Maximum Weight b-Matchings in Random-Order Streams

Chien-Chung Huang ✉

CNRS, DI ENS, École normale supérieure, Université PSL, Paris, France

François Sellier ✉

Université Paris Cité, CNRS, IRIF, F-75013, Paris, France
MINES ParisTech, Université PSL, F-75006, Paris, France

Abstract

We consider the maximum weight b-matching problem in the random-order semi-streaming model. Assuming all weights are small integers drawn from $[1, W]$, we present a $2 - \frac{1}{2W} + \varepsilon$ approximation algorithm, using a memory of $O(\max(|M_G|, n) \cdot poly(\log(m), W, 1/\varepsilon))$, where $|M_G|$ denotes the cardinality of the optimal matching. Our result generalizes that of Bernstein [3], which achieves a $3/2 + \varepsilon$ approximation for the maximum cardinality simple matching. When W is small, our result also improves upon that of Gamlath *et al.* [11], which obtains a $2 - \delta$ approximation (for some small constant $\delta \sim 10^{-17}$) for the maximum weight simple matching. In particular, for the weighted b-matching problem, ours is the first result beating the approximation ratio of 2. Our technique hinges on a generalized weighted version of edge-degree constrained subgraphs, originally developed by Bernstein and Stein [5]. Such a subgraph has bounded vertex degree (hence uses only a small number of edges), and can be easily computed. The fact that it contains a $2 - \frac{1}{2W} + \varepsilon$ approximation of the maximum weight matching is proved using the classical Kőnig-Egerváry's duality theorem.

2012 ACM Subject Classification Theory of computation → Streaming, sublinear and near linear time algorithms; Theory of computation → Approximation algorithms analysis

Keywords and phrases Maximum weight matching, b-matching, streaming, random order

Digital Object Identifier 10.4230/LIPIcs.ESA.2022.68

Related Version *Full Version*: `https://arxiv.org/abs/2207.03863`

Funding This work was funded by the grants ANR-19-CE48-0016 and ANR-18-CE40-0025-01 from the French National Research Agency (ANR).

Acknowledgements The authors thank Claire Mathieu and the anonymous reviewers for their helpful comments.

1 Introduction

The maximum weight (b-)matching problem is a classical problem in combinatorial optimization. In this paper we will study a sparsifier for that problem and use it in order to design a streaming algorithm for randomly-ordered streams of edges.

Our main tool is a generalized weighted version of the *edge-degree constrained subgraph* (EDCS), a graph sparsifier originally designed for the maximum matching problem by Bernstein and Stein [5]. Let us first recall the definition an EDCS H of a graph G [5].

▶ **Definition 1** (from [5]). *Let $G = (V, E)$ be a graph, and H a subgraph of G. Given any integer parameters $\beta \geq 2$ and $\beta^- \leq \beta - 1$, we say that H is a (β, β^-)-EDCS of G if H satisfies the following properties:*

(i) *For any edge $(u, v) \in H$,* $\quad \deg_H(u) + \deg_H(v) \leq \beta$

(ii) *For any edge $(u, v) \in G \backslash H$,* $\deg_H(u) + \deg_H(v) \geq \beta^-$.

30th Annual European Symposium on Algorithms (ESA 2022).
Editors: Shiri Chechik, Gonzalo Navarro, Eva Rotenberg, and Grzegorz Herman; Article No. 68; pp. 68:1–68:14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

An EDCS has a size that can be easily controlled by the parameter β and it somehow "balances" the vertex degrees in the graph. A very nice property of this sparsifier is that, for well-chosen values of β and β^-, it always contains a $3/2+\varepsilon$ approximation of the maximum cardinality matching [2, 6]:

▶ **Theorem 2** (from the recent work of Assadi and Bernstein [2]). *Let $0 < \varepsilon < 1/2$. Set $\lambda = \frac{\varepsilon}{32}$. Let $\beta \geq \beta^- + 1$ be integers such that $\beta \geq 8\lambda^{-2}\log(1/\lambda)$ and $\beta^- \geq (1-\lambda)\cdot\beta$. Then any (β, β^-)-EDCS H of a graph G contains a matching M_H such that $\left(\frac{3}{2}+\varepsilon\right)\cdot|M_H| \geq |M_G|$ where M_G denotes the maximum cardinality matching.*

In our paper, we generalize the EDCS in two ways:
- we handle (small) integer-weighted edges;
- we handle the more general case of b-matchings.

To describe our generalization, first let us introduce some notation. A weighted multi-graph $G = (V, E)$ is defined by its set of vertices V and its multi-set of weighted edges E drawn from $V \times V \times \{1, 2, \ldots, W\}$ (*i.e.*, $e = (u, v, k)$ represents an edge between u and v of weight $w(e) = k$). We emphasize that E is a multi-set: not only can there be multiple edges between two vertices, but also some of these edges can have the same weight. We assume that the multi-graph does not contain any self-loop. For a given vertex $v \in V$ and a given subgraph H of G, $\delta_H(v)$ denotes the multi-set of incident edges to v in H, $\deg_H(v)$ the *degree* of v in the multi-graph H and $\mathbf{w}\deg_H(v)$ its *weighted degree* $\sum_{(u,v,w)\in\delta_H(v)} w$ in H. Given a weighted multi-graph $G = (V, E)$ and a set of capacities $b_v \in \mathbb{Z}_+$ associated to each vertex $v \in V$, a multi-set of weighted edges M is called a *b-matching* if for all $v \in V$ the number of edges incident to v in M is smaller than or equal to b_v. For a given subgraph H of G, we denote by M_H an arbitrary maximum weight b-matching included in H. The concept of b-matching encompasses that of matching and allows us to tackle a larger variety of real situations where the vertices have different capacities, *e.g.* [20]. In this paper we will assume that the number of edges between any two vertices u and v is at most $\min(b_u, b_v)$.[1]

▶ **Definition 3.** *Let $G = (V, E)$ be a weighted multi-graph, where E is a multiset of edges drawn from $V \times V \times \{1, 2, \ldots, W\}$, $\{b_v\}_{v\in V}$ be a set of constraints, and H be a subgraph of G. Given any integer parameters $\beta \geq 3$ and $\beta^- \leq \beta - 2$, we say that H is a (β, β^-)-w-b-EDCS of G if H satisfies the following properties:*

(i) *For any edge $(u, v, w_{uv}) \in H$,* $\frac{\mathbf{w}\deg_H(u)}{b_u} + \frac{\mathbf{w}\deg_H(v)}{b_v} \leq \beta \cdot w_{uv}$

(ii) *For any edge $(u, v, w_{uv}) \in G\backslash H$,* $\frac{\mathbf{w}\deg_H(u)}{b_u} + \frac{\mathbf{w}\deg_H(v)}{b_v} \geq \beta^- \cdot w_{uv}$.

An EDCS is a special case of a *weighted b-EDCS* when all the b_vs and all the weights are equal to 1. We can show that such w-b-EDCSes as described in Definition 3 always exist. Moreover, we can also prove that it uses only a reasonable number of edges (up to $2\beta\cdot|M_G|$) and it contains a relatively large weighted b-matching:

▶ **Theorem 4.** *Let $0 < \varepsilon < 1/2$ and let W be an integer parameter. Set $\lambda = \frac{\varepsilon}{100W}$. Let $\beta \geq \beta^- + 2$ be integers such that $\frac{\beta+6W}{\log(\beta+6W)} \geq 2W^2\lambda^{-2}$ and $\beta^- - 6W \geq (1-\lambda)\cdot(\beta+6W)$. Then any (β, β^-)-w-b-EDCS H of a weighted multi-graph G with integer edge weights bounded by W contains a b-matching M_H such that $\left(2 - \frac{1}{2W} + \varepsilon\right)\cdot w(M_H) \geq w(M_G)$.*

[1] This is actually a reasonable assumption as the maximum number of edges that are relevant between two given vertices u and v to construct a b-matching is at most $\min(b_u, b_v)$. This assumption is more debatable in the streaming setting, and this is why we explain how to handle this case in the full version of this paper.

In the full version of this paper, we give a whole class of tight examples reaching the bound of Theorem 4. Compared with the previous results of [2, 6] (Theorem 2) when $W = 1$, we can observe that the approximation ratio is the same, even though the constraints on β and β^- are a bit stricter here. Nonetheless, we can deal with b-matchings as well, even when $W > 1$. As a side note, we note that to satisfy the conditions stated in Theorem 4, it suffices that β is of order $poly(W, 1/\varepsilon)$ for some polynomial.

The *semi-streaming* model of computation [10] has been motivated by the recent rise of massive graphs, where we cannot afford to store the entire input in memory. Given that the graph is made of $|V| = n$ vertices and $|E| = m$ edges, in the semi-streaming model the graph is presented to the algorithm as a stream of edges $e_1, \dots, e_m$. The algorithm is allowed to make a single pass over that stream and can use a memory roughly proportional to the output size (up to a poly-logarithmic factor).

We note that in the most general model where an adversary decides the order of the elements, even for the maximum cardinality simple matching, it is still unclear whether it is possible to beat the approximation ratio of 2.

Our focus here is on the *random-order* streaming model, where the permutation of the edges in the stream is assumed to be chosen uniformly at random. This is a quite reasonable assumption as real-world data have little reason of being ordered in an adversarial way. In fact, as mentioned in [17], the random-order streaming model might better explain why certain algorithms perform better in practice than their theoretical bounds under an adversary model. It is noteworthy that under the random-order streaming model, there are already quite a few evidences to show that it is possible to beat the approximation factor of 2 [1, 3, 11, 17], at least for the simple matching.

Using an adaptation of EDCS, Bernstein [3] obtained a $3/2 + \varepsilon$ approximation in the random-order semi-streaming framework (with probability $1 - 2n^{-3}$ and using $O(n \cdot \log(n) \cdot poly(1/\varepsilon))$ memory). Similarly, we can adapt our w-b-EDCSes to design a semi-streaming algorithm for randomly-ordered streams of weighted edges:

▶ **Theorem 5.** *Let $0 < \varepsilon < \frac{1}{2}$ and let W be an integer parameter. There exists an algorithm that can extract with high probability (at least $1 - 2m^{-3}$) from a randomly-ordered stream of weighted edges having integer weights in $\{1, \dots, W\}$ a weighted b-matching with an approximation ratio of $2 - \frac{1}{2W} + \varepsilon$, using $O(\max(|M_G|, n) \cdot poly(\log(m), W, 1/\varepsilon))$ memory.*

Theorem 5 is the first result for the maximum (integer-weighted) b-matching problem in the random-order semi-streaming framework. For the special case of simple matching, when $W = 1$, we essentially re-capture the result of Bernstein [3] (albeit using slightly more memory). When $W > 1$, we note that prior to our work, Gamlath *et al.* [11] have obtained an approximation ratio of $2 - \delta$ for some small $\delta \sim 10^{-17}$. Our result gives a better approximation when W is reasonably small (but using a memory depending polynomially in W) and we believe that our approach is significantly simpler.

▶ Remark 6. Another generalization of EDCS has been developed by Bernstein *et al.* [4] to maintain a $3/2 + \varepsilon$ approximation of the optimal weighted matching in a dynamic graph. However it is still unknown if their construction can actually lead to an algorithm in the random-order one-pass semi-streaming model [4], or applied to b-matchings.

Technical Overview

To generalize the EDCS to the weighted case, a natural first idea is to build multiple EDCSes, one for each edge-weight from 1 to W, and then take their union. We show in the full version of this paper that such an idea does not lead to a subgraph containing a matching that is better than a 2 approximation.

Our approach is a proper generalization of EDCS, as defined in Definition 3. In Theorem 4, we show that such a w-b-EDCS contains a matching of good approximation ratio. The proof of this theorem is technically the most innovative part of the present work. In order to handle integer-weighted matchings (see Section 2) we make use of Kőnig-Egerváry's duality theorem [7] and a specially-constructed auxiliary graph. The fact that the weights of the edges are integers is critical to get an approximation ratio better than 2 (especially for Claim 13). Then, to handle b-matchings (see Section 3), we build a reduction to simple matchings and show that by doing so we do not lose too much in the approximation ratio.

Regarding Theorem 5, when we design a semi-streaming algorithm to extract a b-matching there is an additional challenge: we do not know in advance the actual size of M_G, which cannot be bounded by n (for instance $|M_G|$ could be of size $n^{1.2}$ or even larger), but we still want to use as little memory as possible, *i.e.*, $O(\max(|M_G|, n) \cdot poly(\log(m), W, 1/\varepsilon))$. We tackle this issue by using a guessing strategy in the early phase of the stream (see Section 4).

Related Work

In the adversarial semi-steaming setting, for the unweighted case, the simple greedy algorithm building a maximal matching provides a 2 approximation, which is the best known approximation ratio. Knowing whether it is possible to achieve a better approximation ratio is a major open question in the field of streaming algorithms. For weighted matchings an approximation ratio of $2 + \varepsilon$ can be achieved [12, 18, 19]. For weighted b-matchings the approximation ratio $2 + \varepsilon$ can also be attained [14]. On the hardness side, we know that an approximation ratio better than $1 + \ln 2 \approx 1.69$ cannot be achieved [15].

In contrast, for the *random-order* stream, a first result was obtained by Konrad, Magniez, and Mathieu [17] with an approximation ratio strictly below 2 for unweighted simple matchings. The approximation ratio was then improved in a sequence of papers [11, 16, 9, 3]. Currently the best result is due to Assadi and Behnezhad [1], who obtained the ratio of $3/2 - \delta$ for some small constant $\delta \sim 10^{-14}$. Regarding weighted simple matchings, Gamlath *et al.* [11] obtained an approximation ratio of $2 - \delta$ for some small constant $\delta \sim 10^{-17}$. Regarding b-matchings, to our knowledge the only result is an approximation ratio of $2 - \delta$ *in expectation* for random-order online matroid intersection by Guruganesh and Singla [13] (hence it applies for unweighted bipartite b-matchings).

2 EDCS for Weighted Matchings

In this section we consider the problem of finding a maximum weight matching in an edge-weighted graph $G = (V, E)$ where the edges have integer weights in $[1, W]$. For ease of presentation, we will use simplified notations for *simple* graphs in this section. Here $w(u, v)$ denotes edge weight between vertices u and v. For a subgraph H of G and a vertex $u \in V$, we denote by $N_H(v)$ the set of vertices adjacent to v in H, by $\deg_H(v)$ the degree of v in H, *i.e.*, $\deg_H(v) = |N_H(v)|$, and by $\mathbf{w}\deg_H(v)$ the weighted degree of v in H, *i.e.*, $\mathbf{w}\deg_H(v) = \sum_{u \in N_H(v)} w(u, v)$. For a subgraph H of G, we will denote by M_H an arbitrary maximum weight matching in H. Then we define the notion of edge-degree constrained subgraphs for weighted graphs (w-EDCS), which in fact is just Definition 3 specialized to the setting in this section.

▶ **Definition 7.** *Let $G = (V, E)$ be a graph with weighted edges, and H be a subgraph of G. Given any integer parameters $\beta \geq 3$ and $\beta^- \leq \beta - 2$, we say that H is a (β, β^-)-w-EDCS of G if H satisfies the following properties:*

(i) *For any edge $(u, v) \in H$,* $\quad \mathbf{w}\deg_H(u) + \mathbf{w}\deg_H(v) \leq \beta \cdot w(u, v)$

(ii) *For any edge $(u, v) \in G \backslash H$,* $\mathbf{w}\deg_H(u) + \mathbf{w}\deg_H(v) \geq \beta^- \cdot w(u, v)$.

Here is a first simple proposition on (β, β^-)-w-EDCS (coming from Property (i)).

▶ **Proposition 8.** *Let H be a (β, β^-)-w-EDCS of a given graph G. Then, for all $v \in V$, we have $\deg_H(v) \leq \beta$.*

Proof. Let $v \in V$. If $N_H(v) = \emptyset$, the stated property is trivial. Otherwise, pick a vertex u such that $w(u,v) = \min_{u' \in N_H(v)} w(u',v)$. Then, by Property (i), $\beta \cdot w(u,v) \geq \mathbf{w}\mathrm{deg}_H(v)$. Therefore, $\deg_H(v) \leq \frac{\mathbf{w}\mathrm{deg}_H(v)}{w(u,v)} \leq \beta$, as any edge incident to v in H has a weight larger than or equal to $w(u,v)$. ◀

We show the existence of w-EDCSes by construction, using a local search algorithm. The following proof closely follows the argument of [2].

▶ **Proposition 9.** *Any graph $G = (V, E)$ with weighted edges contains a (β, β^-)-w-EDCS for any parameters $\beta \geq \beta^- + 2$. Such a (β, β^-)-w-EDCS can be found in $O(\beta^2 W^2 \cdot n)$ local search steps.*

Proof. Start with an empty subgraph H. Then try the following local improvements of H, until it is no longer possible. If there is an edge in H violating Property (i) of Definition 7, then fix that edge by removing it from H. Otherwise, if there is an edge in $G \backslash H$ violating Property (ii), then fix that edge by inserting it in H.

Observe that we give the priority to the correction of violations of Property (i), so that at each step of the algorithm all the vertices have degrees bounded by $\beta + 1$ (as after inserting an edge, Proposition 8 may be violated). To prove that this algorithm terminates in finite time and to show the existence of a w-EDCS, we introduce a potential function:

$$\Phi(H) = (2\beta - 2) \sum_{(u,v) \in H} w(u,v)^2 - \sum_{u \in V} (\mathbf{w}\mathrm{deg}_H(u))^2.$$

As the vertices have degrees bounded by $\beta + 1$ and the edges have weights bounded by W, the value of that potential function is bounded by $2\beta^2 W^2 \cdot n$. Then we can show that after each local improvement step, the value of $\Phi(H)$ increases at least by 2 (see the full version of this paper for details). Therefore, the algorithm terminates in $O(\beta^2 W^2 \cdot n)$ steps. ◀

We also introduce the notion of w-vertex-cover of the edge-weighted graph.

▶ **Definition 10.** *We say that the non-negative integer variables $(\alpha_v)_{v \in V}$ represent a w-vertex-cover of a subgraph H of G if for all $(u,v) \in H$, we have $w(u,v) \leq \alpha_u + \alpha_v$. The sum $\sum_{v \in V} \alpha_v$ is called the weight of the w-vertex-cover.*

To use this data structure for the maximum weight problem, we will use the theorem of Kőnig-Egerváry [7], which is a classic duality theorem.

▶ **Theorem 11** (Kőnig-Egerváry)**.** *In any edge-weighted bipartite subgraph H of G, the maximum weight of a matching equals the smallest weight of a w-vertex-cover.*

This theorem allows us to prove the following lemma, which is technically the most important part of the present work.

▶ **Lemma 12.** *Let $0 < \varepsilon < 1/2$ and W be an integer parameter. For $\beta \geq \beta^- + 2$ integers such that $\frac{\beta}{\beta^-} \leq 1 + \frac{\varepsilon}{5W}$ and $\beta^- \geq \frac{4W}{\varepsilon}$, we have that any (β, β^-)-w-EDCS H of a bipartite graph G (with integer edge weights bounded by W) contains a matching M_H such that $\left(2 - \frac{1}{2W} + \varepsilon\right) \cdot w(M_H) \geq w(M_G)$.*

Proof. Using Kőnig-Egerváry's theorem in the bipartite graph H, we know that there exist integers $(\alpha_v)_{v\in V}$ such that:

- $\sum_{v\in V} \alpha_v = w(M_H)$
- for all $(u,v) \in H$, $w(u,v) \le \alpha_u + \alpha_v$

Now consider the optimal matching M_G in G. The first idea is to use the duality theorem to relate $w(M_G)$ to $w(M_H)$, with a leftover term that will be analyzed in the second part of the proof. We introduce the notion of *good* and *bad* edges:

- the edges $(u,v) \in M_G$ such that $\beta^- \cdot w(u,v) \le \beta \cdot (\alpha_u + \alpha_v)$, which are called *good edges*; the set of good edges is denoted as M_{good};
- the edges $(u,v) \in M_G$ such that $\beta^- \cdot w(u,v) > \beta \cdot (\alpha_u + \alpha_v)$, which are called *bad edges*; the set of bad edges is denoted as M_{bad}.

A key observation is that the edges in $M_G \cap H$ are necessarily good edges by the definition of the w-vertex-cover $(\alpha_v)_{v\in V}$ and the fact that $\beta^- < \beta$. Therefore, the bad edges (u,v) are in $G\backslash H$ and as a consequence they satisfy Property (ii) of Definition 7, *i.e.*, $\beta^- \cdot w(u,v) \le \mathbf{w}\text{deg}_H(u) + \mathbf{w}\text{deg}_H(v)$.

Hence we can write the following:

$$\begin{aligned}
\beta^- \cdot w(M_G) &= \sum_{(u,v)\in M_{good}} \beta^- \cdot w(u,v) + \sum_{(u,v)\in M_{bad}} \beta^- \cdot w(u,v) \\
&\le \sum_{(u,v)\in M_{good}} \beta \cdot (\alpha_u + \alpha_v) + \sum_{(u,v)\in M_{bad}} (\mathbf{w}\text{deg}_H(u) + \mathbf{w}\text{deg}_H(v)) \\
&= \sum_{(u,v)\in M_G} \beta \cdot (\alpha_u + \alpha_v) + \sum_{(u,v)\in M_{bad}} (\mathbf{w}\text{deg}_H(u) + \mathbf{w}\text{deg}_H(v) - \beta \cdot (\alpha_u + \alpha_v)) \\
&\le \beta \cdot w(M_H) + \sum_{(u,v)\in M_{bad}} ((\mathbf{w}\text{deg}_H(u) - \beta \cdot \alpha_u)_+ + (\mathbf{w}\text{deg}_H(v) - \beta \cdot \alpha_v)_+),
\end{aligned}$$

where $(x)_+$ denotes the non-negative part $\max(x,0)$. In the last inequality we also used the fact that $\sum_{(u,v)\in M_G} (\alpha_u + \alpha_v) \le \sum_{v\in V} \alpha_v = w(M_H)$, as each vertex of V is counted at most once in that sum. Now, denoting by V_{bad} the set of vertices which are the endpoints of a bad edge and such that $\mathbf{w}\text{deg}_H(u) - \beta \cdot \alpha_u > 0$, we get

$$\beta^- \cdot w(M_G) \le \beta \cdot w(M_H) + \sum_{v\in V_{bad}} (\mathbf{w}\text{deg}_H(v) - \beta \cdot \alpha_v). \tag{1}$$

Naturally, we want to upper-bound the value of $\sum_{v\in V_{bad}} (\mathbf{w}\text{deg}_H(v) - \beta \cdot \alpha_v)$ and we will do so *via* a specially-constructed graph. Before we describe this graph, we can first easily observe that for any $v \in V_{bad}$, for any $u \in N_H(v)$, we have $w(u,v) \ge \frac{\mathbf{w}\text{deg}_H(v)}{\beta} > \alpha_v$ (by Property (i) of Definition 7 and the definition of V_{bad}); moreover, as $(\alpha_v)_{v\in V}$ is a w-vertex-cover of H, we obtain that $\alpha_u > 0$. These observations will be useful in the following.

The new graph is $H_{bad} = (V_{bad} \cup \tilde{V}, \tilde{E})$. The vertices in H_{bad} are the vertices of V_{bad} as well as copies of the vertices of V such that $\alpha_v > 0$, *i.e.*, $\tilde{V} = \{\tilde{v} : v \in V, \alpha_v > 0\}$. We build the set of edges $\tilde{E}$ as follows. For each $v \in V_{bad}$, for each $u \in N_H(v)$, we create in $\tilde{E}$ an edge $(v, \tilde{u})$ such that $w(v,\tilde{u}) = w(v,u) - \alpha_v$ (note that if u is also in V_{bad}, then $\tilde{E}$ will also contain another edge $(u,\tilde{v})$ such that $w(u,\tilde{v}) = w(u,v) - \alpha_u$). Note that $w(v,\tilde{u}) \in \mathbb{Z}_{>0}$, since $w(u,v) > \alpha_v$ as observed above. Therefore the graph H_{bad} still has non-negative integer-valued edge weights. We next remove some edges from $\tilde{E}$: while there exists a vertex $v \in V_{bad}$ such that $\mathbf{w}\text{deg}_{H_{bad}}(v) > \mathbf{w}\text{deg}_H(v) - \beta \cdot \alpha_v + W$, we pick an arbitrary edge $(v,\tilde{u}) \in \tilde{E}$ incident to v and remove it from H_{bad}. This process guarantees the following property:

$$\forall v \in V_{bad}, \mathbf{wdeg}_H(v) - \beta \cdot \alpha_v \leq \mathbf{wdeg}_{H_{bad}}(v) \leq \mathbf{wdeg}_H(v) - \beta \cdot \alpha_v + W. \tag{2}$$

This finishes the description of the graph H_{bad}. By (2), for any $(v, \tilde{u}) \in \tilde{E}$ we have:

$$\begin{aligned}\beta \cdot w(v, \tilde{u}) + W = \beta \cdot (w(v,u) - \alpha_v) + W &\geq \mathbf{wdeg}_H(v) - \beta \cdot \alpha_v + W + \mathbf{wdeg}_H(u) \\ &\geq \mathbf{wdeg}_{H_{bad}}(v) + \mathbf{wdeg}_{H_{bad}}(\tilde{u}).\end{aligned}$$

Summing this inequality over all the edges in $\tilde{E}$ we obtain:

$$\begin{aligned}\beta \cdot w(\tilde{E}) + W \cdot |\tilde{E}| &\geq \sum_{(v,\tilde{u}) \in \tilde{E}} (\mathbf{wdeg}_{H_{bad}}(v) + \mathbf{wdeg}_{H_{bad}}(\tilde{u})) \\ &= \sum_{v \in V_{bad}} \deg_{H_{bad}}(v) \cdot \mathbf{wdeg}_{H_{bad}}(v) + \sum_{\tilde{u} \in \tilde{V}} \deg_{H_{bad}}(\tilde{u}) \cdot \mathbf{wdeg}_{H_{bad}}(\tilde{u}) \\ &\geq \sum_{v \in V_{bad}} \frac{\mathbf{wdeg}_{H_{bad}}(v)}{W} \cdot \mathbf{wdeg}_{H_{bad}}(v) + \sum_{\tilde{u} \in \tilde{V}} \frac{\mathbf{wdeg}_{H_{bad}}(\tilde{u})}{\alpha_u} \cdot \mathbf{wdeg}_{H_{bad}}(\tilde{u}) \\ &= \sum_{v \in V_{bad}} \frac{(\mathbf{wdeg}_{H_{bad}}(v))^2}{W} + \sum_{\tilde{u} \in \tilde{V}} \frac{(\mathbf{wdeg}_{H_{bad}}(\tilde{u}))^2}{\alpha_u} \\ &\geq \sum_{v \in V_{bad}} \frac{1}{W} \cdot \left(\frac{w(\tilde{E})}{|V_{bad}|}\right)^2 + \sum_{\tilde{u} \in \tilde{V}} \frac{1}{\alpha_u} \cdot \left(\frac{w(\tilde{E}) \cdot \alpha_u}{\sum_{\tilde{u}' \in \tilde{V}} \alpha_{u'}}\right)^2 \\ &= \frac{w(\tilde{E})^2}{W \cdot |V_{bad}|} + \frac{w(\tilde{E})^2}{\sum_{\tilde{u}' \in \tilde{V}} \alpha_{u'}}.\end{aligned}$$

The second inequality comes from the fact that the degree of a vertex can be lower-bounded by the weighted degree of that vertex divided by the weight of the largest edge incident to it (for $v \in V_{bad}$ this weight is W, and for $\tilde{u} \in \tilde{V}$ it is α_u, as $w(v, \tilde{u}) = w(v,u) - \alpha_v \leq \alpha_u$ for v adjacent to $\tilde{u}$ in H_{bad}). The third inequality comes from the minimization of the function over the constraints $\sum_{v \in V_{bad}} \mathbf{wdeg}_{H_{bad}}(v) = \sum_{\tilde{u} \in \tilde{V}} \mathbf{wdeg}_{H_{bad}}(\tilde{u}) = w(\tilde{E})$. Now observing that $|\tilde{E}| \leq w(\tilde{E})$, we derive the following:

$$\beta + W \geq \frac{w(\tilde{E})}{W \cdot |V_{bad}|} + \frac{w(\tilde{E})}{\sum_{\tilde{u} \in \tilde{V}} \alpha_u}. \tag{3}$$

The following claim will help us lower bound the average weighted degree of the vertices of V_{bad} in H_{bad}, namely, $w(\tilde{E})/|V_{bad}|$. For this part it is crucial that the weights are integers.

▷ Claim 13. For all $(u,v) \in M_{bad}$, $(\mathbf{wdeg}_H(u) - \beta \cdot \alpha_u)_+ + (\mathbf{wdeg}_H(v) - \beta \cdot \alpha_v)_+ \geq \frac{\beta^-}{1+\varepsilon/4}$

Proof. We proceed by contradiction. Suppose that there exists $(u,v) \in M_{bad}$ such that $(\mathbf{wdeg}_H(u) - \beta \cdot \alpha_u)_+ + (\mathbf{wdeg}_H(v) - \beta \cdot \alpha_v)_+ < \frac{\beta^-}{1+\varepsilon/4}$. Then, as $\beta^- \cdot w(u,v) \leq \beta \cdot (\alpha_u + \alpha_v) + (\mathbf{wdeg}_H(u) - \beta \cdot \alpha_u)_+ + (\mathbf{wdeg}_H(v) - \beta \cdot \alpha_v)_+$, it means that

$$\beta \cdot (\alpha_u + \alpha_v) < \beta^- \cdot w(u,v) < \beta \cdot (\alpha_u + \alpha_v) + \frac{\beta^-}{1+\varepsilon/4},$$

and therefore by dividing by β^- we obtain

$$\frac{\beta}{\beta^-} \cdot (\alpha_u + \alpha_v) < w(u,v) < \frac{\beta}{\beta^-} \cdot (\alpha_u + \alpha_v) + \frac{1}{1+\varepsilon/4}.$$

As $(\alpha_u + \alpha_v) \in \{0, 1, \ldots, W\}$ (recall that (u, v) is a bad edge) and because $\frac{\beta}{\beta^-} \leq 1 + \frac{\varepsilon}{5W} < 1 + \frac{\varepsilon}{4W\cdot(1+\varepsilon/4)}$, there cannot be any integer in the open interval

$$\left]\frac{\beta}{\beta^-} \cdot (\alpha_u + \alpha_v),\ \frac{\beta}{\beta^-} \cdot (\alpha_u + \alpha_v) + \frac{1}{1+\varepsilon/4}\right[,$$

implying that $w(u, v)$, which is an integer, cannot exist. The proof follows. $\lhd$

Recall that u of $(u, v) \in M_{bad}$ is part of V_{bad} only if $\mathbf{w}\mathrm{deg}_H(u) - \beta \cdot \alpha_u > 0$. Claim 13 then implies that given $(u, v) \in M_{bad}$, if both u and v are in V_{bad}, then $\mathbf{w}\mathrm{deg}_{H_{bad}}(u) + \mathbf{w}\mathrm{deg}_{H_{bad}}(v) \geq \frac{\beta^-}{1+\varepsilon/4}$; if only u is in V_{bad}, then $\mathbf{w}\mathrm{deg}_{H_{bad}}(u) \geq \frac{\beta^-}{1+\varepsilon/4}$. We can thus infer that $\frac{w(\tilde{E})}{|V_{bad}|} \geq \frac{\beta^-}{2\cdot(1+\varepsilon/4)}$ and we can rewrite (3) as $\beta + W \geq \frac{\beta^-}{2W\cdot(1+\varepsilon/4)} + \frac{w(\tilde{E})}{\sum_{\tilde{u}\in\tilde{V}} \alpha_u}$, and therefore

$$\left(\beta + W - \frac{\beta^-}{2W\cdot(1+\varepsilon/4)}\right) \cdot \sum_{\tilde{u}\in\tilde{V}} \alpha_u \geq w(\tilde{E}). \tag{4}$$

We now can rebound the expression of (1) as follows:

$$\begin{aligned}
\beta^- \cdot w(M_G) &\leq \beta \cdot w(M_H) + \sum_{v\in V_{bad}} (\mathbf{w}\mathrm{deg}_H(v) - \beta \cdot \alpha_v) \\
&\leq \beta \cdot w(M_H) + \sum_{v\in V_{bad}} \mathbf{w}\mathrm{deg}_{H_{bad}}(v) && \text{by (2)} \\
&\leq \beta \cdot w(M_H) + w(\tilde{E}) \\
&\leq \left(2\beta + W - \frac{\beta^-}{2W\cdot(1+\varepsilon/4)}\right) \cdot w(M_H). && \text{by (4) and } \sum_{\tilde{u}\in\tilde{V}} \alpha_u \leq w(M_H)
\end{aligned}$$

Re-arranging,

$$\left(2\frac{\beta}{\beta^-} + \frac{W}{\beta^-} - \frac{1}{2W\cdot(1+\varepsilon/4)}\right) \cdot w(M_H) \geq w(M_G).$$

As $\frac{\beta}{\beta^-} \leq 1 + \varepsilon/4$ and $\beta^- \geq \frac{4W}{\varepsilon}$ we obtain the desired result. ◀

Then we can generalize this result to non-bipartite graphs.

▶ **Theorem 14.** *Let $0 < \varepsilon < 1/2$ and W be an integer parameter. Set $\lambda = \frac{\varepsilon}{100W}$. For $\beta \geq \beta^- + 2$ integers such that $\frac{\beta}{\log(\beta)} \geq 2W^2\lambda^{-2}$ and $\beta^- \geq (1-\lambda)\cdot\beta$, we have that any (β, β^-)-w-EDCS H of a graph G (with integer edge weights bounded by W) contains a matching M_H such that $\left(2 - \frac{1}{2W} + \varepsilon\right) \cdot w(M_H) \geq w(M_G)$.*

Proof. The proof of this theorem relies on Lemma 12 and on the same construction as the one in [2], using the probabilistic method and Lovasz Local Lemma [8]. We provide the details of the proof in the full version of this paper. ◀

3 EDCS for Weighted b-Matchings

From now on we consider the problem of finding a maximum weight b-matching in an edge-weighted multi-graph $G = (V, E)$. Hence we will use the notations described in the introduction. Here we recall the generalization of edge-degree constrained subgraphs (EDCS) to an edge-weighted multi-graph $G = (V, E)$ in the context of the b-matching problem.

▶ **Definition 15.** *Let $G = (V, E)$ be a weighted multi-graph, where E is a multiset of edges drawn from $V \times V \times \{1, 2, \ldots, W\}$, $\{b_v\}_{v \in V}$ be a set of constraints, and H be a subgraph of G. Given any integer parameters $\beta \geq 3$ and $\beta^- \leq \beta - 2$, we say that H is a (β, β^-)-w-b-EDCS of G if H satisfies the following properties:*

(i) *For any edge $(u, v, w_{uv}) \in H$, $\frac{\mathbf{w}\deg_H(u)}{b_u} + \frac{\mathbf{w}\deg_H(v)}{b_v} \leq \beta \cdot w_{uv}$*

(ii) *For any edge $(u, v, w_{uv}) \in G \backslash H$, $\frac{\mathbf{w}\deg_H(u)}{b_u} + \frac{\mathbf{w}\deg_H(v)}{b_v} \geq \beta^- \cdot w_{uv}$.*

As for a w-EDCS (Proposition 8), we can bound the degree of a vertex in a w-b-EDCS H (with almost the same proof as that of Proposition 8).

▶ **Proposition 16.** *Let H be a (β, β^-)-w-b-EDCS of a given graph G. Then, for all $v \in V$, we have $\deg_H(v) \leq \beta \cdot b_v$.*

▶ **Proposition 17.** *Let H be a (β, β^-)-w-b-EDCS of a given graph G. Then H contains at most $2\beta \cdot |M_G|$ edges.*

Proof. A vertex $v \in V$ is called *saturated* by M_G if $|\delta_G(v) \cap M_G| = b_v$. We denote by V_{sat} the set of vertices saturated by M_G. As M_G is a maximal matching in G, it means that for all $(u, v, w_{uv}) \in G \backslash M_G$, either u or v is in V_{sat}. We denote by $M_{sat} \subseteq M_G$ the subset of edges in M_G that are incident to a vertex of V_{sat}. By this definition, we get:

$$|H| = |H \cap (M_G \backslash M_{sat})| + |H \backslash (M_G \backslash M_{sat})| \leq |M_G| - |M_{sat}| + \sum_{v \in V_{sat}} \deg_H(v)$$

$$\leq |M_G| - |M_{sat}| + \sum_{v \in V_{sat}} \beta \cdot b_v \leq |M_G| - |M_{sat}| + 2 \cdot |M_{sat}| \cdot \beta \leq 2\beta \cdot |M_G|,$$

as for all $v \in V$, $\deg_H(v) \leq \beta \cdot b_v$ and $\sum_{v \in V_{sat}} b_v \leq 2 \cdot |M_{sat}|$. ◀

We can also show that such w-b-EDCSes always exist.

▶ **Proposition 18.** *Any multi-graph $G = (V, E)$, along with a set of constraints $\{b_v\}_{v \in V}$, contains a (β, β^-)-w-b-EDCS for any parameters $\beta \geq \beta^- + 2$. Such a (β, β^-)-w-b-EDCS can also be found in $O(\beta^2 W^2 \cdot |M_G|)$ local search steps.*

Proof. As in the proof of Proposition 9, we follow closely the argument of [2]. We use the same local-search algorithm as the one in Proposition 9, except that the properties violated are those of Definition 15. Here we also give the priority to the correction of violations of Property (i), so that the at each step of the algorithm all the vertices $v \in V$ have degrees bounded by $\beta \cdot b_v + 1$. To prove that this algorithm terminates and show the existence of a w-b-EDCS, we introduce the following potential function:

$$\Phi(H) = (2\beta - 2) \sum_{(u,v,w_{uv}) \in H} w_{uv}^2 - \sum_{u \in V} \frac{(\mathbf{w}\deg_H(u))^2}{b_u}.$$

Observe that because of Proposition 17, the value of $\Phi(H)$ is bounded by $2\beta W^2 \cdot 2\beta \cdot |M_G|$. We can also show that after each local improvement, the value of $\Phi(H)$ increases by at least $3/2$ (see full version for details). Hence the algorithm terminates in $O(\beta^2 W^2 \cdot |M_G|)$ steps. ◀

The main interest of these w-b-EDCSes is that they contain an (almost) $2 - \frac{1}{2W}$ approximation, as in the case of w-EDCSes in simple graphs (Theorem 14).

▶ **Theorem 4.** *Let $0 < \varepsilon < 1/2$ and let W be an integer parameter. Set $\lambda = \frac{\varepsilon}{100W}$. Let $\beta \geq \beta^- + 2$ be integers such that $\frac{\beta+6W}{\log(\beta+6W)} \geq 2W^2\lambda^{-2}$ and $\beta^- - 6W \geq (1-\lambda)\cdot(\beta+6W)$. Then any (β, β^-)-w-b-EDCS H of a weighted multi-graph G with integer edge weights bounded by W contains a b-matching M_H such that $\left(2 - \frac{1}{2W} + \varepsilon\right)\cdot w(M_H) \geq w(M_G)$.*

Proof. Consider a maximum weight b-matching M_G. We will build from H and M_G two simple graphs $G' = (V', E')$ and $H' = (V', E'_H)$. The set of vertices V' contains, for each vertex $v \in V$, b_v vertices $v_1, \ldots, v_{b_v}$, so that V' contains $\sum_{v\in V} b_v$ vertices in total. To construct E', for each $v \in V$, we will distribute the edges of $\delta(v) \cap (H \cup M_G)$ among the b_v vertices $v_1, \ldots, v_{b_v}$ in such a way so that the following three properties hold:

(i) each v_i has a most one edge of M_G incident to it;

(ii) G' is a simple graph;

(iii) each v_i has a weighted degree in the interval $\left[\frac{\mathbf{w}\mathrm{deg}_H(v)}{b_v} - 2W, \frac{\mathbf{w}\mathrm{deg}_H(v)}{b_v} + 3W\right]$.

The existence of such a distribution is achieved by a greedy procedure (see the full version of this paper for a proof of this fact). For Property (ii), it is crucial that the graph G has at most $\min(b_u, b_v)$ edges between any vertices u and v. This property is important in the proof of Theorem 14 (where negative association is used). Then, for H', we just consider the restriction of G' to the edges corresponding to H (ignoring those from $M_G \backslash H$ in the preceding construction).

Observe that M_G corresponds a simple matching in G', and that any simple matching in H' corresponds to a b-matching in H. Next we show that H' is a $(\beta + 6W, \beta^- - 6W)$-EDCS for simple graph G'. Consider an edge $(u_i, v_j) \in H'$. It corresponds to an edge (u, v, w_{uv}) of H so $\mathbf{w}\mathrm{deg}_{H'}(u_i) + \mathbf{w}\mathrm{deg}_{H'}(v_j) \leq \frac{\mathbf{w}\mathrm{deg}_H(u)}{b_u} + \frac{\mathbf{w}\mathrm{deg}_H(v)}{b_v} + 6W \leq (\beta + 6W)\cdot w_{uv}$, so Property (i) of Definition 7 holds. Consider next an edge $(u_i, v_j) \in G' \backslash H'$. It corresponds to an edge (u, v, w_{uv}) of $M_G \backslash H$, so $\mathbf{w}\mathrm{deg}_{H'}(u_i) + \mathbf{w}\mathrm{deg}_{H'}(v_j) \geq \frac{\mathbf{w}\mathrm{deg}_H(u)}{b_u} + \frac{\mathrm{deg}_H(v)}{b_v} - 6W \geq (\beta^- - 6W)\cdot w_{uv}$ (as there can be a difference of at most W between the weighted degree of u in G' and in H'). Thus Property (ii) of Definition 7 holds as well. To conclude, H' is a $(\beta + 6W, \beta^- - 6W)$-$w$-EDCS of G', so by Theorem 14 we have that $(2 - \frac{1}{2W} + \varepsilon)\cdot w(M_{H'}) \geq w(M_{G'}) = w(M_G)$. As $w(M_H) \geq w(M_{H'})$ (because any matching in H' corresponds to a b-matching of the same weight in H), completing the proof. ◀

4 Application to b-Matchings in Random-Order Streams

In this section we consider the random-order semi-streaming model and we show how our results in the preceding section can be adapted to get a $2 - \frac{1}{2W} + \varepsilon$ approximation.

As our algorithm builds on that of Bernstein [3] for the unweighted simple matching, let us briefly summarize his approach. In the first phase of the streaming, he constructs a subgraph that satisfies only a weaker definition of EDCS in Definition 1 (only Property (i) holds). In the second phase of the streaming, he collects the "underfull" edges, which are those edges that violate Property (ii). He shows that in the end, the union of the subgraph built in the first phrase and the underfull edges collected in the second phase, with high probability, contains a $3/2 + \varepsilon$ approximation and that the total memory used is in the order of $O(n \cdot \log n)$. As we will show below, this approach can be adapted to our context of edge-weighted b-matching. Our main technical challenge lies in the fact that unlike the simple matching, the size of M_G can vary a lot. We need a "guessing" strategy to ensure that the required memory is proportional to $|M_G|$.

▶ **Definition 19.** *We say that a graph H has bounded weighted edge-degree β if for every edge $(u, v, w_{uv}) \in H$, $\frac{\mathbf{w}\mathrm{deg}_H(u)}{b_u} + \frac{\mathbf{w}\mathrm{deg}_H(v)}{b_v} \leq \beta \cdot w_{uv}$.*

▶ **Definition 20.** *Let G be a edge-weighted multi-graph, and let H be a subgraph of G with bounded weighted edge-degree β. For any parameter β^-, we say that an edge $(u, v, w_{uv}) \in G \backslash H$ is (H, β, β^-)-underfull if $\frac{\mathbf{wdeg}_H(u)}{b_u} + \frac{\mathbf{wdeg}_H(v)}{b_v} < \beta^- \cdot w_{uv}$.*

▶ **Lemma 21.** *Let $0 < \varepsilon < 1/2$ be any parameter and W be an integer parameter. Set $\lambda = \frac{\varepsilon}{100W}$. Suppose that β^- and $\beta \geq \beta^- + 2$ are integers so that $\frac{\beta+8W}{\log(\beta+8W)} \geq 2W^2\lambda^{-2}$ and $\beta^- - 6W \geq (1-\lambda)\cdot(\beta+8W)$. Given an edge-weighted multi-graph G with integer weights in $1, \ldots, W$, and a subgraph H with bounded weighted edge-degree β, if X contains all edges in $G\backslash H$ that are (H, β, β^-)-underfull, then $(2 - \frac{1}{2W} + \varepsilon)\cdot w(M_{H\cup X}) \geq w(M_G)$.*

Proof. First, observe that $H \cup X$ is not necessarily a w-b-EDCS of G. Thereby we use another argument from [3]. Let M_G be a maximum weight b-matching in G, let $M_G^H = M_G \cap H$ and $M_G^{G\backslash H} = M_G \cap (G\backslash H)$. Let $X^M = X \cap M_G^{G\backslash H}$. We can observe that $w(M_G) = w(M_{H\cup M_G^{G\backslash H}})$. Then we can show that $H \cup X^M$ is a $(\beta + 2W, \beta^-)$-w-b-EDCS of $H \cup M_G^{G\backslash H}$ (see the full version of this paper). As a result, Theorem 4 can be applied in this case and we get that $(2 - \frac{1}{2W} + \varepsilon)\cdot w\left(M_{H\cup X^M}\right) \geq w\left(M_{H\cup M_G^{G\backslash H}}\right) = w(M_G)$, thus concluding the proof. ◀

▶ Remark 22. One can easily notice that there exist integers β and β^- that are $O(poly(W, 1/\varepsilon))$ satisfying the conditions of Lemma 21. From now on, we will use the parameters λ, β, and β^- satisfying the conditions of Lemma 21 and they are of values $O(poly(W, 1/\varepsilon))$.

Algorithm 1 Main algorithm computing a weighted b-matching for a random-order stream.

1: $H \leftarrow \emptyset$
2: $\forall\, 0 \leq i \leq \log_2 m$, $\alpha_i \leftarrow \left\lfloor \frac{\varepsilon \cdot m}{\log_2(m)\cdot(2^{i+2}\beta^2W^2+1)} \right\rfloor$
3: **for** $i = 0 \ldots \log_2 m$ **do**
4:     ProcessStopped $\leftarrow$ False
5:     **for** $2^{i+2}\beta^2W^2 + 1$ iterations **do**
6:         FoundUnderfull $\leftarrow$ False
7:         **for** α_i iterations **do**
8:             let (u, v, w_{uv}) be the next edge in the stream
9:             **if** $\frac{\mathbf{wdeg}_H(u)}{b_u} + \frac{\mathbf{wdeg}_H(v)}{b_v} < \beta^- \cdot w_{uv}$ **then**
10:                 add edge (u, v, w_{uv}) to H
11:                 FoundUnderfull $\leftarrow$ True
12:                 **while** there exists $(u', v', w_{u'v'}) \in H : \frac{\mathbf{wdeg}_H(u')}{b_{u'}} + \frac{\mathbf{wdeg}_H(v')}{b_{v'}} > \beta \cdot w_{u'v'}$ **do**
13:                     remove $(u', v', w_{u'v'})$ from H
14:         **if** FoundUnderfull = False **then**
15:             ProcessStopped $\leftarrow$ True
16:             **break** from the loop
17:     **if** ProcessStopped = True **then**
18:         **break** from the loop
19: $X \leftarrow \emptyset$
20: **for** each (u, v, w_{uv}) remaining edge in the stream **do**
21:     **if** $\frac{\mathbf{wdeg}_H(u)}{b_u} + \frac{\mathbf{wdeg}_H(v)}{b_v} < \beta^- \cdot w_{uv}$ **then**
22:         add edge (u, v, w_{uv}) to X
23: **return** the maximum weight b-matching in $H \cup X$

The algorithm, formally described in Algorithm 1, consists of two phases. The first phase, corresponding to Lines 3-18, constructs a subgraph H of bounded weighted edge-degree β using only a ε fraction of the stream E^{early}. In the second phase, the algorithm collects the underfull edges in the remaining part of the stream E^{late}. As in [3] we use the idea that if no underfull edge was found in an interval of size α (see Lines 6-13), with high probability the number of underfull edges remaining in the stream is bounded by some value $\gamma = 4\log(m)\frac{m}{\alpha}$. The issue is therefore to choose the right size of interval α, because we do not know the order of magnitude of $|M_G|$ in the b-matching problem: if we do as in [3] by choosing only one fixed size of intervals α, then if α is too small, the value of γ will be too big compared to $|M_G|$, whereas if the value of α is too large we will not be able to terminate the first phase of the algorithm within the early fraction of size εm. Therefore, the idea in the first phase of the algorithm is to "guess" the value of $\log_2|M_G|$ by trying successively larger and larger values of i (see Line 3). In fact, by setting $i_0 = \lceil \log_2|M_G| \rceil$, we know that the number of operations that can be performed on a w-b-EDCS is bounded by $2^{i_0+2}\beta^2 W^2$ (see the proof of Proposition 18). As a result we know that the first phase should always stop at a time where i is smaller than or equal to i_0, and therefore at a time when $\alpha_i \geq \alpha_{i_0}$. Then we can prove that with high probability the number of remaining underfull edges in the stream is at most $\gamma_i = 4\log(m)\frac{m}{\alpha_i}$.

Algorithm 1 works when M_G is neither too small nor too big. Here we will first argue that the other border cases can be handled anyway. We first have this easy lemma (its proof is very similar to that of Proposition 17, see the full version of this paper).

▶ **Lemma 23.** *We have the inequality $|G| \leq 2n \cdot |M_G|$.*

Then we use it to handle the case of small b-matchings.

▷ Claim 24. We can assume that $w(M_G) \geq \frac{3W^2}{2\varepsilon^2}\log(m)$.

Proof. In fact, if $w(M_G) < \frac{3W^2}{2\varepsilon^2}\log(m)$, then $|M_G| < \frac{3W^2}{2\varepsilon^2}\log(m)$ and by Lemma 23 the graph has only $m = O(n \cdot \frac{3W^2}{2\varepsilon^2} \cdot \log(m))$ edges, so the whole graph can be stored only using $O(n \cdot poly(\log(m), W, 1/\varepsilon))$ memory, implying that we can compute an exact solution. ◁

▷ Claim 25. Assuming Claim 24, with probability at least $1 - m^{-3}$ the late part of the steam E^{late} contains at least a $(1-2\varepsilon)$ fraction of the optimal b-matching.

Proof. Consider a maximum weight b-matching $M_G = \{f_1, \dots, f_{|M_G|}\}$. We define the random variables $X_i = \mathbb{1}_{f_i \in E^{early}} \cdot w(f_i)$. Hence we have $\mathbb{E}[\sum X_i] = \varepsilon \cdot w(M_G)$. Moreover, the random variables X_i are negatively associated, so we can use Hoeffding's inequality to get $\mathbb{P}\left[\sum_{i=1}^{|M_G|} X_i \geq 2\varepsilon \cdot w(M_G)\right] \leq \exp\left(-\frac{2\cdot\varepsilon^2\cdot w(M_G)^2}{|M_G|\cdot W^2}\right) \leq \exp\left(-\frac{2\cdot\varepsilon^2\cdot w(M_G)}{W^2}\right) \leq m^{-3}$, as we now assume that $w(M_G) \geq \frac{3W^2}{2\varepsilon^2}\log(m)$ (see Claim 24). ◁

Recall that we defined $i_0 = \lceil \log_2|M_G| \rceil$.

▷ Claim 26. We can assume that $\frac{\varepsilon\cdot m}{\log_2(m)\cdot(2^{i_0+2}\beta^2W^2+1)} \geq 1$.

Proof. If this is not the case, then we can just store all the edges of G as the number of edges m is bounded by $\frac{\log_2(m)\cdot(2^{i_0+2}\beta^2W^2+1)}{\varepsilon} = O(|M_G| \cdot poly(\log(m), W, 1/\varepsilon))$ (as β is $O(poly(W, 1/\varepsilon))$, see Remark 22). As a result, if at some point of the first phase we have not stopped and we have $\alpha_i = 0$, then we store all the remaining edges of E^{late} and we will be able to get a $(1-2\varepsilon)$ approximation with high probability (because of Claim 25) using $O(|M_G| \cdot poly(\log(m), W, 1/\varepsilon))$ memory. ◁

Then we can move on to our main algorithm. The following lemma is very similar to the one used in [3] (see the proof in the full version of this paper). It can then be combined with previous lemmas and claims to prove that a $2 - \frac{1}{2W} + \varepsilon$ approximation can be achieved with high probability.

▶ **Lemma 27.** *The first phase of Algorithm 1 uses $O(\beta \cdot |M_G|)$ memory and constructs a subgraph H of G, satisfying the following properties:*

1. *The first phase terminates within the first εm edges of the stream.*
2. *When the first phase terminates after processing some edge, we have:*
   - **a.** *H has bounded weighted edge degree β, and contains at most $O(\beta \cdot |M_G|)$ edges.*
   - **b.** *With probability at least $1 - m^{-3}$, the total number of (H, β, β^-)-underfull edges in the remaining part of the stream is at most $\gamma = O(|M_G| \cdot (\log(m))^2 \cdot \beta^2 W^2 \cdot 1/\varepsilon)$.*

▶ **Theorem 28.** *Let $\varepsilon > 0$. Using Algorithm 1, with probability $1 - 2m^{-3}$, one can extract from a randomly-ordered stream of edges a weighted b-matching with an approximation ratio of $2 - \frac{1}{2W} + \varepsilon$, using $O(\max(|M_G|, n) \cdot poly(\log(m), W, 1/\varepsilon))$ memory.*

Proof. Applying Lemma 21 to the graph $H \cup G^{late}$ we can get, choosing the right values β and β^- (which are $O(poly(W, 1/\varepsilon))$), $H \cup X$ contains a $(1-2\varepsilon)^{-1} \cdot (2 - \frac{1}{2W} + \varepsilon)$ approximation of the optimal b-matching (with probability at least $1 - m^{-3}$, see Claim 25), and with a memory consumption of $O(|M_G| \cdot poly(\log(m), W, 1/\varepsilon))$ (with probability at least $1 - m^{-3}$, see Lemma 27), with probability at least $1 - 2m^{-3}$ (union bound). Hence the proof. ◀

References

1 Sepehr Assadi and Soheil Behnezhad. Beating two-thirds for random-order streaming matching. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 19:1–19:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ICALP.2021.19.

2 Sepehr Assadi and Aaron Bernstein. Towards a unified theory of sparsification for matching problems. In Jeremy T. Fineman and Michael Mitzenmacher, editors, *2nd Symposium on Simplicity in Algorithms, SOSA 2019, January 8-9, 2019, San Diego, CA, USA*, volume 69 of *OASICS*, pages 11:1–11:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi:10.4230/OASIcs.SOSA.2019.11.

3 Aaron Bernstein. Improved Bounds for Matching in Random-Order Streams. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*, volume 168 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 12:1–12:13, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ICALP.2020.12.

4 Aaron Bernstein, Aditi Dudeja, and Zachary Langley. A framework for dynamic matching in weighted graphs. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 668–681. ACM, 2021. doi:10.1145/3406325.3451113.

5 Aaron Bernstein and Cliff Stein. Fully dynamic matching in bipartite graphs. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Automata, Languages, and Programming – 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, volume 9134 of *Lecture Notes in Computer Science*, pages 167–179. Springer, 2015. doi:10.1007/978-3-662-47672-7_14.

6 Aaron Bernstein and Cliff Stein. Faster fully dynamic matchings with small approximation ratios. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 692–711. SIAM, 2016. doi:10.1137/1.9781611974331.ch50.

7 Jeno Egerváry. Matrixok kombinatorius tulajdonságairól. *Matematikai és Fizikai Lapok*, 38(1931):16–28, 1931.

8 Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In *Colloquia Mathematica Societatis Janos Bolyai 10. Infinite and Finite Sets, Keszthely (Hungary)*. Citeseer, 1973.

9 Alireza Farhadi, Mohammad Taghi Hajiaghayi, Tung Mai, Anup Rao, and Ryan A. Rossi. Approximate maximum matching in random streams. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1773–1785. SIAM, 2020. doi:10.1137/1.9781611975994.108.

10 Joan Feigenbaum, Sampath Kannan, Andrew McGregor, Siddharth Suri, and Jian Zhang. On graph problems in a semi-streaming model. *Theoretical Computer Science*, 348:207–216, December 2005. doi:10.1016/j.tcs.2005.09.013.

11 Buddhima Gamlath, Sagar Kale, Slobodan Mitrovic, and Ola Svensson. Weighted matchings via unweighted augmentations. In *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing*, PODC '19, pages 491–500, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3293611.3331603.

12 Mohsen Ghaffari and David Wajc. Simplified and Space-Optimal Semi-Streaming $(2+\varepsilon)$-Approximate Matching. In Jeremy T. Fineman and Michael Mitzenmacher, editors, *2nd Symposium on Simplicity in Algorithms (SOSA 2019)*, volume 69 of *OpenAccess Series in Informatics (OASIcs)*, pages 13:1–13:8, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.SOSA.2019.13.

13 Guru Prashanth Guruganesh and Sahil Singla. Online matroid intersection: Beating half for random arrival. In Friedrich Eisenbrand and Jochen Könemann, editors, *Integer Programming and Combinatorial Optimization – 19th International Conference, IPCO 2017, Waterloo, ON, Canada, June 26-28, 2017, Proceedings*, volume 10328 of *Lecture Notes in Computer Science*, pages 241–253. Springer, 2017. doi:10.1007/978-3-319-59250-3_20.

14 Chien-Chung Huang and François Sellier. Semi-streaming algorithms for submodular function maximization under b-matching constraint. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*, 2021.

15 Michael Kapralov. Space lower bounds for approximating maximum matching in the edge arrival model. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10–13, 2021*, pages 1874–1893. SIAM, 2021. doi:10.1137/1.9781611976465.112.

16 Christian Konrad. A simple augmentation method for matchings with applications to streaming algorithms. In Igor Potapov, Paul G. Spirakis, and James Worrell, editors, *43rd International Symposium on Mathematical Foundations of Computer Science, MFCS 2018, August 27-31, 2018, Liverpool, UK*, volume 117 of *LIPIcs*, pages 74:1–74:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.MFCS.2018.74.

17 Christian Konrad, Frédéric Magniez, and Claire Mathieu. Maximum matching in semi-streaming with few passes. In Anupam Gupta, Klaus Jansen, José D. P. Rolim, and Rocco A. Servedio, editors, *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*, volume 7408 of *Lecture Notes in Computer Science*, pages 231–242. Springer, 2012. doi:10.1007/978-3-642-32512-0_20.

18 Roie Levin and David Wajc. Streaming submodular matching meets the primal-dual method. In *SODA*, pages 1914–1933, 2021.

19 Ami Paz and Gregory Schwartzman. A $(2+\varepsilon)$-approximation for maximum weight matching in the semi-streaming model. In *Proceedings of the 2017 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2153–2161, 2017. doi:10.1137/1.9781611974782.140.

20 Moshe Tennenholtz. Some tractable combinatorial auctions. In *AAAI/IAAI*, pages 98–103, 2000.