this model, see [25]. Specifically, a sequence of papers have presented algorithms using $O(n \operatorname{polylog} n)$ bits of space that have steadily reduced the best known approximation ratio for maximum weighted matching: Feigenbaum et al. [13] initially presented a 6 approximation; McGregor [24] then presented a 5.828 approximation; this was reduced to 5.585 by Zelke [26]; and then to 4.911 by Epstein et al. [11]; and the best known result is a 4 approximation

* Supported by NSF CAREER Award CCF-0953754 and CCF-1320719 and a Google Faculty Research Award.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016).
Editors: Klaus Jansen, Claire Matthieu, José D. P. Rolim, and Chris Umans; Article No. 17; pp. 17:1–17:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

due to Crouch and Stubbs [7]. The best known result for maximum cardinality matching is a trivial 2 factor that follows by constructing a greedy matching. Konrad et al. [21] showed that this can be slightly improved if the edges are ordered randomly. Kapralov [18] proved a lower bound of $e/(e-1) \approx 1.58$ on the best possible approximation factor when using only $O(n \operatorname{polylog} n)$ space. Note that all the above algorithms return the large matching rather than just estimating its weight.

A natural question is whether constant approximation is possible using $o(n)$ space if we only need to estimate the weight of the matching. Recently, Esfandiari et al. [12] showed the surprising result that it was indeed possible in the case of planar graphs and more generally, *bounded arboricity graphs.* Recall that a graph G has arboricity α if the set of edges of G can be partitioned into at most α forests. For example, a planar graph has arboricity $\alpha = 3$. Esfandiari et al. presented a $(5\alpha+9)(1+\epsilon)$ approximation using $O(\alpha\epsilon^{-2}n^{2/3} \operatorname{polylog} n)$ space. In the case of planar graphs this corresponds to a $24+\epsilon$ approximation. Very recently Chitnis et al. [5] and Bury and Schwiegelshohn [4], showed that the same approximation was possible in the dynamic graph model where the stream consists of edges being added and deleted from the underlying graph. Bury and Schwiegelshohn [4] also showed that it was possible to extend the result to weighted graphs but with an $O(\alpha^4)$ approximation factor. All of these results rely on an interesting structural result proved by Esfandiari et al. [12] that relates the size of the maximum cardinality matching in a graph G of arboricity α to the number of nodes of "high" degree and the number of edges whose endpoints are both "low" degree. Specifically:

▶ **Theorem 1** (Esfandiari et al. [12]). *Let* $\operatorname{match}(G)$ *be the size of the maximum cardinality matching in* G*. Then*

$$\operatorname{match}(G)/2 \le (h+s)/2 \le \max(h,s) \le (2.5\alpha + 4.5)\operatorname{match}(G)$$

where h *is the number of nodes with degree greater than* $2\alpha+3$ *and* s *is the number of edges that remain after all such nodes are removed.*

The result then reduces the problem of estimating the matching to estimating the quantities h and s. A similar, but weaker bound, is implicit in Czygrinow et al. [9].

1.1 Our Results

The first contribution of this paper is to identify a new quantity that a) yields tighter bounds for $\operatorname{match}(G)$ and b) can be approximated in the data stream model.

▶ **Theorem 2** (Structural Result). *For an edge* $e = \{u,v\}$ *define* $x_e = \min\left(\frac{1}{\deg(u)}, \frac{1}{\deg(v)}, \frac{1}{\alpha+1}\right)$. *Then,*

$$\operatorname{match}(G) \le (\alpha+1)\sum_{e\in E} x_e \le (\alpha+2)\operatorname{match}(G)\,.$$

For example, for a planar graph $\alpha = 3$ and hence $\sum_{e\in E} x_e$ determines $\operatorname{match}(G)$ up to a factor of 5. For a bipartite planar graph (for such graphs, $\alpha = 2$) the result can be further improved to a factor of 3. The proof of Theorem 2 can be found in Section 2.2 and has the advantage of being conceptually simpler than the proof of Theorem 1. The main idea is to prove the result via consideration of fractional matchings, specifically "local" fractional matching where the value of any edge e can be determined by only considering the edges incident to e. We also show a result for local fractional matchings for weighted graphs.

Using Theorem 2, we show that $\mathrm{match}(G)$ on unweighted graphs can be approximated up to a factor $(\alpha+2)(1+\epsilon)$ using $O(\epsilon^{-2}\alpha n^{2/3}\,\mathrm{polylog}\, n)$ bits of space. Furthermore, this result can be generalized to weighted graphs with approximation factor $2(\alpha+2)(1+\epsilon)$ and to the dynamic graph stream model with a slight increase in space. We also show that it is possible to estimate $\mathrm{match}(G)$ up to a factor $(\alpha+2)^2/2$ using only the degree sequence of G. This result immediately leads to a $O(\log n)$ space algorithm in the adjacency list stream model where the stream is a concatenation of the adjacency lists of the graph. The result can also be generalized to weighted graphs while losing only an additional factor of 2.

2 Graph Properties

In this section we present a variety of results relating the size or weight of a maximum matching in a low arboricity graph to "simpler" quantities. We start with some necessary preliminaries about fractional matchings.

2.1 Preliminaries

Define the fractional matching polytope for a graph G as:

$$\mathsf{FM}(G) = \{\mathbf{x} \in \mathbb{R}^E : x_e \geq 0 \text{ for all } e \in E, \sum_{e\in E: u \in e} x_e \leq 1 \text{ for all } u \in V\}\,.$$

We say any $\mathbf{x} \in \mathsf{FM}(G)$ is a fractional matching. The size of this fractional matching is $\sum_{e\in E} x_e$ and for a graph where edge e has weight w_e, the weight of the matching is $\sum_{e\in E} w_e x_e$. A standard result on fractional matching is that the maximum size of a fractional matching is at most a factor $3/2$ larger than the maximum size of an (integral) matching. We will also make use of the following lemma which is a simple corollary of Edmonds Matching Polytope theorem [10].

▶ **Lemma 3.** *For $U \subset V$, let $G[U]$ denote the induced subgraph on U. Let $\mathbf{x} \in \mathsf{FM}(G)$ and suppose there exist $\lambda_3, \lambda_5, \lambda_7 \ldots$ such that*

$$\forall U \subset V \text{ where } |U| \in \{3,5,7,\ldots\}\ ,\ \sum_{e \in G[U]} x_e \leq \lambda_{|U|} \left(\frac{|U|-1}{2}\right)\,.$$

Then for any edge weights $\{w_e\}_{e\in E}$,

$$\sum_{e\in E} w_e x_e \leq \max(1, \lambda_3, \lambda_5, \ldots)\,\mathrm{match}(G)$$

where $\mathrm{match}(G)$ is the weight of the maximum weighted (integral) matching.

Proof. By Edmonds theorem, $\mathrm{match}(G) = \max_{\mathbf{z}\in \mathsf{IM}(G)} \sum_e w_e z_e$ where

$$\mathsf{IM}(G) = \{\mathbf{x} \in \mathbb{R}^E : x_e \geq 0 \text{ for all } e \in E, \sum_{e\in E: u\in e} x_e \leq 1 \text{ for all } u \in V, \sum_{e\in G[U]} x_e \leq \left(\frac{|U|-1}{2}\right) \text{ for all } U \subset V \text{ of odd size}\}\,.$$

But $\frac{\mathbf{x}}{\max(1,\lambda_3,\lambda_5,\ldots)} \in \mathsf{IM}(G)$ and so $\sum_{e\in E} w_e x_e \leq \max(1, \lambda_3, \lambda_5, \ldots)\,\mathrm{match}(G)$ as required. ◀

For the streaming applications we will be interested in fractional matchings that can be computed locally.

▶ **Definition 4.** For a given graph G, we say a fractional matching $\mathbf{x} \in \mathsf{FM}(G)$ is *local* if every x_e is only a function of the edges (and their weights in the case of a weighted graph) that share an end point with e.

2.2 Local Fractional Matching

Define $\mathbf{x} \in \mathbb{R}^E$ where for $e = \{u, v\} \in E$, we set

$$x_e = \min\left(\frac{1}{\deg(u)}, \frac{1}{\deg(v)}, \frac{1}{\alpha+1}\right) .$$

The next two theorems show that $\mathbf{x}$ is a local fractional matching and

$$\frac{1}{\alpha+1} \cdot \mathrm{match}(G) \le \mathrm{score}(\mathbf{x}) \le \frac{\alpha+2}{\alpha+1} \cdot \mathrm{match}(G)$$

where $\mathrm{score}(\mathbf{x}) = \sum_e x_e$. This proves Theorem 2 and we note that the upper bound can be improved slightly if α is even. In Section 3.1, we show that it is possible to efficiently estimate $\mathrm{score}(\mathbf{x})$ in the data stream model.

▶ **Theorem 5.** $\mathbf{x} \in \mathsf{FM}$ *and*

$$\frac{\mathrm{score}(\mathbf{x})}{\mathrm{match}(G)} \le \begin{cases} \frac{\alpha+2}{\alpha+1} & \textit{if } \alpha \textit{ odd,} \\ \frac{\alpha+3}{\alpha+2} & \textit{if } \alpha \textit{ even.} \end{cases}$$

Furthermore, if G is bipartite then $\mathrm{score}(\mathbf{x}) \le \mathrm{match}(G)$.

Proof. First note that $x_e \ge 0$ for each $e \in E$ and for any $u \in V$,

$$\sum_{e \in E: u \in e} x_e \le \sum_{e \in E: u \in e} 1/\deg(u) = 1 .$$

and hence $\mathbf{x} \in \mathsf{FM}$. The bound for bipartite graphs follows because the maximum size of a fractional matching in a bipartite graph equals the maximum size of an integral matching. For the rest of the result, we appeal to Lemma 3. Since $\mathbf{x} \in \mathsf{FM}$, it is simple to show that $\mathbf{x}$ satisfies the conditions of the lemma with $\lambda_t \le t/(t-1)$; this follows because $\sum_{e \in G[U]} x_e \le |U|/2$ for any $\mathbf{x} \in \mathsf{FM}$. Furthermore, since there are at most $\binom{|U|}{2}$ edges in $G[U]$ and $x_e \le 1/(\alpha+1)$ for all e,

$$\sum_{e \in G[U]} x_e \le \binom{|U|}{2} \frac{1}{\alpha+1} = \frac{|U|-1}{2} \cdot \frac{|U|}{\alpha+1} .$$

Therefore, $\lambda_t \le \min\left(t/(t-1), t/(\alpha+1)\right)$. Consequently,

$$\max_{t \text{ odd}} \lambda_t = \begin{cases} \frac{\alpha+2}{\alpha+1} & \text{if } \alpha \text{ odd,} \\ \frac{\alpha+3}{\alpha+2} & \text{if } \alpha \text{ even.} \end{cases}$$ ◀

We next bound $\mathrm{score}(\mathbf{x})$ in terms of the number of high degree vertices and edges that are not incident to high degree vertices. As observed in previous work, these two quantities can then easily be related the size of the maximum matching.

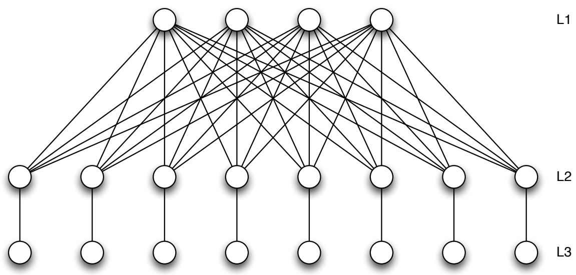

Figure 1 A tight example for Theorem 6. Let L_1 consist of α nodes whereas L_2 and L_2 consist of $n \gg \alpha$ nodes. The edges are a complete bipartite graph of L_1 and L_2 and a matching between L_2 and L_3. Then $\operatorname{score}(\mathbf{x}) = \alpha n \times 1/n + n \times 1/(\alpha+1)$ and $\operatorname{match}(G) = n$. Hence $\operatorname{match}(G)/\operatorname{score}(\mathbf{x})$ tends to $\alpha + 1$ as n tends to infinity.

▶ **Theorem 6.** *Let h be the number of "heavy" nodes with degree at least $\alpha + 2$ and s be the number of "shallow" edges whose endpoints are both not heavy. Then,*

$$\operatorname{score}(\mathbf{x}) \geq 2h/(\alpha+2) + s/(\alpha+1)\,.$$

Furthermore, $\operatorname{match}(G) \leq (\alpha+1)\operatorname{score}(\mathbf{x})$.

Proof. Let d_i be the degree of node i and assume $d_1 \geq d_2 \geq d_3 \geq \ldots$. Let $b_i = |\{j < i : \{i,j\} \in E\}|$ and $c_i = |\{i < j : \{i,j\} \in E\}|$, i.e., the number of neighbors of node i that have higher or lower degree respectively than node i where ties are broken by the ordering supposed in the above line. Consider labeling an edge e with weight x_e where we first label edges incident to node 1, then the (remaining unlabeled) edges incident to node 2, etc. Then $c_1 = d_1$ edges get labeled with $\min(1/d_1, 1/(\alpha+1))$, c_2 edges get labeled with $\min(1/d_2, 1/(\alpha+1))$, c_3 edges get labeled with $\min(1/d_3, 1/(\alpha+1))$ etc. Let $\theta = \alpha + 2$, then

$$\begin{aligned}
\operatorname{score}(\mathbf{x}) &= \sum_i c_i \min(1/d_i, 1/(\alpha+1)) \\
&= \sum_{i:d_i \geq \theta} c_i/d_i + \sum_{i:d_i \leq \theta-1} c_i/(\alpha+1) \\
&= h - \sum_{i:d_i \geq \theta} b_i/d_i + \sum_{i:d_i \leq \theta-1} c_i/(\alpha+1) \\
&\geq h - (\sum_{i:d_i \geq \theta} b_i)/\theta + (\sum_{i:d_i \leq \theta-1} c_i)/(\alpha+1)
\end{aligned}$$

Note that $\sum_{i:d_i \geq \theta} b_i$ is the number of edges in the induced subgraph on heavy nodes. This is at most αh because these edges in this induced subgraph can be partitioned into at most α forests. Similarly, $\sum_{i:d_i \leq \theta-1} c_i$ is the number of shallow edges. Therefore

$$\operatorname{score}(\mathbf{x}) \geq h(1 - \alpha/\theta) + s/(\alpha+1) = 2h/(\alpha+2) + s/(\alpha+1)$$

as required. Note that $h + s \geq \operatorname{match}(G)$ because every edge in a matching is either shallow or has at least one heavy node as an endpoint. Therefore

$$\operatorname{score}(\mathbf{x}) \geq (h+s)/(\alpha+1) \geq \operatorname{match}(G)/(\alpha+1)\,. \qquad ◀$$

See Figure 1 for an example that shows that the above theorem is tight.

2.3 Local Fractional Matchings for Weighted Graphs

In this section we show how to find a good local fractional matching for weighted graphs. We will not use this result in our algorithm for approximating the maximum weighted matching in Section 3 since a better approximation can be achieved using other ideas combined with the fractional matching proposed for the unweighted case. However, we think the structural result is interesting and could be useful in other computational models.

Define $\mathbf{y} \in \mathbb{R}^E$ where for $e = \{u, v\} \in E$, we set

$$y_e = \min\left(\frac{1}{\deg^e(u)\cdot H(\deg(u))}, \frac{1}{\deg^e(v)\cdot H(\deg(v))}, \frac{1}{\alpha+1}\right)$$

where $\deg^e(u)$ and $\deg^e(v)$ are the number of edges at least as heavy as e that are incident to u and v respectively and $H(r) = 1/1 + 1/2 + \ldots + 1/r$ is the harmonic function.

The next two theorems show that $\mathbf{y}$ is a local fractional matching and

$$\frac{1}{H(D)\cdot(\alpha+1)}\,\mathrm{match}(G) \le \mathrm{score}(\mathbf{y}) \le \frac{\alpha+2}{\alpha+1}\,\mathrm{match}(G)$$

where $\mathrm{score}(\mathbf{y}) = \sum_e w_e y_e$ and D is the maximum degree of the graph. Note that D can be as large as $n-1$ even for a low arboricity graph. However, since the average degree of G is at most 2α, we expect D to typically be much smaller for many low arboricity graphs of interest.

▶ **Theorem 7.** $\mathbf{y} \in \mathsf{FM}$ *and*

$$\frac{\mathrm{score}(\mathbf{y})}{\mathrm{match}(G)} \le \begin{cases} \frac{\alpha+2}{\alpha+1} & \text{if } \alpha \text{ odd,} \\ \frac{\alpha+3}{\alpha+2} & \text{if } \alpha \text{ even.} \end{cases}$$

Furthermore, if G is bipartite then $\mathrm{score}(\mathbf{y}) \le \mathrm{match}(G)$.

Proof. For all $u \in V$,

$$\sum_{e\in E:u\in e} x_e \le \frac{1}{H(\deg(u))}\sum_{e\in E:u\in e}\frac{1}{\deg^e(u)} \le \frac{1}{H(\deg(u))}(1/1+1/2+\ldots+1/\deg(u)) = 1\ ,$$

and hence $\mathbf{y} \in \mathsf{FM}$. The result of the proof follows as in the proof of Theorem 5 since $y_e \le 1/(\alpha+1)$ for all e. ◀

▶ **Theorem 8.** $\mathrm{match}(G) \le H(D)(\alpha+1)\,\mathrm{score}(\mathbf{y})$ *where D is the maximum degree.*

Proof. Let z_e be the optimum weighted integral matching. Let $0 < w_1 < w_2 < w_3 < \ldots$ be the distinct weights in the graph and let $w_0 = 0$. Let G_k be the unweighted graph formed from the original weighted graph where all edges whose weight is $< w_k$ are deleted and the other edges are given weight 1. Let z_e^k be the optimum unweighted integral matching for G_k and let $\deg_k(u)$ be the degree of node u in G_k.

Then,

$$\begin{aligned}\mathrm{score}(\mathbf{z}) = \sum_e z_e w_e \quad &\le \quad \sum_k (w_k - w_{k-1})\sum_{e\in G_k} z_e^k \\ &\le \quad (\alpha+1)\sum_k (w_k - w_{k-1})\sum_{e\in G_k}\min\left(\frac{1}{\deg_k(u)}, \frac{1}{\deg_k(v)}, \frac{1}{\alpha+1}\right)\end{aligned}$$

where the last inequality follows by our result for the unweighted case.

But for any $e \in E$,

$$\begin{aligned}
& \sum_{k:e\in G_k} (w_k - w_{k-1}) \min\left(\frac{1}{\deg_k(u)}, \frac{1}{\deg_k(v)}, \frac{1}{\alpha+1}\right) \\
\leq\ & \sum_{k:e\in G_k} (w_k - w_{k-1}) \min\left(\frac{1}{\deg^e(u)}, \frac{1}{\deg^e(v)}, \frac{1}{\alpha+1}\right) \\
\leq\ & w_e \min\left(\frac{1}{\deg^e(u)}, \frac{1}{\deg^e(v)}, \frac{1}{\alpha+1}\right) \\
\leq\ & H(D) w_e y_e
\end{aligned}$$

where the first inequality follows because $\deg_k(u) \geq \deg^e(u)$ for all k such that $e \in G_k$. Therefore $\mathrm{match}(G) \leq H(D)(\alpha+1)\,\mathrm{score}(\mathbf{y})$ as claimed. ◀

2.4 Exact Degree Distribution

Using ideas from the previous sections, we now show that the size of the maximum matching can be approximated up to a $O(\alpha^2)$ factor given just the degree distribution of G. Specifically, consider the following estimate:

$$\tilde{M} = \sum_{u\in V} \min(\alpha + 1 - \deg(u)/2, \deg(u)/2)\,.$$

The next theorem shows that $\tilde{M}$ is a $O(\alpha^2)$ approximation for $\mathrm{match}(G)$.

▶ **Theorem 9.** $\mathrm{match}(G) \leq \tilde{M} \leq \frac{(\alpha+2)^2}{2} \cdot \mathrm{match}(G)$.

Proof. Let h be the number of "heavy" nodes with degree at least $\alpha+2$. Partition the edges E into E_0, E_1, and E_2 depending on whether the edge has zero, one, or two heavy endpoints. Note that E_0 is just the set of shallow edges. Then,

$$\begin{aligned}
& \sum_{u\in V} \min(\alpha+1-\deg(u)/2, \deg(u)/2) \\
=\ & \sum_{u\in V} \deg(u)/2 - \max(\deg(u) - \alpha - 1, 0) \\
=\ & |E_0| + |E_1| + |E_2| - \left(\sum_{u:\deg(u)\geq\alpha+2} \max(\deg(u) - \alpha - 1, 0)\right) \\
=\ & |E_0| + |E_1| + |E_2| - \left(\sum_{u:\deg(u)\geq\alpha+2} \deg(u)\right) + h(\alpha+1) \\
=\ & |E_0| + |E_1| + |E_2| - |E_1| - 2|E_2| + h(\alpha+1) \\
=\ & |E_0| - |E_2| + h(\alpha+1)
\end{aligned}$$

First note that $|E_2| \leq \alpha h$ because the number of edges in any induced subgraph is at most α times the number of nodes in that subgraph. Hence,

$$|E_0| - |E_2| + h(\alpha+1) \geq |E_0| + h \geq \mathrm{match}(G)\,.$$

By appealing to Theorem 6 and Theorem 5

$$\begin{aligned}|E_0| - |E_2| + h(\alpha+1) &\le |E_0| + h(\alpha+1)\\ &\le \frac{(\alpha+2)(\alpha+1)}{2} \cdot (|E_0|/(\alpha+1) + 2h/(\alpha+2))\\ &\le \frac{(\alpha+2)(\alpha+1)}{2} \cdot \frac{\alpha+2}{\alpha+1} \cdot \mathrm{match}(G)\\ &\le \frac{(\alpha+2)^2}{2} \cdot \mathrm{match}(G)\,.\end{aligned}$$ ◀

3 Data Stream Algorithms

In this section we briefly discuss the improved algorithmic results that can be achieved via the results from the previous section.

3.1 Arbitrary Order Graph Streams

In the arbitrary order graph stream model, the stream consists of the edges of the input graph G in arbitrary order. The goal is to estimate the size of the maximum cardinality matching using only a single pass over this stream and limited memory.

From Theorem 2, we know we can estimate the size of the maximum cardinality via the following quantity,

$$A := \sum_{\{u,v\}\in E} \min\left(\frac{1}{\deg(u)}, \frac{1}{\deg(v)}, \frac{1}{\alpha+1}\right).$$

To do this we first show that A can be estimated via the quantity,

$$A_S := \sum_{\{u,v\}\in E: u,v\in S} \min\left(\frac{1}{\deg(u)}, \frac{1}{\deg(v)}, \frac{1}{\alpha+1}\right).$$

where S is a subset of V formed by sampling each node independently with probability p. The next lemma shows that A_S is within a $1+\epsilon$ factor of Ap^2 with probability at least $3/4$ assuming p is sufficiently large. Note that a similar approach is taken in Esfandiari et al. [12] and Chitnis et al. [5] in the context of their algorithm to estimate the number of high degree vertices and edges that are not incident to high degree vertices.

▶ **Lemma 10.** *If $p \ge \sqrt{12\epsilon^{-2}A^{-1}}$, then $\mathbb{P}\left[|A_S - Ap^2| \le \epsilon \cdot Ap^2\right] \ge 3/4$.*

Proof. For each edge $e = \{u,v\} \in E$, let $x_e = \min(1/\deg(u), 1/\deg(v), 1/(\alpha+1))$ and define a random variable X_e where $X_e = x_e$ if $u, v \in S$ and $X_e = 0$ otherwise. Note that $A_S = \sum_{e\in E} X_e$. Then, the expectation and variance of A_S are $\mathbb{E}[A_S] = Ap^2$ and

$$\mathbb{V}[A_S] = \sum_{e\in E}\sum_{e'\in E} \mathbb{E}[X_e X_{e'}] - \mathbb{E}[X_e]\,\mathbb{E}[X_{e'}]\,.$$

Note that

$$\sum_{e'\in E} \mathbb{E}[X_e X_{e'}] - \mathbb{E}[X_e]\,\mathbb{E}[X_{e'}] = \begin{cases} x_e^2(p^2-p^4) & \text{if } e = e' \\ x_e x_{e'}(p^3-p^4) & \text{if } e \text{ and } e' \text{ share exactly one endpoint} \\ 0 & \text{if } e \text{ and } e' \text{ share no endpoints}\end{cases}.$$

Since the sum of all $x_{e'}$ that share an endpoint with e is at most 2 because $\mathbf{x} \in \mathsf{FM}$,

$$\mathbb{V}[A_S] \leq \left(\sum_{e \in E} x_e^2(p^2 - p^4)\right) + 2A(p^3 - p^4) \leq 3Ap^2 .$$

We then use Chebyshev's inequality to obtain

$$\mathbb{P}\left[|A_S - Ap^2| \leq \epsilon Ap^2\right] \leq \frac{3Ap^2}{\epsilon^2 A^2 p^4} = \frac{3}{\epsilon^2 Ap^2} \leq 3/4 . \qquad ◀$$

Given this key lemma, the algorithm and analysis proceed similarly to that of Esfandiari et al. [12]. Specifically, two algorithms are run in parallel: a greedy matching algorithm and a sampling-based algorithm. The greedy matching algorithm uses $O(n^{2/3} \log n)$ space to find a maximal matching of size at least $\min(n^{2/3}, \mathrm{match}(G)/2)$. The sampling-based algorithm uses $O(\alpha n^{2/3} \log n)$ space to sample each node with probability $p = \Theta(\epsilon^{-1}/n^{2/3})$ and then find all edges whose endpoints are both sampled along with the degrees of the sampled edges. If the greedy matching has size less than $n^{2/3}$ then it is necessarily a 2 approximation of $\mathrm{match}(G)$. If not, we can use the estimate of A based on the nodes sampled since in this case $A = \Omega(n^{2/3})$. Similarly, extensions of the above approach for dynamic graph streams [5, 4] go through with the improved approximation factor. To summarize:

▶ **Theorem 11.** *There exists a single pass data stream algorithm using $O(\alpha\epsilon^{-1} n^r \log \delta^{-1})$ space that returns a $(\alpha + 2)(1 + \epsilon)$ approximation of the maximum matching with probability at least $1 - \delta$. In the insert-only model, $r = 2/3$ and in the insert-delete model $r = 4/5$.*

3.2 Adjacency List Graph Streams

In the *adjacency list model*[1] the edges incident to each node v appear consecutively in the stream [23, 3, 2]. Thus, every edge $\{u, v\}$ will appear twice: once when we view the adjacency list of u and once for v. Aside from that constraint, the stream is ordered arbitrarily. For example, for the graph consisting of a cycle on three nodes $V = \{v_1, v_2, v_3\}$, a possible ordering of the stream could be $\langle v_3v_1, v_3v_2, v_2v_3, v_2v_1, v_1v_2, v_1v_3\rangle$. Note that in this model it is trivial to compute

$$\tilde{M} = \sum_{u \in V} \min(\alpha + 1 - \deg(u)/2, \deg(u)/2) .$$

in $O(\log n)$ space since the degree of a node can be calculated exactly when the adjacency list of that node appears. The next theorem immediately follows from Theorem 9.

▶ **Theorem 12.** *An $(\alpha + 2)^2/2$-approximation of the size of maximum matching can be computed using $O(\log n)$ in the adjacency list model. In particular, this yields a 12.5-approximation for planar graphs.*

3.3 Extension to Weighted Graphs

Let $G = (V, E)$ be a weighted graph where edge e has weight $w_e \in [1, W]$ where $W = \mathrm{poly}(n)$. In this section we show that it is possible to reduce the problem of finding a large weighted matching in G to finding large cardinality matchings. Specifically, we show that given a

[1] The adjacency list order model is closely related to the vertex arrival model [15, 18] and row-order arrival model considered in the context of linear algebra problems [6, 14].

t-approximation algorithm for the unweighted problem, there is a $2(1+\epsilon)t$-approximation the maximum weighted problem where the latter algorithm using a factor $O(\epsilon^{-1}\log n)$ more space. This reduction uses ideas from work by Crouch and Stubbs [7].[2] This immediately implies a $2(2+\alpha)(1+\epsilon)$-approximation algorithm for weighted graphs in the arbitrary order model and a $(2+\alpha)^2(1+\epsilon)$-approximation algorithm for weighted graphs in the adjacency list model.

Reduction to Unweighted Matchings

For $k = 0, 1, \ldots, \lfloor \log_{1+\epsilon} W \rfloor$, define the unweighted graph $G_k = (V, E_k)$ where $e \in E_k$ iff $w_e \geq (1+\epsilon)^k$ where w_e is the weight of e in the original weighted graph. Note that $E_0 \subseteq E_1 \subseteq E_2 \subseteq \ldots$ and, in particular, $E_0, E_1, \ldots$ is *not* a partition of E.

▶ **Lemma 13.** *Let* $\mathrm{match}(G)$ *be the weight of the maximum weighted matching in* G *and let* $\tilde{m}_k$ *be a* t*-approximation of the size of the maximum cardinality matching in* G_k*. Then,*

$$\mathrm{match}(G)/t \leq \sum_{k\geq 0} f(k)\cdot \tilde{m}_k \leq 2\cdot(1+\epsilon)\cdot \mathrm{match}(G)$$

where

$$f(k) = \begin{cases} (1+\epsilon)^{k+1} - (1+\epsilon)^k & \text{if } k > 0\,, \\ (1+\epsilon) & \text{if } k = 0\,. \end{cases}$$

Proof. Let m_k be the size of the maximum cardinality matching in G_k and let M be the set of edges in the maximum weighted matching in G. To prove the left inequality, observe that

$$\sum_{k\geq 0} f(k)\cdot \tilde{m}_k \geq \sum_{k\geq 0} f(k)\cdot m_k/t \geq \sum_{k\geq 0} f(k)\cdot |M\cap E_k|/t \geq \mathrm{match}(G)/t\ ,$$

where the last inequality follows since

$$(1+\epsilon)w_e \geq \sum_{k: w_e \geq (1+\epsilon)^k} f(k) \geq w_e\,. \tag{1}$$

We now prove the right inequality. Consider the matching R formed by taking a maximal matching in E_r where $r = \lfloor \log_{1+\epsilon} W\rfloor$; extending this to a maximal matching in E_{r-1}; extending this to a maximal matching in E_{r-2} as so on. Note that since $R\cap E_k$ is a maximal matching in E_k, we have $\tilde{m}_k \leq m_k \leq 2|R\cap E_k|$. Therefore,

$$\sum_{k\geq 0} f(k)\cdot \tilde{m}_k \leq 2\sum_{k\geq 0} f(k)\cdot |R\cap E_k| \leq 2(1+\epsilon)\sum_{e\in R} w_e \leq 2(1+\epsilon)\,\mathrm{match}(G)\ ,$$

where the second last inequality follows from Eq. 1. ◀

4 Conclusion

We established better approximation ratios for the data stream problem of estimating the maximum weight and cardinality matchings in graphs of bounded arboricity α. The main technical result is that the relatively simple quantity $\sum_{\{u,v\}\in E} \min\left(1/\deg(u), 1/\deg(v), 1/(\alpha+1)\right)$ determines the size of the maximum cardinality matching up to a factor of $(\alpha+2)$, e.g., 5 in the case of planar graphs, and this quantity can be estimated efficiently in the data stream model. Other results included establishing that the degree distribution determines the size of the maximum cardinality matching up to a factor of $(\alpha+2)^2/2$, e.g., 12.5 in the case of planar graphs.

[2] Concurrent with our work, Grigorescu, Monemizadeh, and Zhou [16] designed a similar reduction.

References

1. Kook Jin Ahn and Sudipto Guha. Linear programming in the semi-streaming model with application to the maximum matching problem. *Inf. Comput.*, 222:59–79, 2013. `doi:10.1016/j.ic.2012.10.006`.
2. Ziv Bar-Yossef, Ravi Kumar, and D. Sivakumar. Reductions in streaming algorithms, with an application to counting triangles in graphs. In *Proc. of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms, January 6-8, 2002, San Francisco, CA, USA.*, pages 623–632, 2002. URL: `http://dl.acm.org/citation.cfm?id=545381.545464`.
3. Luciana S. Buriol, Gereon Frahling, Stefano Leonardi, Alberto Marchetti-Spaccamela, and Christian Sohler. Counting triangles in data streams. In *Proceedings of the Twenty-Fifth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 26-28, 2006, Chicago, Illinois, USA*, pages 253–262, 2006. `doi:10.1145/1142351.1142388`.
4. Marc Bury and Chris Schwiegelshohn. Sublinear estimation of weighted matchings in dynamic data streams. In *Algorithms – ESA 2015 – 23rd Annual European Symposium, Patras, Greece, September 14-16, 2015, Proceedings*, pages 263–274, 2015. `doi:10.1007/978-3-662-48350-3_23`.
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 *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1326–1344, 2016. `doi:10.1137/1.9781611974331.ch92`.
6. Kenneth L. Clarkson and David P. Woodruff. Numerical linear algebra in the streaming model. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 – June 2, 2009*, pages 205–214, 2009. `doi:10.1145/1536414.1536445`.
7. Michael Crouch and Daniel S. Stubbs. Improved streaming algorithms for weighted matching, via unweighted matching. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014, September 4-6, 2014, Barcelona, Spain*, pages 96–104, 2014. `doi:10.4230/LIPIcs.APPROX-RANDOM.2014.96`.
8. Michael S. Crouch, Andrew McGregor, and Daniel Stubbs. Dynamic graphs in the sliding-window model. In *Algorithms – ESA 2013 – 21st Annual European Symposium, Sophia Antipolis, France, September 2-4, 2013. Proceedings*, pages 337–348, 2013. `doi:10.1007/978-3-642-40450-4_29`.
9. Andrzej Czygrinow, Michal Hanckowiak, and Edyta Szymanska. Fast distributed approximation algorithm for the maximum matching problem in bounded arboricity graphs. In *Algorithms and Computation, 20th International Symposium, ISAAC 2009, Honolulu, Hawaii, USA, December 16-18, 2009. Proceedings*, pages 668–678, 2009. `doi:10.1007/978-3-642-10631-6_68`.
10. Jack Edmonds. Maximum matching and a polyhedron with 0,1-vertices. *Journal of Research of the National Bureau of Standards, 69:125-130*, 1965.
11. Leah Epstein, Asaf Levin, Julián Mestre, and Danny Segev. Improved approximation guarantees for weighted matching in the semi-streaming model. *SIAM J. Discrete Math.*, 25(3):1251–1265, 2011. `doi:10.1137/100801901`.
12. Hossein Esfandiari, Mohammad Taghi Hajiaghayi, Vahid Liaghat, Morteza Monemizadeh, and Krzysztof Onak. Streaming algorithms for estimating the matching size in planar graphs and beyond. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 1217–1233, 2015. `doi:10.1137/1.9781611973730.81`.

13 Joan Feigenbaum, Sampath Kannan, Andrew McGregor, Siddharth Suri, and Jian Zhang. On graph problems in a semi-streaming model. *Theor. Comput. Sci.*, 348(2):207–216, 2005. doi:10.1016/j.tcs.2005.09.013.

14 Mina Ghashami, Edo Liberty, Jeff M. Phillips, and David P. Woodruff. Frequent directions: Simple and deterministic matrix sketching. *CoRR*, abs/1501.01711, 2015. URL: http://arxiv.org/abs/1501.01711.

15 Ashish Goel, Michael Kapralov, and Sanjeev Khanna. On the communication and streaming complexity of maximum bipartite matching. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 468–485, 2012. URL: http://portal.acm.org/citation.cfm?id=2095157&CFID=63838676&CFTOKEN=79617016.

16 Elena Grigorescu, Morteza Monemizadeh, and Samson Zhou. Estimating weighted matchings in o(n) space. *CoRR*, abs/1604.07467, 2016. URL: http://arxiv.org/abs/1604.07467.

17 Venkatesan Guruswami and Krzysztof Onak. Superlinear lower bounds for multipass graph processing. In *Proc. of the 28th Conference on Computational Complexity, CCC 2013, Palo Alto, California, USA, 5-7 June, 2013*, pages 287–298, 2013. doi:10.1109/CCC.2013.37.

18 Michael Kapralov. Better bounds for matchings in the streaming model. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 1679–1697, 2013. doi:10.1137/1.9781611973105.121.

19 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 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 734–751, 2014. doi:10.1137/1.9781611973402.55.

20 Christian Konrad. Maximum matching in turnstile streams. In *Algorithms – ESA 2015 – 23rd Annual European Symposium, Patras, Greece, September 14-16, 2015, Proceedings*, pages 840–852, 2015. doi:10.1007/978-3-662-48350-3_70.

21 Christian Konrad, Frédéric Magniez, and Claire Mathieu. Maximum matching in semi-streaming with few passes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques – 15th International Workshop, APPROX 2012, and 16th International Workshop, RANDOM 2012, Cambridge, MA, USA, August 15-17, 2012. Proceedings*, pages 231–242, 2012. doi:10.1007/978-3-642-32512-0_20.

22 Christian Konrad and Adi Rosén. Approximating semi-matchings in streaming and in two-party communication. In *Automata, Languages, and Programming – 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part I*, pages 637–649, 2013. doi:10.1007/978-3-642-39206-1_54.

23 Madhusudan Manjunath, Kurt Mehlhorn, Konstantinos Panagiotou, and He Sun. Approximate counting of cycles in streams. In *Algorithms – ESA 2011 – 19th Annual European Symposium, Saarbrücken, Germany, September 5-9, 2011. Proceedings*, pages 677–688, 2011. doi:10.1007/978-3-642-23719-5_57.

24 Andrew McGregor. Finding graph matchings in data streams. *APPROX-RANDOM*, pages 170–181, 2005.

25 Andrew McGregor. Graph stream algorithms: a survey. *SIGMOD Record*, 43(1):9–20, 2014. doi:10.1145/2627692.2627694.

26 Mariano Zelke. Weighted matching in the semi-streaming model. *Algorithmica*, 62(1-2):1–20, 2012. doi:10.1007/s00453-010-9438-5.