Approximating Semi-matchings in Streaming and in Two-party Communication

Christian Konrad[1] and Adi Rosén[2]

[1] Reykjavik University, Reykjavik, Iceland, `christiank@ru.is`
[2] CNRS and Univerité Paris Diderot - Paris 7, France, `adiro@liafa.univ-paris-diderot.fr`

Abstract. We study the streaming complexity and communication complexity of approximating unweighted semi-matchings. A semi-matching in a bipartite graph $G = (A, B, E)$ with $n = |A|$ is a subset of edges $S \subseteq E$ that matches all A vertices to B vertices with the goal usually being to do this as fairly as possible. While the term *semi-matching* was coined in 2003 by Harvey et al. [WADS 2003, also Journal of Algorithms 2006], the problem had already previously been studied in the scheduling literature under different names.

We present a deterministic one-pass streaming algorithm that for any $0 \le \epsilon \le 1$ uses space $\tilde{O}(n^{1+\epsilon})$ and computes an $O(n^{(1-\epsilon)/2})$-approximation to the semi-matching problem. Furthermore, with $O(\log n)$ passes it is possible to compute an $O(\log n)$-approximation with space $\tilde{O}(n)$.

In the one-way two-party communication setting, we show that for every $\epsilon > 0$, deterministic communication protocols for computing an $O(n^{\frac{1}{(1+\epsilon)c+1}})$-approximation require a message of size more than cn bits. We present two deterministic protocols communicating n and $2n$ edges that compute an $O(\sqrt{n})$ and an $O(n^{1/3})$-approximation respectively.

Finally, we improve on results of Harvey et al. [Journal of Algorithms 2006] and prove new links between semi-matchings and matchings. While it was known that an optimal semi-matching contains a maximum matching, we show that there is a hierachical decomposition of an optimal semi-matching into maximum matchings. A similar result holds for semi-matchings that do not admit length-two degree-minimizing paths.

1 Introduction

Semi-matchings. A *matching* in an unweighted bipartite graph $G = (A, B, E)$ can be seen as a one-to-one assignment matching the A vertices to B vertices. The usual aim is to find a matching that leaves as few A vertices without associations as possible. A *semi-matching* is then an extension of a matching, in that it is required that *all* A vertices are matched to B vertices. This, however, is generally not possible in an injective way, and therefore we now allow the matching of multiple A vertices to the same B vertex. Typical objectives here are to minimize the maximal number of A vertices that are matched to the same B vertex, or to optimize with respect to even stronger balancing constraints. The term 'semi-matching' was coined by [16] and also used in [11], however, the problem had already previously been intensely studied in the scheduling literature [6, 17, 4, 1, 23]. We stick to this term since it nicely reflects the structural property of entirely matching one bipartition of the graph.

The semi-matching problem captures the problem of assigning a set of unit-length jobs to a set of identical machines with respect to assignment conditions expressed through edges between the two sets. The objective of minimizing the maximal number of jobs that a machine receives then corresponds to minimizing the *makespan* of the scheduling problem. Optimizing the cost function $\sum_{b \in B} \binom{\deg_S(b)+1}{2}$, where $\deg_S(b)$ denotes the number of jobs that a machine b receives in the semi-matching S, corresponds to minimizing the *total completion time* of the jobs (optimizing with respect to this cost function automatically minimizes the maximal degree as well).

Optimality of a Semi-matching. It is well known that matchings are of maximal size if they do not admit *augmenting paths* [5]. Augmenting paths for matchings correspond to

degree-minimizing paths for semi-matchings. They first appeared in [16] under the name of *cost-reducing-paths*, and they were used for the computation of a semi-matching that minimizes a certain cost function. We use the term 'degree-minimizing-path' since it is more appropriate in our setting. A degree-minimizing path starts at a B node of high degree, then alternates between edges of the semi-matching and edges outside the semi-matching, and ends at another B node of smaller degree. Flipping the semi-matching and non-semi-matching edges of the path then generates a new semi-matching such that the large degree of the start node of the path is decreased by 1, and the small degree of the end node of the path is increased by 1. We define an *optimal semi-matching* as one that does not admit any degree-minimizing paths. It was shown in [16] that such a semi-matching is also optimal with respect to a large set of cost functions, including the minimization of the maximal degree as well as the minimization of the total completion time.

Since an optimal semi-matching minimizes many convex cost functions, there is not only one meaningful definition of what an approximation to the semi-matching problem should be. We will consider a notion that has already been used in [4]. We say that an algorithm is a c-approximation algorithm to the semi-matching problem if for any input graph, it outputs a semi-matching S such that $\deg\max S \leq c\deg\max S^*$, where S^* is an optimal semi-matching and $\deg\max T$ denotes the maximal degree of a vertex with respect to the set T. This notion of approximation corresponds to approximating the makespan when the semi-matching is seen as a scheduling problem. This setting was already studied in e.g. [4]. In [9], the semi-matching problem is studied in the distributed setting, and the cost function $\sum_{b\in B}\binom{\deg_S(b)+1}{2}$ is used. These notions are not comparable.

Streaming Algorithms and Communication Complexity. *Streaming Algorithms* fall into the category of massive data set algorithms. In many applications, the data that an algorithm is called upon to process is too large to fit into the computer's memory. In order to cope with this problem, a streaming algorithm sequentially scans the input while using a random access memory of size sublinear in the length of the input stream. Multiple passes often help to further decrease the size of the random access memory. *Graph streams* are widely studied in the streaming model (see [24] for a survey), and in the last years matching problems have received particular attention (e.g. [14, 22, 18, 2, 8, 15, 3, 7, 20, 19, 10]). A graph stream is a sequence of the edges of the input graph with a priori no assumption on the order of the edges. Particular arrival orders of the edges are studied in the literature and allow the design of algorithms that depend on that order. Besides uniform random order [22, 19], the *vertex arrival order* [14, 18] of edges of a bipartite graph is studied where edges incident to the same A node arrive in blocks. Deciding basic graph properties such as connectivity already requires $\Omega(|V|)$ space [12], where V denotes the vertex set of a graph. Many works considering graph streams allow an algorithm to use $\mathrm{O}(|V|\operatorname{polylog}|V|)$ space. This setting is usually called the *semi-streaming* setting.

Space lower bounds for streaming algorithms are often obtained via *Communication Complexity*. There is an inherent link between streaming algorithms and one-way k-party communication protocols. A streaming algorithm for a problem P with space s also serves as a one-way k-party communication protocol for P with communication cost $\mathrm{O}(sk)$. Conversely, a lower bound on the size of any message of such a protocol is also a lower bound on the space requirements of a streaming algorithm. Determining the communication complexity of problems is in itself an important task, however, the previously discussed link to streaming algorithms provides an additional motivation.

Our Contributions. We initiate the study of the semi-matching problem in the streaming and the communication settings. We present a deterministic one-pass streaming algorithm that for any $0 \leq \epsilon \leq 1$ uses space $\tilde{\mathrm{O}}(n^{1+\epsilon})$ and computes an $\mathrm{O}(n^{(1-\epsilon)/2})$ approximation to the semi-

matching problem (**Theorem 1**)[3]. Furthermore, we show that with $O(\log n)$ passes we can compute an $O(\log n)$ approximation with space $\tilde{O}(n)$ (**Theorem 2**).

In the one-way two-party communication setting, we show that for any $\epsilon > 0$, deterministic communication protocols that compute an $O(n^{\frac{1}{(1+\epsilon)c+1}})$ approximation to the semi-matching problem require a message of size at least cn bits (**Theorem 5**). We present two deterministic protocols communicating n and $2n$ edges that compute an $O(\sqrt{n})$ approximation and an $O(n^{1/3})$ approximation, respectively (**Theorem 3**).

While it was known that an optimal semi-matching contains a maximum matching [16], we show that there is a hierarchical decomposition of an optimal semi-matching into maximum matchings (**Lemma 14**). Similarly, we show that semi-matchings that do not admit length-two degree-minimizing paths can be decomposed into maximal matchings (**Lemma 13**). The latter result allows us to prove that the maximal degree of a semi-matching that does not admit a length-two degree-minimizing path is at most $\lceil \log(n + 1) \rceil$ times the maximal degree of an optimal semi-matching (**Theorem 6**).

Related Work on the Semi-matching Problem. The semi-matching problem was firstly studied by Horn [17] and independently by Bruno et al. [6], and both designed an $O(|V|^3)$ algorithm. At present, the best existing algorithm for computing an optimal semi-matching runs in time $O(\sqrt{|V|}|E|\log|V|)$ [11, 13] where $V = A \cup B$. Furthermore, in [13] a randomized algorithm with time complexity $O(|V|^\omega \log^{1+o(1)}|V|)$ is given, where ω is the exponent of the best known matrix multiplication algorithm. Since $\omega \leq 2.38$, this algorithm improves on the $O(\sqrt{|V|}|E|\log|V|)$ time algorithm for dense graphs. To the best of our knowledge, the semi-matching problem has not yet been studied in the streaming setting and the communication setting prior to our work. In the online setting, a $\lceil \log(n) + 1 \rceil$-approximation online algorithm is given in [4], where the maximal degree is approximated. In this model, an A vertex comes in together with its incident edges, and the A vertex has to be matched to a B node immediately and irrevocably. It is shown that the greedy algorithm matching an A node to the B node that currently has the smallest degree is $\lceil \log(n + 1) \rceil$ competitive, and that this result is tight. This algorithm can also be seen as a one-pass $\lceil \log(n + 1) \rceil$ approximation semi-streaming algorithm (meaning $\tilde{O}(n)$ space) for the semi-matching problem when the input stream is in vertex arrival order. Note that our one-pass algorithm does not assume any order on the input sequence, and when allowing $\tilde{O}(n)$ space it achieves an $O(\sqrt{n})$-approximation. Recently, the semi-matching problem was studied in the distributed setting [9]. They show that a 2-approximation to the semi-matching problem can be computed in $O(\Delta^5)$ time, where Δ is the maximal degree in the graph. They consider the notion of approximation with respect to the cost function $\sum_{b \in B} \binom{\deg_S(b)+1}{2}$. It can be shown that their algorithm is a $\lceil \log(n + 1) \rceil$-approximation if the cost function $\deg\max S$ for a semi-matching S is considered.

Techniques. Our streaming algorithms are based on the following greedy algorithm. Fix a maximal degree $d_{\max}$ (for instance $d_{\max} = n^{1/4}$) and greedily add edges to a set S_1 such that the maximal degree of a B node in S_1 does not exceed $d_{\max}$, and the degree of any A node in S_1 is at most 1. This algorithm leaves at most $O(n/d_{\max})$ A vertices unmatched in S_1. To match the yet unmatched vertices, we use a second greedy algorithm that we run in parallel to the first one. We fix a parameter d' appropriately (if $d_{\max} = n^{1/4}$ then we set $d' = n^{1/2}$) and for all vertices $a \in A$ we store arbitrary d' edges incident to a in a set E'. Then, we compute an optimal semi-matching S_2 of the unmatched vertices in S_1 and the B nodes only considering the edges in E'. We prove that such a semi-matching has bounded maximal degree (if $d_{\max} = n^{1/4}$ and $d' = n^{1/2}$ then this degree is $n^{1/4}$). The set $S_1 \cup S_2$ is hence a semi-matching of maximal degree $d_{\max} + \deg\max S_2$ and the space requirement of this algorithm is $\tilde{O}(nd')$. In Section 3

---

[3] We write $\tilde{O}(n)$ to denote $O(n \operatorname{polylog} n)$.

we generalize this idea for any $0 \leq \epsilon \leq 1$ to obtain one-pass algorithms with approximation factors $O(n^{(1-\epsilon)/2})$ using space $\tilde{O}(n^{1+\epsilon})$, and a $\log(n)$-pass algorithm with approximation factor $O(\log n)$ using space $\tilde{O}(n)$.

In the one-way two-party communication setting, the edge set E of a bipartite graph $G = (A, B, E)$ is split among two players, Alice and Bob. Alice sends a message to Bob and Bob outputs a semi-matching of G. Our communication upper bounds make use of what we call a *c-semi-matching skeleton* (or simply *c*-skeleton). A *c*-skeleton of a bipartite graph $G = (A, B, E)$ is a subset of edges $S \subseteq E$ such that for any $A' \subseteq A : \deg \max \mathrm{semi}(A', B, S) \leq c \cdot \deg \max \mathrm{semi}(A', B, E)$ where $\mathrm{semi}(A', B, E')$ denotes an optimal semi-matching between A' and B using edges in E'. We show that if Alice sends a *c*-skeleton S of her subgraph to Bob, and Bob computes an optimal semi-matching using his edges and the skeleton, then the resulting semi-matching is a $c+1$ approximation. We show that there is an $O(\sqrt{n})$-skeleton consisting of n edges, and that there is an $O(n^{1/3})$-skeleton consisting of $2n$ edges. It turns out that an optimal semi-matching is an $O(\sqrt{n})$-skeleton, and we show how an $O(n^{1/2})$-skeleton can be improved to an $O(n^{1/3})$-skeleton by adding additional n edges. These skeletons are almost optimal: we show that for any $\epsilon > 0$, an $O(n^{\frac{1}{(1+\epsilon)c+1}})$-skeleton has at least cn edges. Inspired by the prior lower bound, we prove that for any $\epsilon > 0$, the deterministic one-way two-party communication complexity of approximating semi-matchings within a factor $O(n^{\frac{1}{(1+\epsilon)c+1}})$ is at least cn bits.

In order to prove our structure lemmas on semi-matchings, we make use of degree-minimizing paths. Our results on the decomposition of semi-matchings into maximum and maximal matchings directly relate the absence of degree-minimizing paths to the absence of augmenting paths in matchings. See Section 5 for details.

Organization. After presenting notations and definitions in Section 2, we present our streaming algorithms in Section 3. We then discuss the one-way two-party communication setting in Section 4. We conclude with Section 5, where we present our results on the structure of semi-matchings.

2 Notations and Definitions

Let $G = (A, B, E)$ be a bipartite graph and let $n = |A|$. We assume that the graph does not have isolated A-vertices in order to guarantee that the graph has a semi-matching. Furthermore, we assume that $|B| = \mathrm{poly}(n)$. Let $e \in E$ be an edge connecting nodes $a \in A$ and $b \in B$. Then, we write $A(e)$ to denote the vertex a, $B(e)$ to denote the vertex b, and ab to denote e. Furthermore, for a subset $E' \subseteq E$, we define $A(E') = \bigcup_{e \in E'}\{A(e)\}$ (respectively $B(E')$). For subsets $A' \subseteq A$ and $B' \subseteq B$ we write $E'|_{A' \times B'}$ to denote the subset of edges of E' whose endpoints are all in $A' \cup B'$. We denote by $E'(a)$ the set of edges of $E' \subseteq E$ that have an endpoint in vertex a, and $E'(A')$ the set of edges that have endpoints in vertices of A', where $A' \subseteq A$ (similarly we define $E'(B')$ for $B' \subseteq B$).

For a node $v \in A \cup B$, the *neighborhood* of v is the set of nodes that are adjacent to v and we denote it by $\Gamma(v)$. For a subset $E' \subseteq E$, we write $\Gamma_{E'}(v)$ to denote the neighborhood of v in the graph induced by E'. Note that by this definition $\Gamma(v) = \Gamma_E(v)$. For a subset $E' \subseteq E$, we denote by $\deg_{E'}(v)$ the *degree* in E' of a node $v \in V$, which is the number of edges of E' with an endpoint in v. We define $\deg \max E' := \max_{v \in A \cup B} \deg_{E'}(v)$.

Furthermore, for two sets X, Y, we denote by $X \oplus Y$ the symmetric difference $(X \setminus Y) \cup (Y \setminus X)$.

Definition 1 (Semi-matching). *A* semi-matching *in a bipartite graph* $G = (A, B, E)$ *is a subset* $S \subseteq E$ *such that* $\forall a \in A : \deg_S(a) = 1$.

An important notion for the computation of semi-matchings are *degree-minimizing-paths*.

Definition 2 (Degree-minimizing-path). *A* degree-minimizing path P *with respect to a semi-matching* S *is a path* $P = (b_1, a_1, \ldots, b_{k-1}, a_{k-1}, b_k)$ *of length* $2k$ *($k \geq 1$) such that for all* $i \leq k : (a_i, b_i) \in S$, *for all* $i \leq k-1 : (a_i, b_{i+1}) \notin S$, *and* $\deg_S(b_1) > \deg_S(b_2) \geq \deg_S(b_3) \geq \cdots \geq \deg(b_{k-1}) > \deg(b_k)$.

We define optimality of a semi-matching by means of degree-minimizing-paths.

Definition 3 (Optimal Semi-matching). *An* optimal semi-matching $S^* \subseteq E$ *is a semi-matching that does not admit any degree-minimizing-paths.*

As previously mentioned, it is shown in [16] that an optimal semi-matching also minimizes the maximum degree of a B-node.

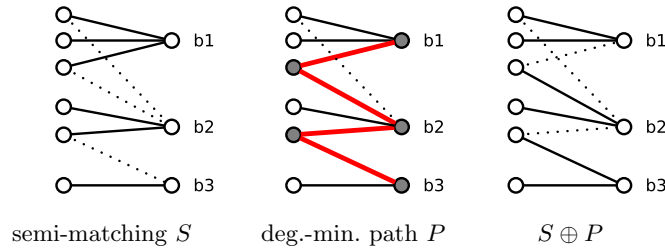The previous notions are illustrated in Figure 1.

semi-matching S deg.-min. path P $S \oplus P$

Fig. 1. Illustration of a semi-matching S. P is a degree-minimizing path of length 4 starting at node b_1 and ending at node b_3. Initially, the degree of b_1 in S is 3 and the degree of b_3 in S is 1. Removing the edges $P \cup S$ from S and inserting the edges $P \setminus S$ into S decreases the degree of b_1 by 1 and increases the degree of b_3 by 1. Here, $S \oplus P$ is an optimal semi-matching.

The SEMI-MATCHING problem consists of computing an optimal semi-matching in a bipartite graph.

For subsets $A' \subseteq A, B' \subseteq B, E' \subseteq E$, we denote by $\text{semi}(A', B', E')$ an optimal semi-matching in the graph $G' = (A', B', E')$, and we denote by $\text{semi}_2(A', B', E')$ a semi-matching that does not admit degree-minimizing paths of length 2 in G'.

Our algorithms for semi-matchings require the notion of *incomplete d-bounded semi-matchings*. These are semi-matchings that do not match all A-vertices and have a bounded maximal degree.

Definition 4 (Incomplete d-bounded Semi-matching). *Let d be an integer. Then an incomplete d-bounded semi-matching of G is a subset $S \subseteq E$ such that $\forall a \in A : \deg_S(a) \leq 1$ and $\forall b \in B : \deg_S(b) \leq d$.*

For subsets $A' \subseteq A, B' \subseteq B, E' \subseteq E$, we write $\text{isemi}_d(A', B', E')$ to denote an incomplete d-bounded semi-matching of maximal size in the graph $G' = (A', B', E')$.

We say that an algorithm (or communication protocol) is a c-approximation algorithm (resp. communication protocol) to SM if it outputs a semi-matching S such that $\deg \max S \leq c \cdot \deg \max S^*$, where S^* denotes an optimal semi-matching. We note that this measure was previously used for approximating semi-matching, e.g, in [4].

3 Streaming Algorithms

To present our streaming algorithms, we describe an algorithm, ASEMI(G, s, d, p) (Algorithm 1), that computes an incomplete $2dp$-bounded semi-matching in the graph G using space $\tilde{O}(s)$,

and makes at most $p \geq 1$ passes over the input stream. If appropriate parameters are chosen, then the output is not only an incomplete semi-matching, but also a semi-matching. We run multiple copies of this algorithm with different parameters in parallel in order to obtain a one-pass algorithm for the semi-matching problem (Theorem 1). Using other parameters, we also obtain a $\log n$-pass algorithm (Theorem 2).

Algorithm 1 Skeleton for approximating semi-matchings: $\textsc{asemi}(G, s, d, p)$

Require: $G = (A, B, E)$ is a bipartite graph
  $S \leftarrow \varnothing$
  repeat at most p times or until $|A(S)| = |A|$
    $S \leftarrow S \cup \textsc{incomplete}(G|_{(A \setminus A(S)) \times B}, s, d)$ {**requires one pass**}
  end repeat
  return S

Algorithm 2 Computing incomplete semi-matchings: $\textsc{incomplete}(G, s, d)$

Require: $G = (A, B, E)$ is a bipartite graph
  $k \leftarrow s/|A|$, $S_1 \leftarrow \varnothing$, $E' \leftarrow \varnothing$
  while $\exists$ an edge ab in stream **do**
    if $ab \notin A \times B$ **then continue**
    if $\deg_{S_1}(a) = 0$ and $\deg_{S_1}(b) < d$ **then** $S_1 \leftarrow S_1 \cup \{ab\}$
    if $\deg_{E'}(a) < k$ **then** $E' \leftarrow E' \cup \{ab\}$
  end while
  $S_2 \leftarrow \text{isemi}_d(E'|_{(A \setminus A(S_1)) \times B})$
  $S \leftarrow S_1 \cup S_2$
  return S

$\textsc{asemi}(G, s, d, p)$ starts with an empty incomplete semi-matching S and adds edges to S by invoking $\textsc{incomplete}(G, s, d)$ (Algorithm 2) on the subgraph of the yet unmatched A vertices in S and all B vertices. Each invocation of $\textsc{incomplete}(G, s, d)$ makes one pass over the input stream and returns a $2d$-bounded incomplete semi-matching while using space $\tilde{\mathrm{O}}(s)$. Since we make at most p passes, the resulting incomplete semi-matching has a maximal degree of at most $2dp$.

$\textsc{incomplete}(G, s, d)$ collects edges greedily from graph G and puts them into an incomplete d-bounded semi-matching S_1 and a set E'. An edge e from the input stream is put into S_1 if $S_1 \cup \{e\}$ is still an incomplete d-bounded semi-matching. An edge $e = ab$ is added to E' if the degree of a in $E' \cup \{e\}$ is less or equal to a parameter k which is chosen to be $s/|A|$ in order to ensure that the algorithm does not exceed space $\tilde{\mathrm{O}}(s)$. The algorithm returns an incomplete $2d$-bounded semi-matching that consists of S_1 and S_2, where S_2 is an optimal incomplete d-bounded semi-matching between the A vertices that are not matched in S_1 and all B vertices, using only edges in E'.

We lower-bound the size of S_2 in Lemma 1. We prove that for any bipartite graph $G = (A, B, E)$ and any $k > 0$, if we store for each $a \in A$ any $\max\{k, \deg_G(a)\}$ incident edges to a, then we can compute an incomplete d-bounded semi-matching of size at least $\min\{kd, |A|\}$ using only those edges, where d is an upper-bound on the maximal degree of an optimal semi-matching between A and B in G.

Lemma 1 is then used in the proof of Lemma 2. In Lemma 2, we apply Lemma 1 in order to obtain a lower bound on the size of S_2 which in turn is used to obtain a lower bound on the size of the output $S_1 \cup S_2$ of $\textsc{incomplete}(G, s, d)$.

Lemma 1. *Let $G = (A, B, E)$ be a bipartite graph, let $k > 0$ and let $d \geq \deg\max\operatorname{semi}(A, B, E)$. Furthermore, let $E' \subseteq E$ be a subset of edges such that for all $a \in A : \deg_{E'}(a) = \min\{k, \deg_E(a)\}$. Then there is an incomplete d-bounded semi-matching $S \subseteq E'$ such that $|S| \geq \min\{kd, |A|\}$.*

Proof. Let $d^* = \deg\max\operatorname{semi}(A, B, E)$. We explicitly construct an incomplete semi-matching S. Let $A_0 \subseteq A$ such that for all $a \in A_0 : \deg_{E'}(a) = \deg_E(a)$, and let $A_1 = A \setminus A_0$. Let $S_0 = \operatorname{semi}(A_0, B, E)$. Clearly, $\deg\max S_0 \leq d^*$. We construct now S as follows.

Start with $S = S_0$, and then add greedily edges in any order from $E'|_{A_1 \times B}$ to S such that S remains an incomplete semi-matching with maximal degree d. Stop as soon as there is no further edge that can be added to S.

We prove that S contains at least $\min\{kd, |A|\}$ edges. To see this, either all nodes of A are matched in S, or there is at least one node $a \in A_1$ that is not matched in S (note that all nodes in A_0 are matched in S). Since $\deg_{E'}(a) = k$, all nodes $b \in \Gamma_{E'}(a)$ have degree d since otherwise a would have been added to S. This implies that there are at least $k \cdot d$ nodes matched in S which proves the lemma. $\square$

Lemma 2. *Let $G = (A, B, E)$ be a bipartite graph, let $s \geq |A|$ and let $d \geq \deg\max\operatorname{semi}(A, B, E)$. Then $\textsc{incomplete}(G, s, d)$ (see Algorithm 2) uses $\tilde{O}(s)$ space and outputs an incomplete $2d$-bounded semi-matching S such that $|S| \geq \min\{|A|\frac{d}{d+d^*} + \frac{ds}{|A|}, |A|\}$, where d^* is the maximum degree in an optimal semi-matching.*

Proof. The proof refers to the variables of Algorithm 2 and the values they take at the end of the algorithm. Furthermore, let $S^* = \operatorname{semi}(A, B, E)$, $d^* = \deg\max S^*$, and let $A' = A \setminus A(S_1)$.

Firstly, we lower-bound $|S_1|$. Let $a \in A'$ and $b = S^*(a)$. Then $\deg_{S_1}(b) = d$ since otherwise a would have been matched in S_1. Hence, we obtain $|A(S_1)| \geq d|B(S^*(A'))| \geq d|A'|/d^*$, where the second inequality holds since the maximal degree in S^* is d^*. Furthermore, since $A' = A \setminus A(S_1)$ and $|S_1| = |A(S_1)|$, we obtain $|S_1| \geq |A|\frac{d}{d+d^*}$. We apply Lemma 1 on the graph induced by the edge set $E'|_{A' \times B}$. We obtain that $|S_2| \geq \min\{ds/|A|, |A'|\}$ and consequently $|S| = |S_1| + |S_2| \geq \min\{|A|\frac{d}{d+d^*} + \frac{ds}{|A|}, |A|\}$.

Concerning space, the dominating factor is the space required for storing the at most $k + 1$ edges for every A vertex. Hence, space is bounded by $\tilde{O}(k|A|) = \tilde{O}(s)$. $\square$

In the proof of Theorem 1, for $0 \leq \epsilon \leq 1$ we show that $\textsc{asemi}(G, n^{1+\epsilon}, n^{(1-\epsilon)/2}d', 1)$ returns a semi-matching if d' is at least the maximal degree of an optimal semi-matching. Using a standard technique, we run $\log(n) + 1$ copies of $\textsc{asemi}$ for all $d' = 2^i$ with $0 \leq i \leq \log n$ and we return the best semi-matching, obtaining a 1-pass algorithm. We use the same idea in Theorem 2, where we obtain a $O(\log n)$ approximation algorithm that makes $\log n$ passes and uses space $\tilde{O}(n)$.

Theorem 1. *Let $G = (A, B, E)$ be a bipartite graph with $n = |A|$. For any $0 \leq \epsilon \leq 1$ there is a one-pass streaming algorithm for $\textsc{Semi-Matching}$ using $\tilde{O}(n^{1+\epsilon})$ space that computes a $4n^{(1-\epsilon)/2}$ approximation.*

Proof. We run $\log(n) + 1$ copies of Algorithm 1 in parallel as follows. For $0 \leq i \leq \lceil \log n \rceil$ let $S_i = \textsc{asemi}(G, n^{1+\epsilon}, n^{(1-\epsilon)/2}2^i, 1)$ and choose among the S_i a semi-matching S_k such that $|S_k| = n$ (this guarantees that S_k is a valid semi-matching), and for any other S_l with $|S_l| = n : \deg\max S_k \leq \deg\max S_l$ (this guarantees that the approximation factor of S_k is best among the computed semi-matchings).

We show now that there is a S_j which is a semi-matching that fulfills the desired approximation guarantee. Let $S^* = \operatorname{semi}(A, B, E)$ and $d^* = \deg\max(S^*)$. Then define j to be such that $d^* \leq 2^j < 2d^*$ and let $d = n^{(1-\epsilon)/2}2^j$. S_j is the output of a call to $\textsc{incomplete}(G, n^{1+\epsilon}, d)$. By Lemma 2, S_j is of size at least $\min\{n\frac{d}{d+d^*} + dn^\epsilon, |A|\}$ which equals $|A|$ for our choice of d.

This proves that all $a \in A$ are matched in S_j. By Lemma 2, $\deg \max S_j \leq 2d$ which is less or equal to $4n^{(1-\epsilon)/2}d^*$. Hence, S_j is a $4n^{(1-\epsilon)/2}$ approximation.

The space requirement is $\log n$ times the space requirement for the computation of a single S_i which is dominated by the space requirements of Algorithm 2. By Lemma 2, this is $\tilde{O}(n^{1+\epsilon})$, and hence the algorithm requires $\tilde{O}(n^{1+\epsilon} \log n) = \tilde{O}(n^{1+\epsilon})$ space. □

Theorem 2. *Let $G = (A, B, E)$ be a bipartite graph with $n = |A|$. There is a $\log n$-pass streaming algorithm for* SEMI-MATCHING *using space $\tilde{O}(n)$ that computes a $4 \log n$ approximation.*

Proof. As in the proof of Theorem 1, we run $\log(n) + 1$ copies of Algorithm 1 in parallel. For $0 \leq i \leq \lceil \log n \rceil$ let $S_i = \text{ASEMI}(G, n, 2^i, \log n)$ and choose among the S_i a semi-matching S_k such that $|S_k| = n$ and for any other S_l with $|S_l| = n$: $\deg \max S_k \leq \deg \max S_l$.

We show now that there is a S_j which is a semi-matching that fulfills the desired approximation guarantee. Let $S^* = \text{semi}(A, B, E)$ and $d^* = \deg \max(S^*)$. Then define j to be such that $d^* \leq 2^j < 2d^*$ and let $d = 2^j$. S_j is the output of a call to $\text{ASEMI}(G, n, d, \log n)$. In each iteration, the algorithm calls $\text{INCOMPLETE}(G', n, d)$, where G' is the subgraph of G of the not yet matched A vertices and the B vertices. By Lemma 2, at least a $\frac{d}{d+d^*} \geq 1/2$ fraction of the unmatched A vertices is matched since $d \geq d^*$, and the maximal degree of the incomplete semi-matching returned by $\text{INCOMPLETE}(G', n, d)$ is at most $2d$. Hence, after $\log n$ iterations, all A vertices are matched. Since $d < 2d^*$ and the algorithm performs at most $\log n$ iterations, the algorithm returns a $4 \log n$ approximation.

Each copy of Algorithm 1 uses space $\tilde{O}(n)$ and since we run $O(\log n)$ the required space is $\tilde{O}(n)$. □

4 One-Way Two-Party Communication

We now consider deterministic one-way two-party protocols which are given a bipartite graph $G = (A, B, E)$ as input, such that $E_1 \subseteq E$ is given to Alice and $E_2 \subseteq E$ is given to Bob. Alice sends a single message to Bob, and Bob outputs a valid semi-matching S for G. A central idea for our upper and lower bounds is what we call a *c-semi-matching skeleton* which we define as follows:

Definition 5. *Given a bipartite graph $G = (A, B, E)$, a c-semi-matching skeleton (or c-skeleton) is a subset of edges $S \subseteq E$ such that $\forall A' \subseteq A$:*

$$\deg \max \text{semi}(A', B, S) \leq c \cdot \deg \max \text{semi}(A', B, E).$$

We show how to construct an $O(\sqrt{n})$-skeleton of size n, and an $O(n^{1/3})$-skeleton of size $2n$. We show that if Alice sends a c-skeleton of her subgraph $G = (A, B, E_1)$ to Bob, then Bob can output a $c + 1$-approximation to the semi-matching problem. Using our skeletons, we thus obtain one-way two-party communication protocols for the semi-matching problem with approximation factors $O(\sqrt{n})$ and $O(n^{1/3})$, respectively (Theorem 3). Then we show that for any $\epsilon > 0$, an $O(n^{\frac{1}{(1+\epsilon)c+1}})$-skeleton requires at least cn edges. This renders our $O(\sqrt{n})$-skeleton and our $O(n^{1/3})$-skeleton tight up to a constant.

4.1 Upper Bound

Firstly, we discuss the construction of two skeletons. In Lemma 5, we show that an optimal semi-matching is an $O(\sqrt{n})$-skeleton, and in Lemma 8, we show how to obtain a $O(n^{1/3})$-skeleton. In these constructions, we use the following key observation: Given a bipartite graph $G = (A, B, E)$,

let $A' \subseteq A$ be such that A' has minimal expansion, meaning that $A' = \arg\min_{A'' \subseteq A} \frac{|\Gamma(A'')|}{|A''|}$. The maximal degree in a semi-matching is then clearly at least $\lceil \frac{|A'|}{|\Gamma(A')|} \rceil$ since all vertices of A' have to be matched to its neighborhood. However, it is also true that the maximal degree of a semi-matching *equals* $\lceil \frac{|A'|}{|\Gamma(A')|} \rceil$. A similar fact was used in [14] for fractional matchings, and also in [21]. We are going to prove this fact in Lemma 4. This proof requires the following technical lemma, Lemma 3.

Lemma 3. *Let $G = (A, B, E)$ be a bipartite graph and let $A' \subseteq A$ such that $|\Gamma(A')| \leq |A'|$. Then:*
$$\forall A'' \subseteq A' : \frac{|\Gamma(A'')|}{|A''|} \geq \frac{|\Gamma(A')|}{|A'|} \Rightarrow \deg\max\mathrm{semi}(A', B, E) \leq \lceil \frac{|A'|}{|\Gamma(A')|} \rceil.$$

Proof. The proof is by contradiction. Let $d = \lceil \frac{|A'|}{|\Gamma(A')|} \rceil$, $S = \mathrm{semi}(A', B, E)$ and suppose that $\deg\max S \geq d+1$. We construct now a set $\tilde{A} \subset A'$ such that $\frac{|\Gamma(\tilde{A})|}{|\tilde{A}|} < \frac{|\Gamma(A')|}{|A'|}$ contradicting the premise of the lemma.

To this end, we define two sequences $(A_i)_i$ with $A_i \subseteq A'$ and $(B_i)_i$ with $B_i \subseteq \Gamma(A')$. Let $b \in \Gamma(A')$ be a node with $\deg_S(b) \geq d+1$ and let $B_1 = \{b\}$. We define

$$A_i = \Gamma_S(B_i),$$
$$B_{i+1} = \Gamma(A_i) \setminus \cup_{j \leq i} B_j. \tag{1}$$

This setting is illustrated in Figure 2. Note that all A_i and all B_i are disjoint. Let k be such that $|A_k| > 0$ and $|A_{k+1}| = 0$. Then we set $\tilde{A} = \bigcup_{i=1}^{k} A_i$.

By construction of the sequence $(B_i)_i$, it is clear that for any $b' \in \cup B_i : \deg_S(b') \geq \deg_S(b) - 1$, since otherwise there is a degree-minimizing path from b to b' contradicting the definition of S. Then, by Equation 1, we obtain for all i that $|A_i| \geq |B_i|(\deg_S(b) - 1)$ which implies that $|A_i| \geq d|B_i|$ since $\deg_S(b) \geq d+1$. Recall that $|A_1| \geq d+1$. We compute

$$\frac{|\Gamma(\tilde{A})|}{|\tilde{A}|} = \frac{|B_1| + \sum_{2 \leq i \leq k} |B_i|}{|A_1| + \sum_{2 \leq i \leq k} |A_i|} \leq \frac{1 + \sum_{2 \leq i \leq k} |B_i|}{(d+1) + \sum_{2 \leq i \leq k} |B_i| d} < \frac{1}{d} \leq \frac{|\Gamma(A')|}{|A'|},$$

and we obtain a contradiction to the premise of the lemma. $\square$
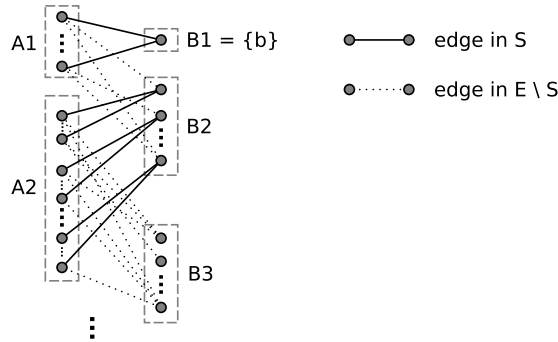

Fig. 2. Illustration of the proof of Lemma 3. All nodes $b' \in \bigcup_{i \geq 2} B_i$ have $\deg_S(b') \geq \deg_S(b) - 1$ since otherwise there is a degree-minimizing path. To keep the figure simple, only those edges of $E \setminus S$ are drawn that connect the A_i to B_{i+1}. Note that in general there are also edges outside S from A_i to $\bigcup_{j < i} B_j$. However, there are no edges in the graph from A_i to $\bigcup_{j \geq i+2} B_j$.

Lemma 4. *Let $G = (A, B, E)$ with $|A| = n$, and let $d = \deg\max \operatorname{semi}(A, B, E)$. Let A' be a subset of A with minimal expansion α, that is*

$$A' = \underset{A'' \subseteq A}{\arg\min} \frac{|\Gamma(A'')|}{|A''|},$$

and let $\alpha = \frac{|\Gamma(A')|}{|A'|}$. Then:

$$d = \lceil \alpha^{-1} \rceil.$$

Proof. We show that $d \geq \lceil \alpha^{-1} \rceil$ and $d \leq \lceil \alpha^{-1} \rceil$ separately.

1. $d \geq \lceil \alpha^{-1} \rceil$**:** The set A' has to be matched entirely to vertices in its neighborhood. Therefore, there is a node $b \in \Gamma(A')$ with degree at least $\lceil \frac{|A'|}{|\Gamma(A')|} \rceil = \lceil \alpha^{-1} \rceil$.

2. $d \leq \lceil \alpha^{-1} \rceil$**:** We construct a semi-matching explicitly with maximal degree d. Since an optimal semi-matching has at most this degree, the claim follows.
Consider a decomposition of A into sets $A_1, A_2, \ldots$ as follows. $A_1 \subseteq A$ is a set with minimal expansion, and for $i > 1$, $A_i \subseteq A \setminus (\bigcup_{j<i} A_j)$ is the set with minimal expansion in $G|_{(A \setminus \bigcup_{j<i} A_j) \times (B \setminus \Gamma(\bigcup_{j<i} A_j))}$.
We construct a semi-matching $\tilde{S} = S_1 \cup S_2 \ldots$ as follows. Firstly, match A_1 to $\Gamma(A_1)$ in S_1. By Lemma 3, the maximal degree in S_1 is at most $\lceil \frac{|A_1|}{|\Gamma(A_1)|} \rceil = \lceil \alpha^{-1} \rceil$.
For a general S_i, we match A_i to vertices in $\Gamma(A_i) \setminus \Gamma(\bigcup_{j<i} A_j)$. By Lemma 3, the maximal degree in S_i is at most $\lceil \frac{|A_i|}{|\Gamma(A_i) \setminus \Gamma(\bigcup_{j<i} A_j)|} \rceil$.
This decomposition is illustrated in Figure 3. Furthermore, it holds

$$\frac{|A_i|}{\Gamma(A_i) \setminus \Gamma(\bigcup_{j<i} A_j)|} \leq \frac{|A_{i+1}|}{\Gamma(A_{i+1}) \setminus \Gamma(\bigcup_{j<i+1} A_j)|},$$

since if this was not true, then the set $A_i \cup A_{i+1}$ would have smaller expansion in the graph $G|_{(A \setminus \bigcup_{j<i} A_j) \times (B \setminus \Gamma(\bigcup_{j<i} A_j))}$ than A_i. This implies that $\deg\max \tilde{S} = \deg\max S_1$ which in turn is $\lceil \alpha^{-1} \rceil$.
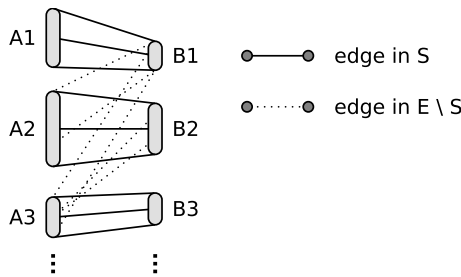
$\square$

Fig. 3. Illustration of the graph decomposition used in the proof of Lemma 4. Here, B_i is the set $\Gamma(A_i) \setminus \Gamma(\bigcup_{j<i} A_j)$. The neighborhood of A_i in G is a subset of $\bigcup_{j \leq i} B_i$. In S, however, A_i is matched entirely to vertices in B_i.

We prove now that an optimal semi-matching is a $O(\sqrt{n})$-skeleton.

Lemma 5. *Let $G = (A, B, E)$ with $n = |A|$, and let $S = \operatorname{semi}(A, B, E)$. Then:*

$$\forall A' \subseteq A : \deg \max \operatorname{semi}(A', B, S) < \sqrt{n} \, (\deg \max \operatorname{semi}(A', B, E))^{1/2} + 1.$$

Proof. Let $A' \subseteq A$ be an arbitrary subset. Let $A'' = \arg \min_{A''' \subseteq A'} \frac{|\Gamma_S(A''')|}{|A'''|}$, and let $k = |\Gamma_S(A'')|$. Let $d = \deg \max \operatorname{semi}(A', B, S)$. Then by Lemma 4, $d = \lceil \frac{|A''|}{k} \rceil$. Furthermore, since A'' is the set of minimal expansion in S, for all $b \in \Gamma_S(A'') : \deg_S(b) = d$, and hence $|A''| = kd$.

Let $d^* = \deg \max \operatorname{semi}(A'', B, E)$. Then $d^* \leq \deg \max \operatorname{semi}(A', B, E)$, since $A'' \subseteq A'$. It holds that $\forall x \in \Gamma_E(A'') \setminus \Gamma_S(A'') : \deg_S(x) \geq d-1$ since otherwise there was a degree-minimizing path of length 2 in S. Figure 4 illustrates this setting. The sum of the degrees of the vertices in $\Gamma_E(A'')$ is upper-bounded by the number of A nodes. We obtain hence $(|\Gamma_E(A'')| - k)(d-1) + kd \leq n$, and this implies that $|\Gamma_E(A'')| \leq \frac{n-k}{d-1}$. Clearly, $d^* \geq |A''| / |\Gamma_E(A'')|$, and using the prior upper bound on $|\Gamma_E(A'')|$ and the equality $|A''| = kd$, we obtain $d^* \geq \frac{kd(d-1)}{n-k}$ which implies that $d < \sqrt{n}\sqrt{d^*} + 1$ for any $k \geq 1$. □
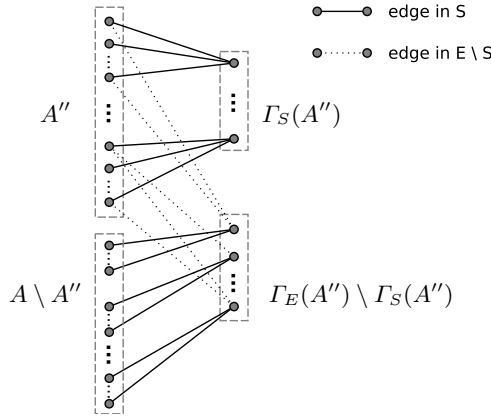

Fig. 4. Illustration of the proof of Lemma 5. All nodes $b \in \Gamma_S(A'')$ have $\deg_S(b) = d$, and all nodes $b' \in \Gamma_E(A'') \setminus \Gamma_S(A'')$ have $\deg_S(b) \geq d - 1$.

In order to obtain an $\mathrm{O}(n^{1/3})$-skeleton, for each $a \in A$ we add one edge to the $\mathrm{O}(\sqrt{n})$-skeleton. Let $S = \operatorname{semi}(A, B, E)$ be the $\mathrm{O}(\sqrt{n})$-skeleton, let $B' = B(S)$ be the B nodes that are matched in the skeleton, and for all $b \in B'$ let $A_b = \Gamma_S(b)$ be the set of A nodes that are matched to b in S. Intuitively, in order to obtain a better skeleton, we have to increase the size of the neighborhood in the skeleton of all subsets of A, and in particular of the subsets A_b for $b \in B'$. We achieve this by adding additional optimal semi-matchings $S_b = \operatorname{semi}(A_b, B, E)$ for all subsets A_b with $b \in B'$ to S, see Lemma 8. We firstly prove a technical lemma, Lemma 6, that points out an important property of the interplay between the matchings S and the matchings S_b for $b \in B'$. Then, we state in Lemma 7 an inequality that is an immediate consequence of Hölder's inequality. Lemma 7 is then used in the proof of Lemma 8, which proves that our construction is an $\mathrm{O}(n^{1/3})$-skeleton.

Lemma 6. *Let $G = (A, B, E)$, $A' \subseteq A$, $A'' \subseteq A'$, and let $S = \operatorname{semi}(A', B, E)$. Furthermore, let $\Gamma_S(A') = \{b_1, \ldots, b_k\}$, and $\forall b_i \in \Gamma_S(A') :$ let $A'_i = \Gamma_S(b_i) \cap A$, and $A''_i = \Gamma_S(b_i) \cap A''$. Then:*

$$\deg \max \operatorname{semi}(A'', B, E)^{-1} \sum_{i : b_i \in \Gamma_S(A'')} |A''_i|(|A'_i| - 1) \leq |A'|.$$

Proof. Let $S'' = \text{semi}(A'', B, E)$, and denote $d = \deg\max S''$. Clearly,

$$\sum_{b'' \in B(S'')} \deg_S(b'') \le |A'|. \tag{2}$$

Consider any $b'' \in B(S'')$. We bound $\deg_S(b'')$ from above as follows

$$\deg_S(b'') \ge \max\{|A_i'| - 1 \ : \ \exists a \in A_i'' \text{ with } b'' \in \Gamma_E(a)\}. \tag{3}$$

Let j be such that $|A_j'| - 1$ poses the maximum of the set in the right hand side of Inequality 3. Note that if Inequality 3 was not true, then there would be a length two degree minimizing path in S connecting b'' and b_j. The setup is visualized in Figure 5. We bound now the right hand side of Inequality 3 as follows

$$
\begin{aligned}
(|A_j'| - 1) &= \max\{|A_i'| - 1 \ : \ \exists a \in A_i'' \text{ with } b'' \in \Gamma_E(a)\} \\
&\ge \sum_{a \in \Gamma_{S''}(b'')} \frac{1}{\deg_{S''}(b'')}(|A_{B(S(a))}'| - 1).
\end{aligned} \tag{4}
$$

We used here that $|A_{B(S(a))}'| \le |A_j'|$ for any $a \in \Gamma_{S''}(b'')$, and $|a \in \Gamma_{S''}(b'')| = \deg_{S''}(b'')$. Since $d = \deg\max S''$, and using Inequalities 3 and 4 we obtain

$$\deg_S(b'') \ge \sum_{a \in \Gamma_{S''}(b'')} \frac{1}{d}(|A_{B(S(a))}'| - 1). \tag{5}$$

We combine Inequalities 2 and 5, and the result follows

$$
\begin{aligned}
|A'| \ge \sum_{b'' \in B(S'')} \deg_S(b'') &\ge \sum_{b'' \in B(S'')} \sum_{a \in \Gamma_{S''}(b'')} \frac{1}{d}(|A_{B(S(a))}'| - 1) \\
&= \frac{1}{d} \sum_{A_i''} |A_i''||A_i' - 1|.
\end{aligned}
$$

$\square$

In the proof of Lemma 8, we also need the following inequality.

Lemma 7. *Let $x_1, \ldots, x_k \ge 0$, and let $p > 0$ be an integer. Then:*

$$\frac{(\sum_{i=1}^k x_i)^p}{k^{p-1}} \le \sum_{i=1}^k x_i^p.$$

Proof. This is an immediate consequence of Hölder's inequality:

$$\sum_{i=1}^k x_i \le (\sum_{i=1}^k x_i^p)^{1/p} k^{\frac{p-1}{p}}.$$

$\square$

Lemma 8. *Let $G = (A, B, E)$ be a bipartite graph with $n = |A|$. Let $S = \text{semi}(A, B, E)$, and for all $b \in B(S) : S_b = \text{semi}(\Gamma_S(b), B, E)$. Then:*

$$\forall A' \subseteq A : \deg\max\text{semi}(A', B, S \cup \bigcup_{b \in B(S)} S_b) \le \lceil 2n^{1/3} \deg\max\text{semi}(A', B, E)\rceil.$$
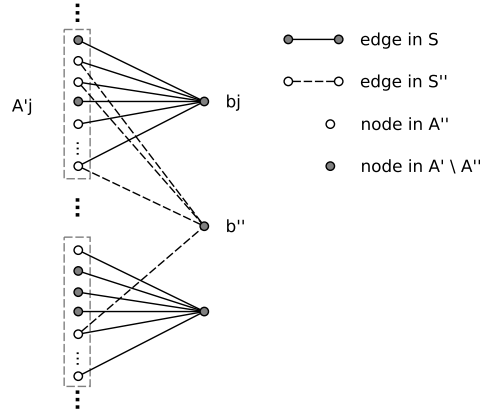
Fig. 5. Illustration of the proof of Lemma 6. The degree of b'' in S is at least $|A'_j| - 1$. Otherwise there would be a length two degree-minimizing path between b'' and b_j.

Proof. Let $A' \subseteq A$. Let $\tilde{S} = S \cup \bigcup_{b \in B(S)} S_b$. Let $A'' = \arg\min_{A''' \subseteq A'} \frac{|\Gamma_{\tilde{S}}(A''')|}{|A'''|}$ and let $k = |\Gamma_{\tilde{S}}(A'')|$. From Lemma 4 it follows that $\deg\max\operatorname{semi}(A', B, \tilde{S}) = \lceil \frac{|A''|}{k} \rceil$. Furthermore, let $d = \deg\max\operatorname{semi}(A'', B, E)$. For a node $b \in \Gamma_{\tilde{S}}(A'')$, let $A''_b = \{a \in A : \tilde{S}(a) = b\}$. For two nodes $b_i, b_j \in \Gamma_{\tilde{S}}(A'')$, let $A''_{b_i, b_j} = \{a \in A'' : S(a) = b_i, S_{b_i}(a) = b_j\}$.

We consider the cases $k \geq n^{1/3}$ and $k < n^{1/3}$ separately.

1. $k \geq n^{1/3}$. Consider the semi-matching S. From Lemma 6 we obtain the condition

$$1/d \sum_{i=1}^{k} |A''_i|(A_i - 1) \leq n,$$

and since $A''_i \leq A_i$ we obtain from the prior Inequality that

$$1/d \sum_{i=1}^{k} (|A''_i| - 1)^2 < n.$$

Using $\sum_{i=1}^{k} |A''_i| = |A''|$ and Lemma 7, we obtain

$$\frac{1}{d}\frac{1}{k}(|A''| - k)^2 < n, \quad \Rightarrow$$
$$|A''| < \sqrt{ndk} + k. \tag{6}$$

Then, since $\deg\max\operatorname{semi}(A'', B, \tilde{S}) = \lceil \frac{|A''|}{k} \rceil$, we obtain from Inequality 6 $\deg\max\operatorname{semi}(A'', B, \tilde{S}) \leq \lceil \frac{\sqrt{nd}}{\sqrt{k}} \rceil + 1$. Since $k \geq n^{1/3}$, we conclude that

$$\deg\max\operatorname{semi}(A'', B, \tilde{S}) \leq n^{1/3}\sqrt{d} + 2.$$

2. $k < n^{1/3}$. We consider here the two subcases $|A''| < 2dk^2$ and $|A''| \geq 2dk^2$.
   (a) $|A''| < 2dk^2$. Then since $\deg\max\operatorname{semi}(A'', B, \tilde{S}) = \lceil \frac{|A''|}{k} \rceil$, we conclude that

$$\deg\max\operatorname{semi}(A'', B, \tilde{S}) \leq \lceil 2dk \rceil < \lceil 2dn^{1/3} \rceil.$$

13

(b) $|A''| \geq 2dk^2$. Let $b \in B(S)$ and consider the semi-matching S_b matching A''_b to B. From Lemma 6 and the fact that $A''_{b,b_i} \subseteq A'_{b,b_i}$ we obtain

$$\frac{1}{d}\sum_{i=1}^{k}|A''_{b,b_i}|(|A''_{b,b_i}|-1) \leq |A_b|,$$

$$\left(\frac{1}{d}\sum_{i=1}^{k}|A''_{b,b_i}|^2\right) - \frac{1}{d}|A''_b| \leq |A_b|.$$

By Lemma 7, we obtain

$$\frac{1}{dk}|A''_b|^2 - \frac{1}{d}|A''_b| \leq |A_b|. \tag{7}$$

Consider now the semi-matching S. From Lemma 6 we obtain the condition

$$\frac{1}{d}\sum_{i=1}^{k}|A''_i|(|A_i|-1) \leq n. \tag{8}$$

Using Inequality 7 in Inequality 8 and simplifying, we obtain

$$\frac{1}{d}\sum_{i=1}^{k}|A''_i|\left(\left(\frac{1}{dk}|A''_i|^2 - \frac{1}{d}|A''_i|\right)-1\right) \leq n,$$

$$\frac{1}{d^2 k}\sum_{i=1}^{k}|A''_i|^3 - \sum_{i=1}^{k}\frac{1}{d^2}|A''_i|^2 - \sum_{i=1}^{k}\frac{1}{d}|A''_i| \leq n,$$

$$\frac{1}{d^2 k^3}|A''|^3 - \underbrace{\frac{1}{d^2 k}|A''|^2}_{I} - \underbrace{\frac{1}{d}|A''|}_{II} \leq n. \tag{9}$$

Since $|A''| \geq 2dk^2$, we can upper bound the terms I and II from Inequality 9 as follows

$$\frac{1}{2d^3 k^3}|A''|^3 \geq I, \text{ and} \tag{10}$$

$$\frac{1}{4d^3 k^4}|A''|^3 \geq II. \tag{11}$$

Using bounds 10 and 11 in Inequality 9 and simplifying, we obtain

$$\frac{1}{4d^2 k^3}|A''|^3 < n, \Rightarrow$$

$$|A''| < 2^{2/3}n^{1/3}d^{2/3}k. \tag{12}$$

Since $\deg\max\text{semi}(A'', B, \tilde{S}) = \lceil\frac{|A''|}{k}\rceil$, and using Inequality 12, we conclude that

$$\deg\max\text{semi}(A'', B, \tilde{S}) \leq \lceil 2^{2/3}n^{1/3}d^{2/3}\rceil.$$

Combining the bounds from cases 1, 2a and 2b, the result follows. $\square$

We mention that there are graphs for which adding further semi-matchings $S_{b_1 b_2} = \text{semi}(A_{b_1 b_2}, B, E)$ to our $O(n^{1/3})$-skeleton, where $A_{b_1 b_2}$ is the set of A vertices whose neighborhood in our $O(n^{1/3})$-skeleton is the set $\{b_1, b_2\}$, does not help to improve the quality of the skeleton. Before stating our main theorem, Theorem 3, we show in Lemma 9 that if Alice sends a c-matching skeleton, then Bob can compute a $c+1$ approximation. Then, we state our main theorem.

Lemma 9. *Let $G = (A, B, E)$ be a bipartite graph and let E_1, E_2 be a partition of the edge set E. Furthermore, let $E_1' \subseteq E_1$ such that for any $A' \subseteq A(E_1)$:*

$$\deg \max \operatorname{semi}(A(E_1), B, E_1') \le c \deg \max \operatorname{semi}(A(E_1), B, E_1').$$

Then:
$$\deg \max \operatorname{semi}(A, B, E_1' \cup E_2) \le (c + 1) \deg \max \operatorname{semi}(A, B, E).$$

Proof. We construct a semi-matching S between A and B with edges from $E_1' \cup E_2$ explicitly and we show that $\deg \max S \le (c+1) \deg \max \operatorname{semi}(A, B, E)$. Since $\deg \max \operatorname{semi}(A, B, E_1' \cup E_2) \le \deg \max S$, the result then follows.

Let $S_2 = \operatorname{semi}(A, B, E) \cap E_2$, and let $S_1 = \operatorname{semi}(A \setminus A(S_2), B, E_1)$. Then $S = S_1 \cup S_2$. Clearly, $\deg \max S_2 \le \deg \max \operatorname{semi}(A, B, E)$. Furthermore, by the premise of the lemma we obtain $\deg \max S_1 \le c \deg \max \operatorname{semi}(A, B, E)$. Since $\deg \max S \le \deg \max S_1 + \deg \max S_2$ and $\deg \max S_1 + \deg \max S_2 \le (c + 1) \deg \max(A, B, E)$ the result follows. □

Theorem 3. *Let $G = (A, B, E)$ with $n = |A|$ and $m = |B|$. Then there are one-way two-party deterministic communication protocols for* Semi-Matching, *one with*

1. *message size $n \log m$ and approximation factor $n^{1/2} + 2$, and another one with*
2. *message size $2n \log m$ and approximation factor $2n^{1/3} + 2$.*

Proof. Alice computes the skeletons as in Lemma 5 or in Lemma 8 and sends them to Bob. Bob computes an optimal semi-matching considering his edges and the edges received from Alice. By Lemma 9 the results follow. □

4.2   Lower Bounds for Semi-matching-skeletons

We present now a lower bound that shows that the skeletons of the previous subsection are essentially optimal. For an integer c, we consider the complete bipartite graph $K_{n,m}$ where m is a carefully chosen value depending on c and n. We show in Lemma 10 that for any subset of edges E' of $K_{n,m}$ such that for all $a \in A : \deg_{E'}(a) \le c$, there is a subset $A' \subseteq A$ with $|A'| \le m$ such that an optimal semi-matching that matches A' using edges in E' has a maximal degree of $\Omega(n^{\frac{1}{c+1}})$. Note that since $|A'| \le m$, there is a matching in $K_{n,m}$ that matches all A' vertices. This implies that such an E' is only an $\Omega(n^{\frac{1}{c+1}})$-skeleton.

Lemma 10. *Let $G = (A, B, E)$ be the complete bipartite graph with $|A| = n$ and $|B| = (c!)^{\frac{1}{c+1}} n^{\frac{1}{c+1}}$ for an integer c. Let $E' \subseteq E$ be an arbitrary subset such that $\forall a \in A : \deg_{E'}(a) \le c$. Then there exists an $A' \subseteq A$ with $|A'| \le |B|$ and*

$$\deg \max \operatorname{semi}(A', B, E') \ge \frac{(c!)^{\frac{1}{c+1}}}{c} n^{\frac{1}{c+1}} > e^{-1.3} n^{\frac{1}{c+1}}. \tag{13}$$

Proof. Let $E' \subseteq E$ be as in the statement of the lemma. Let E'' be an arbitrary superset of E' such that $\forall a \in A : \deg_{E''}(a) = c$. Since $\deg \max \operatorname{semi}(A', B, E'') \le \deg \max \operatorname{semi}(A', B, E')$ it is enough to show the lemma for E''. Denote by $A_{\{i_1, \ldots, i_c\}}$ the subset of A such that $\forall a \in A_{\{i_1, \ldots, i_c\}} : \Gamma_{E''}(a) = \{b_{i_1}, \ldots, b_{i_c}\}$. Then

$$|A| = \sum_{\substack{A_i : i = \{i_1, \ldots, i_c\} \text{ and} \\ \{b_{i_1}, \ldots, b_{i_c}\} \text{ is a } c\text{-subset of } B}} |A_i|, \tag{14}$$

since $\forall a \in A : \deg_{E''}(a) = c$. Suppose for the sake of a contradiction that Inequality 13 is not true. Then for all A_i on the right side of Inequality 14 we have $|A_i| < (c!)^{\frac{1}{c+1}} n^{\frac{1}{c+1}}$. There are at most $\binom{|B|}{c}$ such sets. This implies that:

$$|A| \leq \binom{|B|}{c} \cdot (c!)^{\frac{1}{c+1}} n^{\frac{1}{c+1}} < \frac{|B|^c}{c!} (c!)^{\frac{1}{c+1}} n^{\frac{1}{c+1}} < \frac{(c!)^{\frac{c}{c+1}} n^{\frac{c}{c+1}}}{c!} (c!)^{\frac{1}{c+1}} n^{\frac{1}{c+1}} = n.$$

This is a contradiction to the fact that $|A| \geq n$ and proves the first inequality in Inequality 13. To proof the second, we apply Stirling's formula, and we obtain

$$\frac{(c!)^{\frac{1}{c+1}}}{c} > \frac{(\sqrt{2\pi} c^{c+1/2} e^{-c})^{\frac{1}{c+1}}}{c} = e^{\frac{1/2 \ln(2\pi) - 1/2 \ln(c) - c}{c+1}}.$$

It can be shown that for any $c > 0$, $\frac{1/2 \ln(2\pi) - 1/2 \ln(c) - c}{c+1} > -1.3$ which proves the result. $\square$

We extend Lemma 10 now to edge sets of bounded cardinality without restriction on the maximal degree of an A node, and we state then our lower-bound result in Theorem 4.

Lemma 11. *Let $c > 0$ be an integer, let $\epsilon > 0$ be a constant, and let $c' = (1 + \epsilon)c$. Let $G = (A, B, E)$ be the complete bipartite graph with $|A| = n$ and $|B| = (c'!)^{\frac{1}{c'+1}} (\frac{\epsilon}{1+\epsilon} \cdot n)^{\frac{1}{c'+1}}$. Let $E' \subseteq E$ be an arbitrary subset of size at most $c \cdot n$. Then there exists an $A' \subseteq A$ with $|A'| \leq |B|$ and*

$$\deg \max \operatorname{semi}(A', B, E') > e^{-1.3} \left(\frac{\epsilon}{1+\epsilon} n \right)^{\frac{1}{c'+1}}. \tag{15}$$

Proof. Split A into $A_>$ and $A_\leq$ such that for all $a \in A_> : \deg_{S'}(a) > c'$, and for all $a \in A_\leq : \deg_{S'}(a) \leq c'$. Then $|A_>|c' + |A_\leq| \leq cn$ which implies that $|A_\leq| \geq \frac{\epsilon}{1+\epsilon} n$. Let $G' = G|_{A_\leq \times B}$. Then by Lemma 10 applied on G' there is a subset $A' \subseteq A_\leq$ with $|A'| \leq |B|$ such that

$$\deg \max \operatorname{semi}(A', B, E'|_{A_\leq \times B}) > e^{-1.3} |A_\leq|^{\frac{1}{c'+1}},$$

and since $\deg \max \operatorname{semi}(A', B, E'|_{A_\leq \times B}) = \deg \max \operatorname{semi}(A', B, E')$, the result follows. $\square$

Theorem 4. *Let $c > 0$ be an integer. Then for all $\epsilon > 0$, an $O(n^{\frac{1}{(1+\epsilon)c+1}})$-semi-matching skeleton requires at least cn edges.*

4.3 One-way Two-party Communication Lower Bound

To prove a lower bound on the deterministic communication complexity we define a family of bipartite graphs. For given integers n and m, let $\mathcal{G}_1 = \{G_1(x) | x \in \{0,1\}^{n \times m}\}$ be defined as follows. Let $B_0 = \{b_1^0, \ldots, b_m^0\}$, $B_1 = \{b_1^1, \ldots, b_m^1\}$ and $A = \{a_1, \ldots, a_n\}$. Given $x \in \{0,1\}^{n \times m}$, let $E_x = \{(a_i, b_j^{x_{i,j}}) | 1 \leq i \leq n, 1 \leq j \leq m\}$ (i.e, the entries of the matrix x determine if there is an edge (a_i, b_j^0) or an edge (a_i, b_j^1) for all i, j). Then, we define $G_1(x) = (A, B_0 \cup B_1, E_x)$. From Lemma 11 we immediately obtain the following lemma.

Lemma 12. *Let $c > 0$ be an integer, let $\epsilon > 0$ be a constant, and let $c' = (1 + \epsilon)c$. Let n be a sufficiently large integer, and let $m = (c'!)^{\frac{1}{c'+1}} (\frac{\epsilon}{1+\epsilon} \cdot n)^{\frac{1}{c'+1}}$. Let $G = (A, B_0 \cup B_1, E)$ be a graph $G \in \mathcal{G}_1$, and let $E' \subseteq E$ be such that $|E'| \leq cn$. Then there exists a set of nodes $A' \subseteq A$ with $|A'| \leq m$ and $\deg \max \operatorname{semi}(A', B_0 \cup B_1, E') > 1/2 e^{-1.3} (\frac{\epsilon}{1+\epsilon} n)^{\frac{1}{c'+1}}$.*

We further define a second family of bipartite graphs $\mathcal{G}_2$ on the sets of nodes A and C, $|A| = |C| = n$. For a set $A' \subseteq A$ we define the graph $G_2(A')$ to be an arbitrary matching from all the nodes of A' to nodes of C. The family of graphs $\mathcal{G}_2$ is defined as $\mathcal{G}_2 = \{G_2(A') | A' \subseteq A\}$.

Our lower bound will be proved using a family of graphs $\mathcal{G}$. Slightly abusing notation, the family of graphs $\mathcal{G}$ is defined as $\mathcal{G} = \mathcal{G}_1 \times \mathcal{G}_2$. That is, the graphs in $\mathcal{G}$ are all graphs $G = (A, B_0 \cup B_1 \cup C, E_1 \cup E_2)$ built from a graph $G_1 = (A, B_0 \cup B_1, E_1) \in \mathcal{G}_1$ and a graph $G_2 = (A, C, E_1) \in \mathcal{G}_2$ where the set of nodes A is the same for G_1 and G_2. We now prove our lower bound.

Theorem 5. *Let $c > 0$ be an integer and let $\epsilon > 0$ be an arbitrarily small constant. Let $\mathcal{P}$ be a β-approximation one-way two-party protocol for semi matching that has communication complexity at most α. If $\beta \leq \gamma = 1/2 \frac{1}{e^{1.3}} (\frac{\epsilon}{\epsilon+1} n)^{\frac{1}{(1+\epsilon)c+1}}$, then $\alpha > cn$, where n is the number of nodes to be matched.*

Proof. Take n sufficiently large. Let $c' = (1+\epsilon)c$ and let $m = (c'!)^{\frac{1}{c'+1}} (\frac{\epsilon}{1+\epsilon} \cdot n)^{\frac{1}{c'+1}}$. We consider as possible inputs the graphs in $\mathcal{G}$ (for n and m). Given an input graph, Alice will get as input all edges between A and $B_0 \cup B_1$ (i.e., a graph in $\mathcal{G}_1$) and Bob will get all edges between A and C (i.e., a graph in $\mathcal{G}_2$)

Assume towards a contradiction that the communication complexity of $\mathcal{P}$ is at most cn. Then there is a set of graphs $\mathcal{G}^* \subseteq \mathcal{G}_1$, $|\mathcal{G}^*| \geq 2^{nm-cn}$, such that on all graphs in $\mathcal{G}^*$ Alice sends the same message to Bob. Consider the set $X^* \subseteq \{0,1\}^{n \times m}$ such that $\mathcal{G}^* = \{G_1(x) \mid x \in X^*\}$, Since there is a one-to-one correspondence between $\mathcal{G}^*$ and X^*, $|X^*| \geq 2^{nm-cn}$, and there are at most cn entries which are constant over all matrices in X^*, otherwise $|X^*| < 2^{nm-cn}$. This means that there are at most cn edges that exist in all graphs in $\mathcal{G}^*$. Let E' be the set of all these edges.

Consider now the graph $G = (A, B_0 \cup B_1, E')$. Since $|E'| \leq cn$, by Lemma 12 there exists a set $A' \subseteq A$ with $|A'| \leq m$ and $\deg \max \text{semi}(A', B_0 \cup B_1, E') > \gamma$. We now define $G_2^* \in \mathcal{G}_2$ to be $G_2^* = G_2(A \setminus A')$.

Now observe that on any of $G \in \mathcal{G}^* \times \{G_2^*\} \subseteq \mathcal{G}$, $\mathcal{P}$ gives the same output semi-matching S. S can include, as edges matching the nodes in A', only edges from E', since for any other edge there exists an input in $\mathcal{G}^* \times \{G_2^*\}$ in which that edge does not exist and $\mathcal{P}$ would not be correct on that input. It follows (by Lemma 12) that the maximum degree of S is greater than γ. On the other hand, since $|A'| \leq m$, there is a perfect matching in any graph in $\mathcal{G}^* \times \{G_2^*\}$. The approximation ratio of $\mathcal{P}$ is therefore greater than γ. A contradiction. $\qquad\square$

Finally, the previous communication lower bound immediately implies a lower bound on the space of every one-pass streaming algorithm.

Corollary 1. *For an integer c and an arbitrary small constant $\epsilon > 0$, every one-pass streaming algorithm for the semi-matching problem with approximation factor $\mathrm{O}(n^{\frac{1}{(1+\epsilon)c+1}})$ uses space $\Omega(cn)$.*

5 The Structure of Semi-Matchings

We now present our results concerning the structure of semi-matchings. Firstly, we show in Lemma 13 that a semi-matching that does not admit length 2 degree-minimizing paths can be decomposed into maximal matchings. In Lemma 14, we show that if a semi-matching does not admit *any* degree-minimizing paths, then there is a similar decomposition into maximum matchings.

Lemma 13 is then used to prove that semi-matchings that do not admit length 2 degree-minimizing paths approximate optimal semi-matchings within a factor $\lceil \log(n+1) \rceil$. To this end, we firstly show in Lemma 15 that the first d^* maximal matchings of the decomposition of such a semi-matching match at least $1/2$ of the A vertices, where d^* is the maximal degree of an optimal semi-matching. In Theorem 6, we then apply this result $\lceil \log(n+1) \rceil$ times, showing that the maximal degree of a semi-matching that does not admit length 2 degree-minimizing paths is at most $\lceil \log(n+1) \rceil$ times the maximal degree of an optimal semi-matching.

Lemma 13. *Let $S = \text{semi}_2(A, B, E)$ be a semi-matching in G that does not admit a length 2 degree-minimizing path, and let $d = \deg \max S$. Then S can be partitioned into d matchings $M_1, \ldots, M_d$ such that*

$$\forall i : M_i \text{ is a maximal matching in } G|_{A_i \times B_i},$$

where $A_1 = A$, $B_1 = B$, and for $i > 1 : A_i = A \setminus \bigcup_{1 \le j < i} A(M_j)$ and $B_i = B(M_{j-1})$.

Proof. The matchings $M_1, \ldots, M_d$ can be obtained as follows. For each $b \in B(S)$, label its incident edges in S by $1, 2, \ldots, \deg_S(b)$ arbitrarily. Matching M_i is then the subset of edges of S that are labeled by i.

We prove the statement by contradiction. Let i be the smallest index such that M_i is not maximal in $G|_{A_i \times B_i}$. Then there exists an edge $e = ab \in E$ with $a \in A_i$ and $b \in B_i$ such that $M_i \cup \{e\}$ is a matching in $G|_{A_i \times B_i}$. Note that $\deg_S(b) < i$ since b is not matched in M_i. Consider now the edge $e' \in S$ matching the node a to b' in S. Since $a \in A_i$ and a is not matched in M_i, e' is in a matching M_j with $j > i$ and hence $\deg_S(b') \ge j > i$. Then $P = (b', a, b)$ is a length 2 degree-minimizing path since $\deg_S(b') > i$ and $\deg_S(b) < i$ contradicting our assumption. $\square$

Lemma 14. *Let $S^* = \text{semi}(A, B, E)$ be a semi-matching in G that does not admit degree-minimizing paths of any length, and let $d^* = \deg \max S^*$. Then S^* can be partitioned into d^* matchings $M_1, \ldots, M_{d^*}$ such that*

$$\forall i : M_i \text{ is a maximum matching in } G|_{A_i \times B_i},$$

where $A_1 = A$, $B_1 = B$, and for $i > 1 : A_i = A \setminus \bigcup_{1 \le j < i} A(M_j)$ and $B_i = B(M_{j-1})$.

Proof. The proof is similar to the proof of Lemma 13. The matchings $M_1, \ldots, M_{d^*}$ can be obtained as follows. For each $b \in B(S)$, label its incident edges in S by $1, 2, \ldots, \deg_{S^*}(b)$ arbitrarily. Matching M_i is then the subset of edges of S that are labeled by i.

We prove the statement by contradiction. Let i be the smallest index such that M_i is not a maximum matching in $G|_{A_i \times B_i}$. Then there exists an augmenting path $A = (a_1, b_1, \ldots a_l, b_l)$ such that for all $j < l : (a_{j+1}, b_j) \in M_i$ and $\forall i : (a_i, b_i) \notin M_i$. Let b' be the match of a_1 in S^*. Since $a_1 \in A_l$, $\deg_{S^*}(b') > i$. Since $b_l \in B_i$ and b_l is not matched in M_i^*, $\deg_{S^*}(b_l) < i$. Then $P = (b', a_1, b_1, \ldots, a_l, b_l)$ is a degree-minimizing path contradicting our assumption. $\square$

We firstly prove a lemma that is required in the proof of Theorem 6.

Lemma 15. *Let $A' \subseteq A$, let $S = \text{semi}_2(A', B, E)$ be a semi-matching in $G|_{A' \times B}$ that does not admit length 2 degree-minimizing paths and let $S^* = \text{semi}(A', B, E)$ be an optimal semi-matching in $G|_{A' \times B}$. Then $\exists A'' \subseteq A'$ with $|A''| \ge |A'|/2$ such that*

1. $\deg \max S|_{A'' \times B} \le \deg \max S^*$,
2. $S|_{A' \setminus A'' \times B}$ *is a semi-matching of $G|_{A' \setminus A'' \times B}$ and it does not admit length 2 degree-minimizing paths.*

Proof. Let $d = \deg\max S$ and let $d^* = \deg\max S^*$. Partition S into matchings $M_1, \ldots, M_d$ as in Lemma 13. We will show that $A'' = \bigcup_{i \leq d^*} A(M_i)$ fulfills Item 1 and Item 2 of the Lemma.

We have to show that $|A''| \geq |A'|/2$. Let $A''' = A' \setminus A''$ and let $(a, b) \in S^*$ be an edge such that $a \in A'''$. We argue now, that $\deg_S(b) \geq d^*$.

Suppose for the sake of a contradiction that $\deg_S(b) < d^*$. Then (a, b) could have been added to some matching M_j with $j \leq d^*$. Since by Lemma 13 all M_i are maximal, we obtain a contradiction and this proves that $\deg_S(b) \geq d^*$.

This implies further that $|A''| \geq d^* \cdot |B(S^*|_{A''' \times B})| \geq d^* \cdot |A'''|/d^* = |A'''|$, where the last inequality comes from the fact that a node $b \in B(S^*|_{A''' \times B})$ has at most d^* edges incident in S^*. Since A''' and A'' form a partition of A', we obtain $|A''| \geq |A'|/2$.

Since $A'' = A(S|_{A'' \times B})$ and $S|_{A'' \times B}$ is a set of d^* matchings, Item 1 is trivially true. Concerning Item 2, note that if $S|_{A' \setminus A'' \times B}$ admitted a length 2 degree-minimizing path, then that path would also be a degree-minimizing path in S contradicting the premise that S does not admit a length 2 degree-minimizing path. $\square$

Theorem 6. *Let $S = \mathrm{semi}_2(A, B, E)$ be a semi-matching of G that does not admit a length 2 degree-minimizing path. Let S^* be an optimal semi-matching in G. Then:*

$$\deg\max S \leq \lceil \log(n+1) \rceil \deg\max S^*.$$

Proof. We construct a sequence of vertex sets (A_i) and a sequence of semi-matchings (S_i) as follows. Let $A_1 = A$, and let $S_1 = S$. For any i, S_i will be a semi-matching in the graph $G|_{A_i \times B}$ and it will not admit length 2 degree-minimizing paths.

We construct A_{i+1} and S_{i+1} from A_i and S_i as follows. By Item 1 of Lemma 15, there is a subset $A_i' \subseteq A_i$ of size at least $|A_i|/$ such that $S_i|_{A_i' \times B}$ has maximal degree at most d^*. Let $A_{i+1} = A_i \setminus A_i'$, and let $S_{i+1} = S_i|_{A_{i+1} \times B}$. By Item 2 of Lemma 15, S_{i+1} does not comprise length 2 degree-minimizing paths in the graph $G|_{A_{i+1} \times B}$. We stop this construction at iteration l when $A_l' = A_l$ occurs.

Note that $S = \bigcup_i S_i|_{A_i' \times B}$ and hence $\deg\max S \leq \sum_{i=1}^l \deg\max S_i|_{A_i' \times B} \leq l \cdot d^*$. It remains to argue that $l \leq \log(n) + 1$. Since $|A_i'| \geq |A_i|/2$ and $A_{i+1} = A_i \setminus A_i'$, we have $|A_{i+1}| \leq |A_i|/2$. Since $|A_1| = n$, we have $|A_i| \leq (\frac{1}{2})^{i-1} n$. Then, $|A_{\lceil \log(n+1)\rceil}| < 1$ which implies that $|A_{\lceil \log(n+1)\rceil}| = 0$. We obtain hence $l \leq \lceil \log(n+1) \rceil$, which proves the theorem. $\square$

6 Conclusion and Open Problems

In this paper, we presented a one-pass streaming algorithm for the semi-matching problem that computes a $O(n^{(1-\epsilon)/2})$-approximation using $\tilde{O}(n^{1+\epsilon})$ space, for any $0 \leq \epsilon \leq 1$, and we complemented this algorithm with results in the related one-way two-party communication setting. Specifically, we showed that any one-way two-party communication protocol with approximation factor $O(n^{\frac{1}{(1+\epsilon)c+1}})$ uses a message of size $\Omega(cn)$ bits, for arbitrary small $\epsilon > 0$. Furthermore, we showed that there are essentially tight one-way two-party protocols that compute an $O(n^{1/2})$-approximation and an $O(n^{1/3})$-approximation communicating n and $2n$ edges, respectively. Last, we provided structural results on optimal semi-matchings and on semi-matchings that do not admit length two degree-minimizing paths that establish new ties between the semi-matching problem and the matching problem.

We leave the problem in an interesting state. For $\epsilon = 0$, our streaming algorithm runs in the semi-streaming model (i.e., it uses $\tilde{O}(n)$ space) and computes a $O(\sqrt{n})$-approximation, and we are interested in whether this is best possible. Suppose it is. Then, our one-way two-party communication protocol with approximation factor $O(n^{1/3})$ and communication complexity $\tilde{O}(n)$ shows that in order to prove a space lower bound in the one-way communication setting

that matches the space required by our streaming algorithm, one-way protocols with at least three parties have to be considered. Suppose it is not. As our one-way two-party communication protocol with approximation factor $O(n^{1/3})$ only stores $2n$ edges and is essentially best possible, it provides a characterization of *optimal* edges that a streaming algorithm should retain. This characterization may inspire streaming algorithms with improved space complexity.

Determining the precise trade-off between space complexity and approximation factor for one-pass streaming algorithms is the most intriguing open question.

Acknowledgements

References

1. Abraham, D.: Algorithmics of two-sided matching problems. Master's thesis, University of Glasgow (2003)
2. Ahn, K.J., Guha, S.: Linear programming in the semi-streaming model with application to the maximum matching problem. Inf. Comput. 222, 59–79 (Jan 2013)
3. Ahn, K.J., Guha, S.: Near linear time approximation schemes for uncapacitated and capacitated b-matching problems in nonbipartite graphs. In: Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014. pp. 239–258 (2014)
4. Azar, Y., Naor, J.S., Rom, R.: The competitiveness of on-line assignments. J. Algorithms 18(2), 221–237 (Mar 1995)
5. Berge, C.: Two Theorems in Graph Theory. Proc. of the National Academy of Sciences of the United States of America 43(9), 842–844 (1957), http://www.jstor.org/pss/89875
6. Bruno, J., Coffman, Jr., E.G., Sethi, R.: Scheduling independent tasks to reduce mean finishing time. Commun. ACM 17(7), 382–387 (Jul 1974)
7. Crouch, M., Stubbs, D.M.: Improved Streaming Algorithms for Weighted Matching, via Unweighted Matching. In: Jansen, K., Rolim, J.D.P., Devanur, N.R., Moore, C. (eds.) Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014). Leibniz International Proceedings in Informatics (LIPIcs), vol. 28, pp. 96–104. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2014), http://drops.dagstuhl.de/opus/volltexte/2014/4690
8. Crouch, M., McGregor, A., Stubbs, D.: Dynamic graphs in the sliding-window model. In: Bodlaender, H., Italiano, G. (eds.) Algorithms ESA 2013, Lecture Notes in Computer Science, vol. 8125, pp. 337–348. Springer Berlin Heidelberg (2013)
9. Czygrinow, A., Hanćkowiak, M., Szymańska, E., Wawrzyniak, W.: Distributed 2-approximation algorithm for the semi-matching problem. In: Proceedings of the 26th international conference on Distributed Computing. pp. 210–222. DISC'12, Springer-Verlag, Berlin, Heidelberg (2012)
10. Esfandiari, H., Hajiaghayi, M.T., Liaghat, V., Monemizadeh, M., Onak, K.: 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. pp. 1217–1233. SODA '15, SIAM (2015), http://dl.acm.org/citation.cfm?id=2722129.2722210
11. Fakcharoenphol, J., Laekhanukit, B., Nanongkai, D.: Faster algorithms for semi-matching problems. ACM Trans. Algorithms 10(3), 14:1–14:23 (May 2014)
12. Feigenbaum, J., Kannan, S., McGregor, A., Suri, S., Zhang, J.: Graph distances in the streaming model: the value of space. In: Proc. of the 16th annual ACM-SIAM symposium on Discrete algorithms. SODA '05 (2005), http://dl.acm.org/citation.cfm?id=1070432.1070537
13. Galčík, F., Katrenič, J., Semanišin, G.: On computing an optimal semi-matching. In: Proceedings of the 37th international conference on Graph-Theoretic Concepts in Computer Science. pp. 250–261. WG'11, Springer-Verlag, Berlin, Heidelberg (2011)
14. Goel, A., Kapralov, M., Khanna, S.: On the communication and streaming complexity of maximum bipartite matching. In: Proc. of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '12 (2012), http://dl.acm.org/citation.cfm?id=2095116.2095157
15. Guruswami, V., Onak, K.: Superlinear lower bounds for multipass graph processing. In: Proceedings of the 28th Conference on Computational Complexity, CCC 2013, K.lo Alto, California, USA, 5-7 June, 2013. pp. 287–298 (2013)
16. Harvey, N.J.A., Ladner, R.E., Lovász, L., Tamir, T.: Semi-matchings for bipartite graphs and load balancing. J. Algorithms 59(1), 53–78 (Apr 2006)

17. Horn, W.A.: Minimizing average flow time with parallel machines. Operations Research pp. 846–847 (1973)
18. Kapralov, M.: Better bounds for matchings in the streaming model. In: Proc. of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '13 (2013)
19. Kapralov, M., Khanna, S., Sudan, M.: Approximating matching size from random streams. In: Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 734–751. SODA, SIAM (2014)
20. Kapralov, M., Lee, Y.T., Musco, C., Musco, C., Sidford, A.: Single pass spectral sparsification in dynamic streams. In: 55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014. pp. 561–570 (2014)
21. Kleinberg, J., Rabani, Y., Tardos, É.: Fairness in routing and load balancing. Journal of Computer and System Sciences 63(1), 2–20 (2001)
22. Konrad, C., Magniez, F., Mathieu, C.: Maximum matching in semi-streaming with few passes. In: Proc. of 15th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (2012)
23. Lin, Y., Li, W.: Parallel machine scheduling of machine-dependent jobs with unit-length. European Journal of Operational Research 156(1), 261–266 (July 2004)
24. McGregor, A.: Graph stream algorithms: A survey. SIGMOD Rec. 43(1) (2014)