

# The Stochastic Matching Problem: Beating Half with a Non-Adaptive Algorithm

Sepehr Assadi\*

Sanjeev Khanna\*

Yang Li\*

## Abstract

In the *stochastic matching* problem, we are given a general (not necessarily bipartite) graph  $G(V, E)$ , where each edge in  $E$  is *realized* with some constant probability  $p > 0$  and the goal is to compute a *bounded-degree* (bounded by a function depending only on  $p$ ) subgraph  $H$  of  $G$  such that the expected maximum matching size in  $H$  is close to the expected maximum matching size in  $G$ . The algorithms in this setting are considered *non-adaptive* as they have to choose the subgraph  $H$  without knowing any information about the set of realized edges in  $G$ . Originally motivated by an application to *kidney exchange*, the stochastic matching problem and its variants have received significant attention in recent years.

The state-of-the-art non-adaptive algorithms for stochastic matching achieve an approximation ratio of  $\frac{1}{2} - \varepsilon$  for any  $\varepsilon > 0$ , naturally raising the question that if  $1/2$  is the limit of what can be achieved with a non-adaptive algorithm. In this work, we resolve this question by presenting the first algorithm for stochastic matching with an approximation guarantee that is strictly better than  $1/2$ : the algorithm computes a subgraph  $H$  of  $G$  with the maximum degree  $O\left(\frac{\log(1/p)}{p}\right)$  such that the ratio of expected size of a maximum matching in realizations of  $H$  and  $G$  is at least  $1/2 + \delta_0$  for some absolute constant  $\delta_0 > 0$ . The degree bound on  $H$  achieved by our algorithm is essentially the best possible (up to an  $O(\log(1/p))$  factor) for any *constant factor* approximation algorithm, since an  $\Omega(\frac{1}{p})$  degree in  $H$  is necessary for a vertex to acquire at least one incident edge in a realization.

Our result makes progress towards answering an open problem of Blum *et al.* (EC 2015) regarding the possibility of achieving a  $(1 - \varepsilon)$ -approximation for the stochastic matching problem using non-adaptive algorithms. From the technical point of view, a key ingredient of our algorithm is a structural result showing that a graph whose expected maximum matching size is  $\text{OPT}$  always contains a  $b$ -matching of size (essentially)  $b \cdot \text{OPT}$ , for  $b = \frac{1}{p}$ .

---

\*Department of Computer and Information Science, University of Pennsylvania. Supported in part by National Science Foundation grants CCF-1552909, CCF-1617851, and IIS-1447470.  
Email: {sassadi, sanjeev, yangli2}@cis.upenn.edu.

# 1 Introduction

We study the problem of finding a maximum matching in presence of *uncertainty* in the input graph. Specifically, we consider the *stochastic* setting where for an input graph  $G(V, E)$  and a parameter  $p > 0$ , each edge in  $E$  is realized *independently* w.p.<sup>1</sup>  $p$ . We call the graph obtained from this stochastic process (which should be viewed as a random variable) a *realization* of  $G(V, E)$ , denoted by  $G_p(V, E_p)$ . The *stochastic matching* problem can now be defined as follows. Given a general (not necessarily bipartite) graph  $G(V, E)$  and an edge realization probability  $p > 0$ , compute a subgraph  $H$  of  $G$  such that:

- (i) The expected maximum matching size in a realization of  $H$  is close to the expected maximum matching size in a realization of  $G$ .
- (ii) The degree of each vertex in  $H$  is bounded by some function that only depends on  $p$ , independent of the size of  $G$ .

In other words, the stochastic matching problem asks if every graph  $G$  contains a subgraph  $H$  of *bounded degree* (depending only on the realization probability  $p$ ) such that the expected matching size in realizations of  $G$  and  $H$  are close.

**Kidney exchange.** A canonical and arguably the most important application of the stochastic matching problem appears in *kidney exchange*, where patients waiting for kidney transplant can *swap* their incompatible donors to each get a compatible donor. The goal is to identify a maximum set of patient-donor pairs to perform such a swap (i.e., finds a maximum matching). However, through medical records of patients and donors, one can only filter out the patient-donor pairs where donation is *impossible*, and more costly and time consuming tests must be performed before a transplant can be performed.

The stochastic setting captures the essence of the need of extra tests for kidney exchange: an algorithm selects a set of patient-donor pairs to perform the extra tests (i.e., computes a subgraph  $H$ ), while making sure that there is a large matching among the pairs that pass the extra tests. The objective that the subgraph  $H$  has small degree captures the essence of minimizing the number of (costly and time consuming) tests that each patient needs to go through. The kidney exchange problem has been extensively studied in the literature, particularly under stochastic settings (see, e.g., [2, 4, 5, 9, 17–19, 29, 34]). We remark that the stochastic matching problem captures the simplest form of the kidney exchange, referred to as *pairwise exchange*. Modern kidney exchange programs regularly employ swaps between three or patient-donor pairs and this setting has also been studied previously in the literature; we refer the interested reader to [11] for more details.

**Previous work.** Our results are directly related to the results in [11] and [8] which we describe in detail below. Blum *et al.* [11] introduced the (variant of) stochastic matching problem and proposed a  $(\frac{1}{2} - \varepsilon)$ -approximation algorithm (for any  $\varepsilon > 0$ ) which requires the subgraph  $H$  to have maximum degree of  $\frac{\log(1/\varepsilon)}{p^{\Theta(1/\varepsilon)}}$ . The algorithm of Blum *et al.* [11] works as follows: Pick a maximum matching  $M_i$  in  $G$  and remove the edges in  $M_i$ ; repeat for  $R := \frac{\log(1/\varepsilon)}{p^{\Theta(1/\varepsilon)}}$  times. In order to analyze this algorithm, the authors showed that, for any  $i \in [R]$ , if the size of the maximum matching among the realized edges in  $M_1, \dots, M_i$  is less than  $\text{OPT}/2$ , the matching  $M_{i+1}$  contains many augmenting paths of  $M$

---

<sup>1</sup>Throughout, we use *w.p.*, *w.h.p.*, and *prob.* to abbreviate “with probability”, “with high probability”, and “probability”, respectively.

of length  $O(\frac{1}{\varepsilon})$ ; since each such augmenting path is realized w.p.  $p^{O(\frac{1}{\varepsilon})}$ , one needs to repeat this augmentation process for  $\frac{1}{p^{O(\frac{1}{\varepsilon})}}$  time (as is roughly the value of  $R$ ) to increase the matching size to  $(\frac{1}{2} - \varepsilon) \cdot \text{OPT}$ .

In a recent work [8], we showed that in order to obtain a  $(\frac{1}{2} - \varepsilon)$ -approximation algorithm, one only needs a subgraph  $H$  with max-degree of  $O(\frac{\log(1/\varepsilon p)}{\varepsilon p})$ , significantly smaller than the bounds in [11]. Interestingly, the algorithm of [8] and the one in [11] are essentially identical (modulo an extra sparsification part required in [8]) and the main difference is in the analysis. In [8], we completely bypassed the need for using augmenting paths in the analysis and instead, took advantage of structural properties of matchings in a global manner (by using *Tutte-Berge formula*; see, e.g., [28]). In particular, we showed that repeatedly picking  $O(\frac{\log(1/\varepsilon p)}{\varepsilon p})$  maximum matchings (as described before) suffices to ensure that, among the chosen edges, a matching of size (essentially) equal to the size of the last chosen matching would be realized (with high probability). Having this, one can show that running the aforementioned algorithm even for  $R := O(\frac{\log(1/\varepsilon p)}{\varepsilon p})$  suffices to obtain a  $(\frac{1}{2} - \varepsilon)$ -approximation.

*Adaptive* algorithms for stochastic matching have also been studied by [8, 11]. In an adaptive algorithm, instead of a single graph  $H$ , one is allowed to pick a *small* number of bounded-degree graphs  $H_1, \dots, H_k$  where the choice of each  $H_i$  can be made after *probing* the edges in  $H_1, H_2, \dots, H_{i-1}$  to see if they are realized or not. A  $(1 - \varepsilon)$ -approximation adaptive algorithm for this problem was first proposed in [11] and further refined in [8].

**Beating the half approximation.** This state-of-the-art highlights the following natural question:

*Is half-approximation the limit for non-adaptive algorithms or is there a non-adaptive algorithm that achieves approximation guarantee of strictly better than half?*

It is worth mentioning that in many variations, obtaining half approximation for the maximum matching problem is typically a relatively easy task (usually via a greedy approach), while beating half approximation turns out to be a difficult task. Some notable examples include, randomized greedy matching [6, 14, 20, 32], online stochastic matching [26, 30, 31], and semi-streaming matching [21, 27].

## 1.1 Our Contributions

We resolve the aforementioned question of obtaining an algorithm for stochastic matching with an approximation guarantee of *strictly* better than half. Formally,

**Theorem 1.** *There exists an algorithm that given any graph  $G(V, E)$  and any parameter  $p > 0$ , computes a subgraph  $H(V, Q)$  of  $G$  with a maximum degree of  $O\left(\frac{\log(1/p)}{p}\right)$  such that the ratio of the expected maximum matching size of a realization of  $H$  to a realization of  $G$  is at least:*

- (i) *0.52 when  $p \leq p_0$  for an absolute constant  $p_0 > 0$ .*
- (ii)  *$0.5 + \delta_0$  for any  $0 < p < 1$ , where  $\delta_0 > 0$  is an absolute constant.*

Our result in Theorem 1 makes progress towards an open problem posed by Blum *et al.* [11] regarding the possibility of having a non-adaptive  $(1 - \varepsilon)$ -approximation algorithm for stochastic matching. We further remark that the assumption in Part (i) of Theorem 1 is standard in the stochastic matching literature and is referred as the case of *vanishing probabilities*, see, e.g. [30, 31].

It is worth mentioning the max-degree on  $H$  achieved in Theorem 1 is essentially the best possible (up to an  $O(\log \frac{1}{p})$  factor) for any *constant factor* approximation algorithm: suppose  $G$  is a complete graph; in this case the expected matching size in  $G$  is  $n - o(n)$  by standard results on random graphs (see, e.g., [13], Chapter 7); however, if max-degree of  $H$  is  $o(\frac{1}{p})$ , then the expected number of realized edges in  $H$  is  $o(n)$ , implying that the expected matching size in  $H$  is  $o(n)$ .

Our approach to proving Theorem 1 can be divided into two parts. In the first part, we prove a structural result showing that if a realization of  $G$  has expected maximum matching size  $\text{OPT}$ , then  $G$  itself should contain essentially  $\frac{1}{p}$  edge-disjoint matchings of size  $\text{OPT}$  each. This result, established through a characterization of  $b$ -matching size in general graphs (see Section 3), sheds more light into the structure of a graph in terms of its expected maximum matching size, which may be of independent interest.

In the second part, we combine the aforementioned structural result with the  $(\frac{1}{2} - \varepsilon)$ -approximation algorithm of [8] to obtain a matching of size strictly larger than  $\text{OPT}/2$ . In order to do this, we first find a collection of  $\frac{1}{p}$  edge-disjoint matchings of size at least  $\text{OPT}$ , remove them from the graph, and then run the algorithm of [8] on the remaining edges. We show that the edges in this collection of edge-disjoint matchings must form many *length-three augmenting paths* of the matching computed by the algorithm of [8], hence leading to a matching of size strictly larger than  $\text{OPT}/2$ . The analysis is separated into two steps: we first formulate the increment in the matching size (through these augmenting paths) via a (non-linear) minimization program, and then analyze the optimal solution of this minimization program and hence lower bound the increment in the matching size obtained from the augmenting paths.

**Other related work.** Multiple variants of stochastic matching have been considered in the literature. Blum *et al.* [12] studied a similar setting where one can only probe *two* edges incident on any vertex and the goal is to find the optimal set of edges to query. Another well studied setting is the *query-commit* model, whereby an algorithm probes one edge at a time and if an edge  $e$  is probed and realized, then the algorithm must take  $e$  as part of the matching it outputs [1, 10, 15, 16, 23]. We refer the reader to [11] for a detailed description of the related work.

**Organization.** The rest of the paper is organized as follows. We start by providing a high level overview of our algorithm in Section 2. Next, in Section 3, we introduce the notation and preliminaries needed for the rest of the paper. We prove our main structural result, i.e.,  $b$ -matching lemma in Section 4. Our main algorithm and its analysis, i.e., the proof of Part (i) of Theorem 1 are provided in Section 5. Proof of Part (ii) of Theorem 1, i.e., an algorithm that works for the large-probability case appears in Section 6. We conclude the paper in Section 7.

## 2 Technical Overview

In this section, we give a more detailed overview of the main ideas used in our algorithm for stochastic matching. For clarity of exposition, throughout this section, we assume  $p$  is a sufficiently small constant (corresponding to Part (i) of Theorem 1) and the expected maximum matching size in  $G$  (i.e.,  $\text{OPT}$ ) is  $n - o(n)$ , or in other words, a realization of  $G$ ,  $G_p$ , has a near perfect matching in expectation.

Our starting point is the following observation: In order for  $G_p$  to have a (near) perfect matching in expectation, the input graph  $G$  must have many (roughly  $1/p$ ) edge-disjoint (near) perfect matchings. To gain some intuition why this is true, suppose for the moment that the input graph is a bipartite graph  $G(L, R, E)$ . Then, by *Hall's Marriage Theorem*, we know that in order for  $G_p$

to have a matching of size  $n - o(n)$ , for any two subsets  $X \subseteq L$  and  $Y \subseteq R$ , with  $|X| - |Y| \geq o(n)$ , at least one edge from  $X$  to  $\bar{Y}$  should realize in  $G_p$ . However, this requirement implies that in  $G$ , there should be  $1/p$  edges from  $X$  to  $\bar{Y}$  so that at least one of these edges appears in  $G_p$ . One can then show that a bipartite graph  $G$  with such a structure has  $1/p$  edge-disjoint matchings of size at least  $n - o(n)$ .

In general, we need to handle graphs that are not necessarily bipartite. In order to adapt the previous strategy, we slightly relax our requirement of having  $1/p$  edge-disjoint matchings to having one (simple)  $b$ -matching<sup>2</sup> of size  $nb$  for the parameter  $b = \frac{1}{p}$ . We show that,

**b-Matching Lemma.** Any graph  $G$  where  $G_p$  has a matching of size  $n - o(n)$  in expectation, has a  $\frac{1}{p}$ -matching of size (essentially)  $\frac{n}{p}$ .

Next, we combine the fact that a large  $\frac{1}{p}$ -matching, denoted by  $B$ , always exists in  $G$ , with the  $(\frac{1}{2} - \varepsilon)$ -approximation algorithm of [8] to obtain a strictly better than  $\frac{1}{2}$ -approximation algorithm.

To continue, we briefly describe the algorithm of [8], which we refer to as **MatchingCover**. **MatchingCover** works by picking a maximum matching  $M_i$  in  $G$  and removing the edges of  $M_i$  for  $R := \Theta\left(\frac{\log(1/p)}{p}\right)$  times<sup>3</sup>. This collection of matchings, denoted by  $E_{MC}$ , is referred to as a *matching cover* of the original graph  $G$ . The main property of this matching cover, proved in [8], is that the set of realized edges in  $E_{MC}$  has a matching of size (essentially)  $|M_R|$ ; note that  $M_R$  is the smallest size matchings among the matchings in  $E_{MC}$ .

We are now ready to define our main algorithm: Pick a maximum  $\frac{1}{p}$ -matching  $B$  from  $G$ ; run **MatchingCover** over the edges  $E \setminus B$  and obtain a matching cover  $E_{MC}$ ; return  $H(V, B \cup E_{MC})$ . If  $|M_R| < (\frac{1}{2} - \delta_0)n$ , using the fact that  $E_{MC}$  is obtained by repeatedly picking maximum matchings, one can show that any matching  $M$  of size  $n - o(n)$  in  $G$  has more than  $(\frac{1}{2} + \delta_0)n - o(n)$  edges in  $B \cup E_{MC}$ . This also implies that the expected matching size in  $H$  is at least  $(\frac{1}{2} + \delta_0)n - o(n)$ . The more difficult case, which is where we concentrate bulk of our technical effort, is when  $|M_R| \geq (\frac{1}{2} - \delta_0)n$ . For simplicity, assume  $|M_R| = n/2$  from here on.

As stated above, if  $|M_R| = n/2$ , then in almost every realization of the edges in  $E_{MC}$ , there exists a matching  $M$  of size at least  $n/2$ . Our strategy is to *augment* the matching  $M$  using the (realized) edges in  $B$ , so that the matching size becomes  $(\frac{1}{2} + \delta_0)n$ . It is important to note that the set of edges in  $E_{MC}$  and  $B$  are disjoint, and hence whether edges in  $E_{MC}$  and  $B$  are realized are independent of each other.

Let  $U$  be the set of vertices matched by  $M$ . There are two cases here to consider:

- **Case 1.** Nearly all edges in  $B$  are incident on vertices in  $U$ .
- **Case 2.** An  $\varepsilon$ -fraction of edges in  $B$  are not incident on  $U$  (for some constant  $\varepsilon > 0$ ).

The second case is relatively easy to handle: we show that a realization of a  $\frac{1}{p}$ -matching with  $N/p$  edges has a matching of size at least  $N/3$  in expectation. This implies that  $B_p$  has a matching  $M'$  of size  $\varepsilon \cdot \frac{n}{3} = \Theta(\varepsilon) \cdot n$  which is not incident on  $U$ . Consequently,  $B \cup E_{MC}$  has a matching of size  $\frac{n}{2} + \Theta(\varepsilon) \cdot n$  in expectation. The more challenging task is to tackle the first case. To convey the main idea, we make a series of simplifying assumptions here: (i) all edges in  $B$  are incident on  $U$ , (ii) each edge in  $B$  is incident on exactly one vertex in  $U$ , and (iii) every vertex in  $U$  is incident on exactly  $\frac{1}{p}$  edges of  $B$ .

---

<sup>2</sup>Recall that a (simple)  $b$ -matching is simply a graph with degree of each vertex bounded by  $b$ . See Section 3 for more details.

<sup>3</sup>We remark that this algorithm has an extra *sparsification* step which is needed to handle the case where  $\text{OPT} = o(n)$ . However, since in this section we assume  $\text{OPT} = n - o(n)$ , this extra step is not required.

Our goal is to identify a large collection of length-three augmenting paths for the matching  $M$  using the edges of  $B$ . To achieve this, we consider the event that an edge  $(u, v)$  in  $M$  has a length-three augmenting path  $a - u - v - b$  where  $u$  (resp.  $v$ ) is the only neighbor of  $a$  (resp.  $b$ ). We say such an edge  $(u, v)$  is *successful*. Since the length-three augmenting that certifies successful edges are vertex-disjoint by definition, they can all (simultaneously) augment  $M$ . Consequently, it suffices to lower bound the expected number of successful edges, or, equivalently, to lower bound the prob. that each edge is successful.

Let us further assume for the moment that  $G$  is a bipartite graph. In this case,  $u$  and  $v$  do not share a common neighbor and we can consider the neighborhood of  $u$  and  $v$  separately. The prob. that  $u$  has a neighbor  $w$  where  $u$  is the only neighbor of  $w$  (we say  $u$  is successful in this case) is not difficult to bound: enumerate all  $1/p$  neighbors  $w$  of  $u$  and account for the prob. that the edge  $(u, w)$  is realized and the prob. that no other edge incident on  $w$  is realized. A similar argument can be made for  $v$ . Now, the prob. that  $(u, v)$  is successful is simply the product of the prob. that  $u$  is successful and the prob. that  $v$  is successful.

However, in general (non-bipartite) graphs,  $u$  and  $v$  might have common neighbors which results in prob. of  $u$  being successful *not independent* of prob. of  $v$  being successful. Handling this case requires a more careful argument and analysis. Moreover, recall that in the above discussion, we made rather strong simplifying assumptions about how the edges in  $B$  are distributed across the vertices of  $U$ . In order to further remove these assumptions, in the actual analysis, we cast the probability of each edge  $(u, v)$  being successful as a function of the degrees of the vertices  $u$  and  $v$ , and formulate a (non-linear) minimization program to capture the minimum number of possible successful edges. Finally, we analyze the optimal solution of this minimization program, which allows us to achieve the target lower bound on the expected increment in the matching size.

### 3 Preliminaries

**Notation.** For a graph  $G(V, E)$ ,  $n$  denotes the number of vertices in  $G$ . For any  $U \subseteq V$ , we use  $G[U]$  to denote the subgraph of  $G$  induced only on vertices in  $U$ , and use  $E[U]$  to denote the set of edges in  $G[U]$ , i.e., the set of edges with both end points in  $U$ . For any two subsets  $U, W$  of  $V$ , we further use  $E[U, W]$  to denote the set of edges with one end point in  $U$  and another in  $W$ . For any  $X \subseteq E$ , we use  $V(X)$  to denote the set of vertices incident on  $X$ . Finally, we use  $\mu(E)$  to denote the maximum matching size among a set of edge  $E$ .

When sampling from a set of edges  $X$  (resp. a graph  $H$ ) where each edge in  $X$  (resp.  $H$ ) is sampled w.p.  $p$ , we use  $X_p$  (resp.  $H_p$ ) to denote the random variable for the set of sampled edges. We use  $\text{OPT}(G)$  (or shortly  $\text{OPT}$  if the graph  $G$  is clear from the context) to denote the *expected* maximum matching size of a realization of  $G$  (i.e.,  $G_p(V, E_p)$ )<sup>4</sup>. For any algorithm for the stochastic matching problem, we use  $\text{ALG}$  to denote the expected matching size in a realization of  $H$ , where  $H$  is the subgraph computed by the algorithm.

**b-matchings.** For any graph  $G(V, E)$  and any integer  $b \geq 1$ , a subset  $M \subseteq E$  is called a *simple b-matching*, iff the number of edges  $M$  that are incident on each vertex is at most  $b$ . Throughout, we drop the word ‘simple’, and refer to  $M$  as a  $b$ -matching.

We use the following characterization of the maximum  $b$ -matching size in general graphs (see [33], Volume A, Chapter 33).

---

<sup>4</sup>We assume  $\text{OPT} = \omega(1)$  to obtain the desired concentration bounds (for example in Lemma 3.3).

**Theorem 2.** Let  $G(V, E)$  be a graph and  $b \geq 1$  be any integer. The maximum size of a  $b$ -matching is equal to the minimum value of

$$b \cdot |U| + |E[W]| + \sum_K \left\lfloor \frac{1}{2} \left( b \cdot |K| + |E[K, W]| \right) \right\rfloor$$

taken over all disjoint subsets  $U, W$  of  $V$ , where  $K$  ranges over all connected components in the graph  $G[V - U - W]$ .

**Useful inequalities.** We also use the following simple inequalities. The Proofs are provided in Appendix A for completeness.

**Proposition 3.1.** Let  $f(x) := \frac{1-e^{-x}}{x}$ . Then, for any  $c$ , and any  $x \in [0, c]$ ,  $e^{-x} \leq 1 - f(c) \cdot x$ .

**Proposition 3.2.** For any  $x \in (0, 0.43]$ ,  $(1-x)^{\frac{1}{x}} \geq \frac{1-x}{e}$ .<sup>5</sup>

### 3.1 MatchingCover Algorithm

We use the  $(0.5 - \varepsilon)$ -approximation algorithm of [8] (Algorithm 3) as a sub-routine. For simplicity, throughout the paper, we refer to this algorithm as **MatchingCover**. In the following lemma, we summarize the properties of **MatchingCover** that we use in this paper. The proof of this lemma immediately follows from Lemma 3.9 and Lemma 5.2 in [8].

**Lemma 3.3** ([8]). For any graph  $G(V, E)$ , and any input parameter  $\varepsilon > 0$ ,  $\text{MatchingCover}(G, \varepsilon)$  outputs a collection of  $R$  matchings  $M_1, M_2, \dots, M_R$  (denote  $E_{MC} = M_1 \cup M_2 \cup \dots \cup M_R$ ), such that, w.p.  $1 - o(1)$ :

1. The size of a maximum matching among realized edges in  $E_{MC}$  is at least  $(1 - \varepsilon) |M_R|$ .
2.  $|M_1| \geq \dots \geq |M_R| \geq (1 - \varepsilon) \cdot \mu(E \setminus E_{MC})$ .
3.  $R = \Theta(\frac{\log 1/(\varepsilon p)}{\varepsilon p})$ .

We can also prove the following simple claim based on the second property of the **MatchingCover** in Lemma 3.3. Roughly speaking, this claim states that if the **MatchingCover** is not able to extract any further large matching (of size essentially  $\text{OPT}/2$ ) from  $G$ , then the set of extracted edges already provides a matching of size  $\text{OPT}/2$  in any realization. A similar result is proven in [8] (see Lemma 5.3); however, since Claim 3.4 does not follow directly from the results in [8], we provide a self-contained proof of this claim here.

**Claim 3.4.** Fix  $0 < \varepsilon < \delta < 1$ . Let  $G(V, E)$  be a graph,  $X$  be any arbitrary subset of  $E$ , and  $(M_1, \dots, M_R) = \text{MatchingCover}(G(V, E \setminus X), \varepsilon)$ . Define  $E_{MC} = M_1 \cup \dots \cup M_R$ . If  $|M_R| \leq (\frac{1}{2} - \delta) \text{OPT}$ , then the expected maximum matching size in a realization of  $G(V, X \cup E_{MC})$  is at least  $(\frac{1}{2} + \delta - \varepsilon) \text{OPT}$ .

*Proof.* For each realization of  $G_p$ , we fix one maximum matching. Now the expected matching size in  $G_p$  can be written as

$$\text{OPT} = \sum_M \Pr(M \text{ is the fixed maximum matching in } G_p) \cdot |M|$$

---

<sup>5</sup>This inequality actually holds for any  $x \in [0, 1]$ . However, as we only need the range  $(0, 0.43]$  in our proofs and this allows us to provide a simpler proof, we only consider this range.

By property (2) of MatchingCover in Lemma 3.3, the maximum matching size in the graph  $G(V, E \setminus (X \cup E_{MC}))$  is at most  $(1 + \varepsilon)|M_R|$ . Therefore, for any matching  $M$ , at most  $(1 + \varepsilon)|M_R|$  edges of  $M$  is in  $E \setminus (X \cup E_{MC})$ , and hence at least  $|M| - (1 + \varepsilon)L$  edges of  $M$  is in  $X \cup E_{MC}$ . This implies that if all edges in  $M$  are realized, a matching of size at least  $|M| - (1 + \varepsilon)|M_R|$  is realized in  $Q$ . Let  $\text{ALG}$  be the expected maximum matching size in  $G(V, X \cup E_{MC})$ ; we have,

$$\begin{aligned}\text{ALG} &\geq \sum_M \Pr(M \text{ is the fixed maximum matching in } G_p)(|M| - (1 + \varepsilon)|M_R|) \\ &= \text{OPT} - (1 + \varepsilon)|M_R|\end{aligned}$$

Since  $|M_R| \leq (1/2 - \delta)\text{OPT}$ , we have,

$$\text{ALG} \geq \text{OPT} - (1 + \varepsilon) \cdot (1/2 - \delta) \cdot \text{OPT} \geq (1/2 + \delta - \varepsilon) \cdot \text{OPT}$$

which concludes the proof.  $\blacksquare$

## 4 $b$ -Matching Lemma

Here, we develop one of the main ingredients of our algorithm, namely, any input graph  $G$  contains a  $b$ -matching of size almost  $b \cdot \text{OPT}(G)$  for  $b = 1/p$ . Intuitively, if the expected matching size in  $G$  is  $\text{OPT}$ , then since only  $p$  fraction of edges are realized in expectation, one may hope to find up to  $1/p$  edge-disjoint matchings of size  $\text{OPT}$  in  $G$ . The following lemma formalizes this intuition by using  $b$ -matchings (for  $b = 1/p$ ) instead of a collection of edge-disjoint matchings.

**Lemma 4.1** ( $b$ -matching lemma). *Let  $b = \left\lfloor \frac{1}{p} \right\rfloor$ ; any graph  $G(V, E)$  has a  $b$ -matching of size at least  $(b - 1) \cdot \text{OPT}(G)$ .*

*Proof.* Suppose by contradiction that the maximum  $b$ -matching  $B$  in  $G$  is of size less than  $(b - 1) \cdot \text{OPT}$ . Consequently, by Theorem 2, there exist disjoint subsets  $U, W$  of  $V$  such that,

$$b \cdot |U| + |E[W]| + \sum_K \left\lfloor \frac{1}{2} \left( b \cdot |K| + |E[K, W]| \right) \right\rfloor < (b - 1) \cdot \text{OPT} \quad (1)$$

where  $K$  ranges over all connected components in the graph  $G[V - U - W]$ . Let  $c$  be the number of connected components in  $G[V - U - W]$ . We first note that  $c < 2\text{OPT}$ ; otherwise,

$$\sum_K \left\lfloor \frac{1}{2} \left( b \cdot |K| + |E[K, W]| \right) \right\rfloor \geq \sum_K \left\lfloor \frac{b}{2} \right\rfloor \geq c \cdot \left( \frac{b - 1}{2} \right) \geq (b - 1) \cdot \text{OPT}$$

and hence the LHS in Eq (1) would be more than  $(b - 1) \cdot \text{OPT}$ , i.e., the RHS; a contradiction.

Additionally, we have  $|U| + |W| + \sum_K |K| = n$ . Hence, by multiplying each side in Eq (1) by 2 and plugging in this bound, we have,

$$\begin{aligned}2b \cdot \text{OPT} - 2\text{OPT} &> nb - b|W| + b|U| + 2|E[W]| + \sum_K (|E[K, W]| - 1) \\ &\geq nb - b|W| + b|U| + 2|E[W]| + \sum_K |E[K, W]| - 2\text{OPT}\end{aligned}$$

Let  $T := V \setminus (U \cup W)$ , i.e., the set of vertices in connected components  $K$ . Using this notation, we can write the above equation simply as,

$$b \cdot |W| - b \cdot |U| - 2|E[W]| - |E[T, W]| > b \cdot (n - 2\text{OPT}) \quad (2)$$

Now consider the partition  $T, U, W$  in a *realized* graph  $G(V, E_p)$ . Let  $E_p[W]$  and  $E_p[T, W]$  denote, respectively, the set of edges in  $E[W]$  and  $E[T, W]$  after sampling the edges w.p.  $p$ . For any matching  $M$  in  $G_p$ , define  $x(M)$  to be the number of *unmatched* vertices (by  $M$ ) in  $W$ . Finally, define  $x^* := \min_M x(M)$ , where the minimum is taken over all matchings in  $G_p$ . Clearly,  $x^*$  is a random variable depending on the choice of edges in  $G_p$ . We have the following simple claim.

**Claim 4.2.** *For any realization  $G_p$ ,  $x^* \geq |W| - |U| - 2|E_p[W]| - |E_p[T, W]|$ .*

*Proof.* Consider the set of vertices in  $W$ . At most  $|U|$  vertices of  $W$  can be matched to vertices in  $U$ . Additionally, any edge in  $E_p[W]$  can further reduce the number of unmatched vertices in  $W$  by at most 2. Finally, any edge in  $E_p[T, W]$  can reduce the number of remaining unmatched vertices in  $W$  by at most 1. ■

Using the fact that  $\mathbb{E}[x^*] \leq n - 2\text{OPT}$ , we have,

$$\begin{aligned} b \cdot (n - 2\text{OPT}) &\geq b \cdot \mathbb{E}[x^*] \\ &\geq b \cdot |W| - b \cdot |U| - b \cdot \mathbb{E}\left[2|E_p[W]| + |E_p[T, W]|\right] \quad (\text{by Claim 4.2}) \\ &= b \cdot |W| - b \cdot |U| - pb \cdot (2|E[W]| + |E[T, W]|) \\ &\geq b \cdot |W| - b \cdot |U| - 2|E[W]| - |E[T, W]| \quad (\text{since } pb = p\left\lfloor \frac{1}{p} \right\rfloor \leq 1) \\ &> b \cdot (n - 2\text{OPT}) \quad (\text{by Eq (2)}) \end{aligned}$$

a contradiction. ■

We further prove that the bound established in Lemma 4.1 is essentially tight (see Appendix B).

**Claim 4.3.** *For any constant  $0 < p < 1$ , there exist bipartite graphs  $G$  where  $G_p$  has a matching of size  $n - o(n)$  in expectation, but for any  $b \geq \frac{2}{p}$ , there is no  $b$ -matching in  $G$  with (at least)  $b \cdot 0.99n$  edges; here  $n$  is the number of vertices on each side of  $G$ .*

Finally, we establish the following auxiliary lemma.

**Lemma 4.4.** *Let  $B$  be a  $\left\lfloor \frac{1}{p} \right\rfloor$ -matching with  $\left( \left\lfloor \frac{1}{p} \right\rfloor \cdot N \right)$  edges; then,  $\mathbb{E}[\mu(B_p)] \geq (1 - 3p) \cdot \frac{N}{3}$ .*

*Proof.* We first partition the edges of  $B$  into a collection of matchings. Since the degree of each vertex in  $G(V, B)$  is at most  $\left\lfloor \frac{1}{p} \right\rfloor$ , by Vizing's Theorem [35], we can color the edges in  $G(V, B)$  with  $\left\lfloor \frac{1}{p} \right\rfloor + 1$  colors such that no two edges with the same color are incident on a vertex. This ensures that  $B$  can be decomposed into  $R = \left\lfloor \frac{1}{p} \right\rfloor + 1$  matchings  $M_1, \dots, M_R$ .

Next, we define the following process. Define  $M^{(0)} = \emptyset$ ; for  $i = 1$  to  $R$  rounds, let  $M^{(i)}$  be a maximal matching obtained by adding to  $M^{(i-1)}$  the set of realized edges in  $M_i$  that are not incident on vertices in  $M^{(i-1)}$ . Define  $M := M^{(R)}$ .

We argue that  $\mathbb{E}[|M|] \geq (1 - 3p) \cdot \frac{N}{3}$ . To do this, we need the following notation. Define  $Y_i$  as a random variable denoting the *set of edges* in  $M_i$  that are *not* incident on any vertex of matching  $M^{(i-1)}$ . Note that  $Y_i$  depends only on the realization of edges in  $M_1, \dots, M_{i-1}$  and is *independent* of the realization of  $M_i$ . Moreover, define  $X_i$  as a random variable indicating the *number of edges* in (a realization of)  $Y_i$  that are added to  $M^{(i-1)}$  (after updating by edges in  $M_i$ ). We first have,

$$|Y_i| \geq |M_i| - 2|M^{(i-1)}|$$

since any edge in  $M^{(i-1)}$  can be incident on at most two vertices of  $M_i$ . Moreover, conditioned on any valuation for  $Y_i$ , we have  $\mathbb{E}[X_i] = p \cdot |Y_i|$  since each edge in  $M_i$  is realized w.p.  $p$ , independent of the choice of  $Y_i$ . Consequently,

$$\mathbb{E}[X_i] = p \cdot \mathbb{E}[Y_i] \geq p \cdot \left( |M_i| - 2\mathbb{E}\left[ |M^{(i-1)}| \right] \right)$$

We again stress that the expectation for  $X_i$  is taken over the choice of edges in  $M_i$ , while the expectation for  $Y_i$  (and  $M^{(i-1)}$ ) is taken over the choice of edges in  $M_1, \dots, M_{i-1}$ . We now have,

$$\begin{aligned} \mathbb{E}[|M|] &= \sum_{i=1}^R \mathbb{E}[X_i] \geq \sum_{i=1}^R p \cdot \left( |M_i| - 2\mathbb{E}\left[ |M^{(i-1)}| \right] \right) \\ &\geq p \cdot \left( \sum_{i=1}^R |M_i| - 2 \sum_{i=1}^R \mathbb{E}[|M^{(i-1)}|] \right) \quad (\mathbb{E}[|M|] \geq \mathbb{E}\left[ |M^{(i-1)}| \right]) \\ &\geq p \cdot \left( \left\lfloor \frac{1}{p} \right\rfloor \cdot N - 2 \left( \left\lfloor \frac{1}{p} \right\rfloor + 1 \right) \cdot \mathbb{E}[|M|] \right) \quad (R = \left\lfloor \frac{1}{p} \right\rfloor + 1) \end{aligned}$$

This implies that

$$\mathbb{E}[|M|] \geq (1 - 3p) \cdot \frac{N}{3}$$

which concludes the proof. ■

## 5 Main Algorithm and Analysis

We provide our main algorithm for the stochastic matching problem (when  $p$  is sufficiently small) in this section and prove Part (i) of Theorem 1. We assume throughout this section that the edge realization probability  $p \leq p_0$  for some sufficiently small constant  $p_0$ . In this case,  $\left\lfloor \frac{1}{p} \right\rfloor - 1 \geq (1 - O(p_0)) \cdot \frac{1}{p}$  and we use this inequality frequently in the proof. Indeed, throughout this section, one should view  $p_0$  as a negligible constant and hence the term  $(1 - O(p_0))$  can essentially be ignored.

Let  $\delta_0 = 0.02$ , and  $\varepsilon_0 = 0.02001$ . Our algorithm is stated as Algorithm 1 below:

---

**ALGORITHM 1:** A 0.52-Approximation Algorithm for Stochastic Matching

---

**Input:** A graph  $G(V, E)$  and an edge realization probability  $p \leq p_0$ .

**Output:** A subgraph  $H(V, Q)$  of  $G(V, E)$ .

1. Let  $B$  be a maximum  $\left\lfloor \frac{1}{p} \right\rfloor$ -matching in  $G$ .
  2. Let  $(M_1, M_2, \dots, M_R) := \text{MatchingCover}(G(V, E \setminus B), \varepsilon_1)$  for  $\varepsilon_1 = (\varepsilon_0 - \delta_0)/2$ , and  $E_{MC} = M_1 \cup \dots \cup M_R$ .
  3. Return  $H(V, Q)$  where  $Q := B \cup E_{MC}$ .
- 

Each vertex in  $H$  has degree  $O\left(\frac{\log(1/p)}{p}\right)$  – this follows immediately from Lemma 3.3. In what follows, we prove that  $H$  has a matching of size at least  $(0.5 + \delta_0) \cdot \text{OPT} = 0.52 \cdot \text{OPT}$  in expectation, which will complete the proof of Part (i) of Theorem 1.

First notice that if  $|M_R| < (\frac{1}{2} - \frac{\varepsilon_0 + \delta_0}{2})\text{OPT}$  where  $M_R$  is the smallest matching in the matching cover  $E_{MC}$  found by Algorithm 1, then by Claim 3.4, the expected matching size in  $Q$  is at least

$(\frac{1}{2} + \frac{\varepsilon_0 + \delta_0}{2} - \frac{\varepsilon_0 - \delta_0}{2})\text{OPT} = (\frac{1}{2} + \delta_0) \cdot \text{OPT}$ . Therefore, from now on we focus on the case that  $|M_R| \geq (\frac{1}{2} - \frac{\varepsilon_0 + \delta_0}{2})\text{OPT}$ .

In this case, by Lemma 3.3, w.p.  $1 - o(1)$ , there exists a matching  $M$  among the realized edges in  $E_{MC}$  with size at least

$$\begin{aligned} \left(1 - \frac{\varepsilon_0 - \delta_0}{2}\right) \left(\frac{1}{2} - \frac{\varepsilon_0 + \delta_0}{2}\right) \text{OPT} &\geq \left(\frac{1}{2} - \frac{\varepsilon_0 + \delta_0}{2} - \frac{\varepsilon_0 - \delta_0}{4}\right) \text{OPT} \\ &= \left(\frac{1}{2} - \frac{3\varepsilon_0}{4} - \frac{\delta_0}{4}\right) \text{OPT} \geq \left(\frac{1}{2} - \varepsilon_0\right) \text{OPT} \end{aligned}$$

In the following, we assume this event happens<sup>6</sup> and prove that the set of edges realized in the  $\left\lfloor \frac{1}{p} \right\rfloor$ -matching  $B$  can be used to augment the matching  $M$  to create a matching of size  $(\frac{1}{2} + \delta_0) \cdot \text{OPT}$  in expectation. To simplify the analysis, we assume w.l.o.g. that  $|M| = (\frac{1}{2} - \varepsilon_0)\text{OPT}$  (i.e., we only keep  $(\frac{1}{2} - \varepsilon_0)\text{OPT}$  edges of  $M$  and remove any additional edges if there is any). By the  $b$ -matching lemma (Lemma 4.1),  $|B| \geq \left(\left\lfloor \frac{1}{p} \right\rfloor - 1\right) \text{OPT} \geq (1 - O(p_0)) \cdot \frac{\text{OPT}}{p}$ , and hence, to prove Part (1) of Theorem 1, it suffices to prove the following statement.

**Lemma 5.1.** *Let  $M$  be a matching of size  $(\frac{1}{2} - \varepsilon_0)\text{OPT}$ , and  $B$  be a  $\left\lfloor \frac{1}{p} \right\rfloor$ -matching of size at least  $(1 - O(p_0)) \cdot \frac{\text{OPT}}{p}$ ; then the expected maximum matching size in  $M \cup B_{\overline{M}}$  is at least  $(\frac{1}{2} + \delta_0)\text{OPT}$ .*

*Proof.* Let  $B_M$  be the set of edges in  $B$  that are incident on the vertices in the matching  $M$ , and let  $B_{\overline{M}} = B \setminus B_M$ . Let  $s_{\overline{M}}$  be the random variable denoting the maximum matching size of a realization of  $B_{\overline{M}}$ . By Lemma 4.4,

$$\mathbb{E}[s_{\overline{M}}] \geq \frac{|B_{\overline{M}}|}{\left\lfloor \frac{1}{p} \right\rfloor} \cdot \frac{1 - 3p}{3} \geq p |B_{\overline{M}}| \cdot \frac{1 - 3p}{3} \quad (3)$$

Therefore, if  $|B_{\overline{M}}| \geq 6\varepsilon_0 \cdot \frac{\text{OPT}}{p}$ , then

$$\mathbb{E}[s_{\overline{M}}] \geq 6\varepsilon_0 \cdot \text{OPT} \cdot \frac{1 - 3p}{3} = 2\varepsilon_0(1 - 3p) \cdot \text{OPT} \geq (\varepsilon_0 + \delta_0) \cdot \text{OPT} \quad (\text{assuming } p_0 \leq \frac{\varepsilon_0 - \delta_0}{6\varepsilon_0})$$

and since no edge in  $B_{\overline{M}}$  is incident on the vertices in  $M$ , the expected matching size in  $M \cup B_{\overline{M}}$  is at least

$$\left(\frac{1}{2} - \varepsilon_0\right) \text{OPT} + (\varepsilon_0 + \delta_0) \text{OPT} = \left(\frac{1}{2} + \delta_0\right) \text{OPT}$$

as asserted by Lemma 5.1. In the following, we assume  $|B_{\overline{M}}| \leq 6\varepsilon_0 \cdot \frac{\text{OPT}}{p}$ . Furthermore, we fix a realization of  $B_{\overline{M}}$  and fix a maximum matching  $M'$  in the realization of  $B_{\overline{M}}$  (whose size is  $s_{\overline{M}}$  by definition). In other words, we will lower bound the expected maximum matching size in  $M \cup B_{\overline{M}}$  conditioned on *any* realization of  $B_{\overline{M}}$ . The lower bound we obtain would be a linear function of  $s_{\overline{M}}$ , and by linearity of expectation, we can simply replace  $s_{\overline{M}}$  with  $\mathbb{E}[s_{\overline{M}}]$ , use Eq (3) to lower bound  $\mathbb{E}[s_{\overline{M}}]$ , and obtain the desired lower bound of  $(1/2 + \delta_0)\text{OPT}$  on the expected maximum matching size.

Denote by  $M^+$  the matching  $M \cup M'$  (since the matchings  $M$  and  $M'$  are vertex-disjoint,  $M \cup M'$  is indeed a valid matching of size  $|M| + s_{\overline{M}}$ ). We can focus the realizations of  $B_{\overline{M}}$  where

---

<sup>6</sup>This assumption can be removed while losing a negligible factor of  $o(1)$  in the size of final matching.

$s_{\overline{M}} \leq 2\varepsilon_0 \cdot \text{OPT}$  since otherwise the matching  $M^+$  already have size  $(1/2 + \varepsilon_0) \cdot \text{OPT} > (1/2 + \delta_0) \cdot \text{OPT}$ . Therefore, we have  $s_{\overline{M}} \leq 2\varepsilon_0 \cdot \text{OPT} = O(\text{OPT})$  and  $|M^+| = O(\text{OPT})$ , which will be useful in simplifying the presentation.

Now consider the edges in  $B_M$ . We further denote by  $C$  the set of edges in  $B_M$  that are incident on *exactly* one vertex in  $M^+$ . In the following, we first show that  $|C|$  must be large (Claim 5.2) and then show that many edges in  $C$  can be used to augment the matching  $M^+$ , which leads to an increment on the matching size as a function of  $|C|$  (Lemma 5.3 and Lemma 5.4). Combining these two statements completes the proof of Lemma 5.1.

**Claim 5.2.**  $|C| \geq 2|B_M| - \frac{2|M^+|}{p}$ .

*Proof.* Let  $x$  denote the number of edges in  $B_M$  that have degree 2 to  $V(M^+)$  (i.e., are incident on two vertices in  $M^+$ ). By definition, every edge in  $B_M$  is incident on  $M$ , and hence every edge in  $B_M$  is also incident on  $M^+ (= M \cup M')$ . Consequently, there are  $|B_M| - x$  edges in  $B_M$  that have degree 1 to  $V(M^+)$  (i.e. belongs to  $C$ ). Therefore, the total degrees of all vertices  $V(M^+)$  provided by  $B_M$  is at least:  $2 \cdot x + 1 \cdot (|B_M| - x) = x + |B_M|$ .

On the other hand, since  $|V(M^+)| = 2|M^+|$  and  $B$  (hence  $B_M$ ) is a  $\left\lfloor \frac{1}{p} \right\rfloor$ -matching, the total degree of the vertices  $V(M^+)$  provided by  $B_M$  is at most  $\frac{2|M^+|}{p}$ . Therefore,  $x + |B_M| \leq \frac{2|M^+|}{p}$ , which implies  $x \leq \frac{2|M^+|}{p} - |B_M|$ . Therefore, the number of edges in  $B_M$  incident on exactly one vertex in  $V(M^+)$  (i.e.,  $|C|$ ) is at least

$$|B_M| - \left( \frac{2|M^+|}{p} - |B_M| \right) = 2|B_M| - \frac{2|M^+|}{p}$$

completing the proof. ■

The following two lemmas are dedicated to showing that the edges in a realization of  $C$ ,  $C_p$ , form many vertex-disjoint length-three augmenting paths for the matching  $M^+$  in expectation, which is a lower bound on the expected increment on the matching size. We first define some notation. Let  $W := V \setminus V(M^+)$ , i.e.,  $W$  is the set of vertices *not* matched by  $M^+$ . Denote the edges in  $M^+$  by  $\{(u_i, v_i) \mid i \in [|M^+|]\}$ , and denote by  $d(u_i)$  (resp.  $d(v_i)$ ) the number of edges in  $C$  incident on  $u_i$  (resp.  $v_i$ ). Since, by definition, the edges in  $C$  are only incident on one vertex in  $V(M^+)$ ,  $d(u_i)$  (resp.  $d(v_i)$ ) is also the number of edges in  $C$  between  $u_i$  (resp.  $v_i$ ) and  $W$ . In the following, whenever we say “neighbors” or “degrees”, they are only w.r.t. the edges  $C$ . Let  $f(x)$  be the function defined in Proposition 3.1.

**Lemma 5.3.** *For any edge  $(u_i, v_i) \in M$ , w.p. at least*

$$(1 - O(p_0)) \frac{f(1/e) \cdot p}{e^2} \left(1 - e^{-p \cdot d(v_i)}\right) \cdot \max \{d(u_i) - 1, 0\},$$

*there exists a length-three augmenting path  $a_i - u_i - v_i - b_i$  in the realization  $C_p$  of  $C$ , such that  $a_i, b_i$  have no neighbors other than  $u_i$  and  $v_i$ .*

Note that we can use all edges  $(u_i, v_i)$  in  $M^+$  with such an augmenting path  $a_i - u_i - v_i - b_i$  to (simultaneously) augment  $M^+$  since these augmenting paths are vertex-disjoint ( $a_i$  and  $b_i$  are only neighbors of  $u_i$  and  $v_i$ ). Therefore, the expected number of edges in  $M^+$  that has such an augmenting path is a lower bound on the expected increment on the matching size.

*Proof of Lemma 5.3.* We consider three disjoint subsets of edges in  $C$  one by one: (i) the edges between  $v_i$  and  $W$ , (ii) the edges incident on a specific vertex  $w$  in  $W$  (excluding the edge  $(v_i, w)$ ), and (iii) the edges incident on neighbors of  $u_i$  other than  $w$  (excluding the edges incident on  $v_i$ ).

First, consider the edges between  $v_i$  and  $W$ . The prob. that none of these  $d(v_i)$  edges are realized is at most

$$(1 - p)^{d(v_i)} \leq e^{-p \cdot d(v_i)}$$

Therefore, w.p. at least  $1 - e^{-p \cdot d(v_i)}$ , at least one edge between  $v_i$  and  $W$  is realized. We condition on this event and fix any such edge, denoted by  $(v_i, b_i)$ .

Second, consider the edges incident on  $b_i$  (excluding the edge  $(v_i, b_i)$ ). There are at most  $1/p$  such edges, and the prob. that none of them is realized is at least

$$(1 - p)^{1/p} \geq \frac{1 - p}{e} \geq \frac{1 - p_0}{e} \quad (p \leq p_0 \leq 0.43)$$

where the first inequality is by Proposition 3.2. In the following, we further condition on no other edges incident on  $w$  is realized.

Third, consider all neighbors of  $u_i$  other than  $b_i$  (there are at least  $\max\{d(u_i) - 1, 0\}$  such neighbors) and the edges incident on these neighbors (excluding the edges incident on  $v_i$ ). For each one of these neighbors  $w$  of  $u_i$ , the prob. that the edge  $(u_i, w)$  is realized (w.p.  $p$ ) and  $w$  does not have any neighbor other than  $u_i$  (and possibly  $v_i$ ) (w.p. at least  $\frac{1-p_0}{e}$  by Proposition 3.2) is at least  $p \cdot \frac{1-p_0}{e}$ . Therefore, the prob. that at least one neighbor of  $u_i$  satisfies these two properties is at least

$$\begin{aligned} 1 - \left(1 - p \cdot \frac{1 - p_0}{e}\right)^{\max\{d(u_i) - 1, 0\}} &\geq 1 - e^{-(1 - p_0) \cdot p \cdot \max\{d(u_i) - 1, 0\}/e} \\ &\geq f\left(\frac{1}{e}\right) \cdot (1 - p_0) \cdot p \cdot \max\{d(u_i) - 1, 0\} / e \end{aligned}$$

where  $f(x) = \frac{1-e^{-x}}{x}$  and the second inequality is by Proposition 3.1, using the fact that

$$\frac{(1 - p_0) \cdot p \max\{d(u_i) - 1, 0\}}{e} \leq \frac{1}{e} \quad (\text{since } d(u_i) \leq \frac{1}{p})$$

Putting the three steps together, the prob. that there is an augmenting path  $a_i - u_i - v_i - b_i$  where  $a_i$  and  $b_i$  has no neighbors other than  $u_i$  and  $v_i$  is at least

$$\begin{aligned} &\left(1 - e^{-p \cdot d(v_i)}\right) \cdot \frac{1 - p_0}{e} \cdot \frac{f\left(\frac{1}{e}\right) \cdot (1 - p_0) \cdot p \cdot \max\{d(u_i) - 1, 0\}}{e} \\ &= (1 - O(p_0)) \cdot f\left(\frac{1}{e}\right) \cdot \frac{p}{e^2} \left(1 - e^{-p \cdot d(v_i)}\right) \cdot \max\{d(u_i) - 1, 0\} \end{aligned}$$

■

As we pointed out after the statement of Lemma 5.3, we need to lower bound the expected number of edges in  $M^+$  that has such an augmenting path, which, by Lemma 5.3, is lower bounded by the function  $F$  defined below. For the two vectors  $d_u := (d(u_1), \dots, d(u_{|M^+|}))$  and  $d_v := (d(v_1), \dots, d(v_{|M^+|}))$ ,

$$F(d_u, d_v) := \sum_{i \in [|M^+|]} (1 - O(p_0)) \cdot f\left(\frac{1}{e}\right) \cdot \frac{p}{e^2} \left(1 - e^{-p \cdot d(v_i)}\right) \cdot \max\{d(u_i) - 1, 0\}$$

The goal now is to find the smallest value of  $F(d_u, d_v)$ , with the constraint on the vectors  $d_u$  and  $d_v$  formulated in the following (non-linear) minimization program (referred to as MP-(4)).

$$\begin{aligned} & \text{minimize} && F(d_u, d_v) \\ & \text{subject to} && \sum_{i \in [|M^+|]} d(u_i) + d(v_i) = |C| \\ & && d(u_i), d(v_i) \in \left[ \left\lfloor \frac{1}{p} \right\rfloor \right] \quad i = 1, \dots, |M^+| \end{aligned} \tag{4}$$

The constraint on each individual  $d(u_i)$  and  $d(v_i)$  is because  $C$  is a  $\left\lfloor \frac{1}{p} \right\rfloor$ -matching. The following lemma lower bounds the value of the objective function in MP-(4).

**Lemma 5.4.** *Let  $F^*$  denote the optimal value of MP-(4); then,*

$$F^* \geq (p \cdot |C| - |M^+|) \cdot \eta - O(p_0) \cdot \text{OPT}$$

$$\text{where } \eta := f\left(\frac{1}{e}\right) \cdot \frac{1}{e^2} (1 - e^{-1}) > 0.07157.$$

The proof of the Lemma 5.4 is technical, and we defer it to Section 5.1. By Lemma 5.4 and Claim 5.2 (the lower bound on  $|C|$ ) the expected increment (over  $M^+$ ) of the matching size is at least

$$\begin{aligned} F^* &\geq (p \cdot |C| - |M^+|) \cdot \eta - O(p_0) \cdot \text{OPT} \\ &\geq (p \cdot (2|B_M| - 2|M^+|/p) - |M^+|) \cdot \eta - O(p_0) \cdot \text{OPT} \quad (\text{By Claim 5.2}) \\ &= (2p|B_M| - 3|M^+|) \cdot \eta - O(p_0) \cdot \text{OPT} \\ &= (2p(|B| - |B_{\bar{M}}|) - 3(|M| + s_{\bar{M}})) \cdot \eta - O(p_0) \cdot \text{OPT} \\ &= (2p|B| - 3|M| - 2p|B_{\bar{M}}| - 3s_{\bar{M}}) \cdot \eta - O(p_0) \cdot \text{OPT} \\ &= \left( 2p(1 - O(p_0)) \frac{\text{OPT}}{p} - 3\left(\frac{1}{2} - \varepsilon_0\right) \text{OPT} - 2p|B_{\bar{M}}| - 3s_{\bar{M}} \right) \cdot \eta - O(p_0)\text{OPT} \\ &= \left( \left(\frac{1}{2} + 3\varepsilon_0\right) \text{OPT} - 2p|B_{\bar{M}}| - 3s_{\bar{M}} \right) \cdot \eta - O(p_0) \cdot \text{OPT} \end{aligned}$$

Since the original matching  $M^+$  is of size  $(1/2 - \varepsilon_0) \cdot \text{OPT} + s_{\bar{M}}$ , the expected matching size in  $M \cup B_p$ , i.e.,  $\mu(M \cup B_p)$ , is

$$\begin{aligned} \mathbb{E}[\mu(M \cup B_p)] &= \sum_{s_{\bar{M}}} \Pr(s_{\bar{M}}) \mathbb{E}[\mu(M \cup B_p) \mid s_{\bar{M}}] \quad (\mathbb{E}[X] = \sum_Y \Pr(Y) \mathbb{E}[X \mid Y]) \\ &\geq \sum_{s_{\bar{M}}} \Pr(s_{\bar{M}}) (|M^+| + F^*) \\ &\geq \left(\frac{1}{2} - \varepsilon_0\right) \cdot \text{OPT} + \mathbb{E}[s_{\bar{M}}] + \left( \left(\frac{1}{2} + 3\varepsilon_0\right) \text{OPT} - 2p|B_{\bar{M}}| - 3\mathbb{E}[s_{\bar{M}}] \right) \cdot \eta - O(p_0) \cdot \text{OPT} \\ &\geq \left(\frac{1}{2} + \frac{\eta}{2} - \varepsilon_0 + 3\eta\varepsilon_0\right) \cdot \text{OPT} + (1 - 3\eta)\mathbb{E}[s_{\bar{M}}] - 2p\eta|B_{\bar{M}}| - O(p_0) \cdot \text{OPT} \\ &\geq \left(\frac{1}{2} + \frac{\eta}{2} - \varepsilon_0 + 3\eta\varepsilon_0\right) \cdot \text{OPT} + (1 - 3\eta)(1 - O(p_0)) \frac{p|B_{\bar{M}}|}{3} - 2p\eta|B_{\bar{M}}| - O(p_0) \cdot \text{OPT} \quad (\text{By Equation 3}) \\ &\geq \left(\frac{1}{2} + \frac{\eta}{2} - \varepsilon_0 + 3\eta\varepsilon_0\right) \cdot \text{OPT} + \left(\frac{1}{3} - 3\eta\right) p|B_{\bar{M}}| - O(p_0) \cdot \text{OPT} \end{aligned}$$

Since  $\eta \approx 0.07157$ ,  $\frac{1}{3} - 3\eta > 0$ , and we have

$$\begin{aligned}
& \left( \frac{1}{2} + \frac{\eta}{2} - \varepsilon_0 + 3\eta\varepsilon_0 \right) \cdot \text{OPT} + \left( \frac{1}{3} - 3\eta \right) p |B_{\overline{M}}| - O(p_0) \cdot \text{OPT} \\
& \geq \left( \frac{1}{2} + \frac{\eta}{2} - \varepsilon_0 + 3\eta\varepsilon_0 \right) \cdot \text{OPT} - O(p_0) \cdot \text{OPT} \\
& > 0.52 \cdot \text{OPT} \quad (\varepsilon_0 = 0.02001, \eta > 0.07157, \text{ and } p_0 \text{ is sufficiently small.}) \\
& = (1/2 + \delta_0) \cdot \text{OPT}
\end{aligned}$$

completing the proof of Lemma 5.1.

### 5.1 Lower Bounding the Value of MP-(4)

In this section, we prove Lemma 5.4, i.e., the following inequality,

$$F^* (= \min F(d_u, d_v)) \geq (p \cdot |C| - |M^+|) \cdot \eta - O(p_0) \text{OPT}$$

where  $\eta := f(\frac{1}{e}) \cdot \frac{1}{e^2} (1 - e^{-1})$ .

Recall that

$$\begin{aligned}
F(d_u, d_v) &= \sum_i (1 - O(p_0)) \cdot f(\frac{1}{e}) \cdot \frac{p}{e^2} \left( 1 - e^{-p \cdot d(v_i)} \right) \cdot \max \{d(u_i) - 1, 0\} \\
&= (1 - O(p_0)) \cdot f(\frac{1}{e}) \cdot \frac{p}{e^2} \sum_i \left( 1 - e^{-p \cdot d(v_i)} \right) \cdot \max \{d(u_i) - 1, 0\}
\end{aligned}$$

Since the term  $(1 - O(p_0)) \cdot f(\frac{1}{e}) \cdot \frac{p}{e^2}$  is independent of  $d_u$  and  $d_v$ ,

$$\arg \min_{d_u, d_v} F = \arg \min_{d_u, d_v} \sum_i \left( 1 - e^{-p \cdot d(v_i)} \right) \cdot \max \{d(u_i) - 1, 0\}$$

Define  $d(V) := \sum_i d(v_i)$  and  $d(U) := \sum_i d(u_i)$ ; then,  $d(V) + d(U) = |C|$ . We need to prove that for any choice of  $d(V)$  and  $d(U)$ , the lemma statement holds. First of all, we can assume  $d(U) \geq |M^+|$ : otherwise, since  $d(V) \leq \left\lfloor \frac{1}{p} \right\rfloor |M^+|$ , we will have

$$|C| = d(V) + d(U) \leq \left\lfloor \frac{1}{p} \right\rfloor |M^+| + |M^+| \leq \left( \frac{1}{p} + 1 \right) |M^+|$$

Therefore, for the target lower bound on  $F^*$

$$\begin{aligned}
& (p \cdot |C| - |M^+|) \cdot \eta - O(p_0) \text{OPT} \\
& \leq \left( p \left( \frac{1}{p} + 1 \right) |M^+| - |M^+| \right) \cdot \eta - O(p_0) \text{OPT} \\
& \leq p\eta |M^+| - O(p_0) \text{OPT} \\
& \leq p_0\eta |M^+| - O(p_0) \text{OPT}
\end{aligned}$$

which can be made negative by choosing the constant hidden in  $O(p_0)$  to be 1, proving Lemma 5.4.

We further assume  $d(U) - |M^+|$  is an integer multiple of  $\left\lfloor \frac{1}{p} \right\rfloor - 1$  and  $d(V)$  is an integer multiple of  $\left\lfloor \frac{1}{p} \right\rfloor$ . This can be achieved by removing at most  $1/p$  edges from  $d(U)$  and  $d(V)$  respectively.

Since  $F$  is monotonically increasing with any  $d(u_i)$  or  $d(v_i)$ , removing edges from  $d(U)$  and  $d(V)$  can only make  $F^*$  even smaller. Therefore, if we show that after removing these edges, the target lower bound on  $F^*$  holds, then it definitely holds for the original  $d(U)$  and  $d(V)$ . In the following, we fix any  $d(U)$  and  $d(V)$  where  $d(U) - |M^+|$  is an integer multiple of  $\left\lfloor \frac{1}{p} \right\rfloor - 1$ ,  $d(V)$  is an integer multiple of  $\left\lfloor \frac{1}{p} \right\rfloor$ , and  $d(V) + d(U) \geq |C| - \frac{2}{p}$ . We prove the following key property of  $F(d_u, d_v)$ .

**Lemma 5.5.** *There exists  $d_u$  and  $d_v$  that minimizes  $F(d_u, d_v)$  where any entry in  $d_u$  is either 1 or  $\left\lfloor \frac{1}{p} \right\rfloor$  and any entry in  $d_v$  is either 0 or  $\left\lfloor \frac{1}{p} \right\rfloor$ .*

We first show why Lemma 5.5 implies the target lower bound on  $F^*$ , and then prove Lemma 5.5. Fix  $d^*(u_i)$  and  $d^*(v_i)$  that satisfy the property in Lemma 5.5. Since every entry in  $d_u^*$  is either 1 or  $\left\lfloor \frac{1}{p} \right\rfloor$ , the number of 1's in  $d_u^*$  is  $x := |M^+| - (d(U) - |M^+|)/(\left\lfloor \frac{1}{p} \right\rfloor - 1)$ . Similarly, the number of 0's in  $d_v^*$  is  $y := |M^+| - d(V)/\left\lfloor \frac{1}{p} \right\rfloor$ . Therefore, the number of edges in  $M^+$  where  $d^*(v_i) = d^*(u_i) = \left\lfloor \frac{1}{p} \right\rfloor$  is at least

$$\begin{aligned}
|M^+| - x - y &= |M^+| - \left( |M^+| - \frac{d(U) - |M^+|}{\left\lfloor \frac{1}{p} \right\rfloor - 1} \right) - \left( |M^+| - \frac{d(V)}{\left\lfloor \frac{1}{p} \right\rfloor} \right) \\
&= \frac{d(U) - |M^+|}{\left\lfloor \frac{1}{p} \right\rfloor - 1} - |M^+| + \frac{d(V)}{\left\lfloor \frac{1}{p} \right\rfloor} \\
&\geq \frac{d(U) - |M^+|}{\frac{1}{p}} - |M^+| + d(V)p && (\left\lfloor \frac{1}{p} \right\rfloor - 1 \leq \left\lfloor \frac{1}{p} \right\rfloor \leq \frac{1}{p}) \\
&= p \cdot d(U) + p \cdot d(V) - (1 + p) |M^+| \\
&\geq p \left( |C| - \frac{2}{p} \right) - (1 + p) |M^+| && (d(V) + d(U) \geq |C| - \frac{2}{p}..) \\
&= p |C| - |M^+| - O(p_0) \text{OPT} && (|M^+| = O(\text{OPT}))
\end{aligned}$$

And just focusing on these  $p |C| - |M^+| - O(p_0) \text{OPT}$  edges, we have

$$\begin{aligned}
F^* &\geq (p |C| - |M^+| - O(p_0) \text{OPT}) \cdot (1 - O(p_0)) \cdot f\left(\frac{1}{e}\right) \cdot \frac{p}{e^2} (1 - e^{-1}) \cdot (\left\lfloor \frac{1}{p} \right\rfloor - 1) \\
&\geq (p |C| - |M^+| - O(p_0) \text{OPT}) \cdot (1 - O(p_0)) \cdot f\left(\frac{1}{e}\right) \cdot \frac{p}{e^2} (1 - e^{-1}) \cdot \frac{1}{p} \quad (\left\lfloor \frac{1}{p} \right\rfloor - 1 \geq (1 - O(p_0)) \frac{1}{p}) \\
&\geq (p |C| - |M^+| - O(p_0) \text{OPT}) \cdot (1 - O(p_0)) \cdot \eta \\
&\geq (p |C| - |M^+|) \cdot \eta - O(p_0) \text{OPT} && (|M^+| = O(\text{OPT}), p |C| \leq p |B_M| = O(\text{OPT}))
\end{aligned}$$

which proves Lemma 5.4.

We now prove Lemma 5.5 which will complete the proof.

*Proof of Lemma 5.5.* Fix any allocation  $d_u^*$  and  $d_v^*$  that minimizes  $F(d_u, d_v)$ . We will show that first, there exists a sequence of locally reallocating the values (i.e., degrees) in  $d_u^*$  without changing the value of  $F(d_u, d_v)$  such that at the end, every entry in  $d_u^*$  is either 1 or  $\left\lfloor \frac{1}{p} \right\rfloor$ . After changing the vector  $d_u^*$ , we then show that there exists a sequence of locally reallocating the values in  $d_v^*$  without changing the value of  $F(d_u, d_v)$  such that at the end, every entry in  $d_v^*$  is either 0 or  $\left\lfloor \frac{1}{p} \right\rfloor$ .

We first explain how to change  $d_u^*$ . To simplify the presentation, we define  $q_i = 1 - e^{-p \cdot d^*(v_i)}$  and the target expression becomes

$$\sum_i \left(1 - e^{-p \cdot d^*(v_i)}\right) \cdot \max \{d^*(u_i) - 1, 0\} = \sum_i q_i \cdot \max \{d^*(u_i) - 1, 0\} \quad (5)$$

Recall that  $d^*(u_i)$  satisfies  $d^*(u_i) \in \left[\left\lfloor \frac{1}{p} \right\rfloor\right]$  (in MP-(4)) (and hence  $q_i \geq 0$ ) and  $\sum_i d^*(u_i) = d(U)$ . First of all, if there exists some  $i_1$  where  $d^*(u_{i_1}) = 0$ , then since  $d(U) \geq |M^+|$ , there must exist an index  $i_2$  where  $d^*(u_{i_2}) \geq 2$ . Then, we can shift one degree from  $d^*(u_{i_2})$  to  $d^*(u_{i_1})$  and after the shift, (a)  $\max \{d^*(u_{i_1}) - 1, 0\}$  remains 0 and hence  $q_{i_1} \max \{d^*(u_i) - 1, 0\}$  remains 0, and (b)  $\max \{d^*(u_{i_2}) - 1, 0\}$  decreases and hence  $q_{i_2} \max \{d^*(u_{i_2}) - 1, 0\}$  does not increase. Therefore,  $F(d_u, d_v)$  does not increase after the shift, and from now on, we have  $d^*(u_i) \geq 1$  for all  $i \in [|M^+|]$ . To proceed, we need the following property of  $d^*(u_i)$ .

**Claim 5.6.** *For any pair of indices  $i_1, i_2$  where  $q_{i_1} > q_{i_2}$ , either  $d^*(u_{i_2}) = \left\lfloor \frac{1}{p} \right\rfloor$  or  $d^*(u_{i_1}) \leq 1$ .*

*Proof.* Suppose not. We have  $d^*(u_{i_2}) < \left\lfloor \frac{1}{p} \right\rfloor$  and  $d^*(u_{i_1}) > 1$ , for some  $i_1$  and  $i_2$ . We can shift one degree from  $d^*(u_{i_1})$  to  $d^*(u_{i_2})$  and still get a valid allocation. In the following, we show that this new allocation achieves a smaller value of  $F$ , which contradicts to the optimality of  $d^*(u_i)$ .

Since shifting from  $d^*(u_{i_1})$  to  $d^*(u_{i_2})$  only changes the degrees for  $u_{i_1}$  and  $u_{i_2}$ , it suffices for us to prove that

$$\begin{aligned} \Delta := & (q_{i_2} \max \{d^*(u_{i_2}) - 1, 0\} + q_{i_1} \max \{d^*(u_{i_1}) - 1, 0\}) \\ & - (q_{i_2} \max \{d^*(u_{i_2}), 0\} + q_{i_1} \max \{d^*(u_{i_1}) - 2, 0\}) > 0 \end{aligned}$$

Since  $d^*(u_{i_1}) \geq 2$ ,  $\max \{d^*(u_{i_1}) - 1, 0\} = d^*(u_{i_1}) - 1$ ,  $\max \{d^*(u_{i_1}) - 2, 0\} = d^*(u_{i_1}) - 2$ . In addition, since  $d^*(u_{i_2}) \geq 0$ ,  $\max \{d^*(u_{i_2}), 0\} = d^*(u_{i_2})$ . We have

$$\begin{aligned} \Delta = & (q_{i_2} \max \{d^*(u_{i_2}) - 1, 0\} + q_{i_1} (d^*(u_{i_1}) - 1)) - (q_{i_2} d^*(u_{i_2}) + q_{i_1} (d^*(u_{i_1}) - 2)) \\ = & q_{i_2} (\max \{d^*(u_{i_2}) - 1, 0\} - d^*(u_{i_2})) + q_{i_1} \\ \geq & q_{i_2} (d^*(u_{i_2}) - 1 - d^*(u_{i_2})) + q_{i_1} \\ = & q_{i_1} - q_{i_2}. \end{aligned}$$

Since we have  $q_{i_1} > q_{i_2}$ , the value of  $F$  decreases after the shifting according to Eq 5, a contradiction. ■

We use Claim 5.6 to prove the correctness of the following sequence of reallocation of  $d_u^*$ . Now, as long as there exists an index  $i_1$ , where  $d^*(u_{i_1}) \in (1, \left\lfloor \frac{1}{p} \right\rfloor)$ , since  $d(U) - |M^+|$  is an integer multiple of  $(\left\lfloor \frac{1}{p} \right\rfloor - 1)$ , there must exist some index  $i_2$  where  $d^*(u_{i_2}) \in (1, \left\lfloor \frac{1}{p} \right\rfloor)$  (recall that  $d^*(u_i) \geq 1$ ), and we will shift the values between  $d^*(u_{i_1})$  and  $d^*(u_{i_2})$  such that one of them becomes either 1 or  $\left\lfloor \frac{1}{p} \right\rfloor$  and both of them are still at least 1 (it is easy to see this is always possible). First of all, every step of this reallocation reduces the number of vertices with  $d^*(u_i) \in (1, \left\lfloor \frac{1}{p} \right\rfloor)$ , and hence it will terminate. To see that this process never changes  $F(d_u, d_v)$ , (a) it cannot be that  $q_{i_1} \neq q_{i_2}$ , since otherwise the indices  $i_1$  and  $i_2$  will contradict Claim 5.6, and (b) if  $q_{i_1} = q_{i_2}$ , shifting the allocation between  $i_1$  to  $i_2$  will not change  $F(d_u, d_v)$ .

Therefore, we can focus on the case where the entries of  $d^*(u_i)$  are either 1 or  $\left\lfloor \frac{1}{p} \right\rfloor$ . We now consider  $d^*(v_i)$ . Recall that  $d(V) = \sum_i d(v_i)$  and  $d(V)$  is an integer multiple of  $\left\lfloor \frac{1}{p} \right\rfloor$ . The target expression can be written as

$$\begin{aligned} & \sum_i \left(1 - e^{-p \cdot d^*(v_i)}\right) \cdot \max \{d^*(u_i) - 1, 0\} \\ &= \sum_{i: d^*(u_i)=\left\lfloor \frac{1}{p} \right\rfloor} \left(1 - e^{-p \cdot d^*(v_i)}\right) \left(\left\lfloor \frac{1}{p} \right\rfloor - 1\right) + \sum_{i: d^*(u_i)=1} \left(1 - e^{-p \cdot d^*(v_i)}\right) \cdot 0 \\ &= \sum_{i: d^*(u_i)=\left\lfloor \frac{1}{p} \right\rfloor} \left(1 - e^{-p \cdot d^*(v_i)}\right) \end{aligned}$$

Since  $F$  is monotonically increasing when any  $d(u_i)$  increases, for the indices  $i$  where  $d^*(u_i) = 1$ , ideally, one should allocate as many degrees to  $d^*(v_i)$  as possible, i.e.,  $d^*(v_i) = \left\lfloor \frac{1}{p} \right\rfloor$ . However, it might be the case that  $d(V)$  cannot supply  $\left\lfloor \frac{1}{p} \right\rfloor$  degree for all  $i$  where  $d^*(u_i) = 1$ . But in this case, we are done since reallocating between different  $d^*(v_i)$  where  $d^*(u_i) = 1$  does not change the  $F$  (in fact,  $F$  is always 0), and we can shift them such that we have as many  $d^*(v_i) = \left\lfloor \frac{1}{p} \right\rfloor$  as possible and leave the rest equal to 0.

In the following, we assume  $d(V)$  can supply  $\left\lfloor \frac{1}{p} \right\rfloor$  degree for all  $i$  with  $d^*(u_i) = 1$ , and hence  $d^*(v_i) = \left\lfloor \frac{1}{p} \right\rfloor$  whenever  $d^*(u_i) = 1$ . It suffices to only focus on  $d^*(v_i)$  where  $d^*(u_i) = \left\lfloor \frac{1}{p} \right\rfloor$ . We need the following property of  $d^*(v_i)$  to complete the argument.

**Claim 5.7.** *For any pair of indices  $i_1$  and  $i_2$  such that  $d^*(u_{i_1}) = d^*(u_{i_2}) = \left\lfloor \frac{1}{p} \right\rfloor$ , we have that either  $\min \{d^*(v_{i_1}), d^*(v_{i_2})\} = 0$  or  $\max \{d^*(v_{i_1}), d^*(v_{i_2})\} = \left\lfloor \frac{1}{p} \right\rfloor$ .*

*Proof.* Suppose not. Then, for some  $i_1$  and  $i_2$ , we have  $\min \{d^*(v_{i_1}), d^*(v_{i_2})\} > 0$ , and also  $\max \{d^*(v_{i_1}), d^*(v_{i_2})\} < \left\lfloor \frac{1}{p} \right\rfloor$ . Without loss of generality, assume  $1 \leq d^*(v_{i_1}) \leq d^*(v_{i_2}) \leq \left\lfloor \frac{1}{p} \right\rfloor - 1$ . Then, shifting one degree from  $d^*(v_{i_1})$  to  $d^*(v_{i_2})$  leads to a valid allocation, and we prove in the following that the new allocation decreases the objective function which contradicts the optimality of  $d_v^*$ .

Since only the indices  $i_1$  and  $i_2$  are affected, it suffices for us to prove that

$$\Delta := \left(1 - e^{-p \cdot d^*(v_{i_1})}\right) + \left(1 - e^{-p \cdot d^*(v_{i_2})}\right) - \left(1 - e^{-p \cdot (d^*(v_{i_1})-1)}\right) + \left(1 - e^{-p \cdot (d^*(v_{i_2})+1)}\right) > 0$$

We have

$$\Delta \geq e^{-p \cdot d^*(v_{i_1})}(e^p - 1) + e^{-p \cdot d^*(v_{i_2})}(e^{-p} - 1)$$

Since  $d^*(v_{i_1}) \leq d^*(v_{i_2})$ ,  $e^{-p \cdot d^*(v_{i_1})} \geq e^{-p \cdot d^*(v_{i_2})}$ . We further have

$$\begin{aligned} e^{-p \cdot d^*(v_{i_1})}(e^p - 1) + e^{-p \cdot d^*(v_{i_2})}(e^{-p} - 1) &\geq e^{-p \cdot d^*(v_{i_2})}(e^p - 1) + e^{-p \cdot d^*(v_{i_2})}(e^{-p} - 1) \\ &\geq e^{-p \cdot d^*(v_{i_2})}(e^p - 1 + e^{-p} - 1) \\ &> e^{-p \cdot d^*(v_{i_2})}(2\sqrt{e^p \cdot e^{-p}} - 2) \\ &= 0 \end{aligned}$$

where the strict inequality is true since the two terms can only be equal when  $e^p = e^{-p}$  which does not happen for  $p > 0$ .  $\blacksquare$

Using Claim 5.7, we can now show that any  $d^*(v_i)$  where  $d^*(u_i) = \left\lfloor \frac{1}{p} \right\rfloor$ , must be either 0 or  $\left\lfloor \frac{1}{p} \right\rfloor$ . Suppose not. If  $d^*(v_{i_1}) \in (0, \left\lfloor \frac{1}{p} \right\rfloor)$ , then since  $d(V)$  is an integer multiple of  $\left\lfloor \frac{1}{p} \right\rfloor$ , there must exist some other index  $i_2$  where  $d^*(v_{i_2}) \in (0, \left\lfloor \frac{1}{p} \right\rfloor)$ ; hence,  $0 < \min \{d^*(v_{i_1}), d^*(v_{i_2})\} < \max \{d^*(v_{i_1}), d^*(v_{i_2})\} < \left\lfloor \frac{1}{p} \right\rfloor$ , a contradiction to Claim 5.7.  $\blacksquare$

## 6 An Algorithm for Large Values of $p$

In this section, we provide an algorithm, namely, Algorithm 2, with approximation ratio strictly better than  $1/2$  when  $p$  is bounded away from zero. In particular, this algorithm computes a matching of size  $(1/2 + \Theta(p^2)) \cdot \text{OPT}$ . Algorithm 2 is required to handle the case when  $p$  is not small enough for Algorithm 1 to perform well. Using a combination of both of these algorithms, we can prove the second part of Theorem 1.

Let  $p_0$  be any fixed constant independent of  $n$ ,  $\delta = \frac{p^2}{4}$ , and  $\varepsilon = \frac{p_0^2}{10^4}$ . The new algorithm (i.e, Algorithm 2) is similar to Algorithm 1 with the only difference being that instead of a  $\left\lfloor \frac{1}{p} \right\rfloor$ -matching, here, we simply pick a single maximum matching in  $G$ . Our algorithm is stated as Algorithm 2.

---

### ALGORITHM 2: A $(0.5 + \Theta(p^2))$ -Approximation Algorithm for Stochastic Matching

---

**Input:** A graph  $G(V, E)$  and an edge realization probability  $p_0 \leq p < 1$ .

**Output:** A subgraph  $H(V, Q)$  of  $G(V, E)$ .

1. Let  $M$  be a maximum matching in  $G$ .
  2. Let  $(M_1, M_2, \dots, M_R) := \text{MatchingCover}(G(V, E \setminus M), \varepsilon)$  (recall that  $\varepsilon = \frac{p_0^2}{10^4}$ ), and  $E_{MC} = M_1 \cup \dots \cup M_R$ .
  3. Return  $H(V, Q)$  where  $Q := M \cup E_{MC}$ .
- 

The following lemma proves the approximation ratio of Algorithm 2.

**Lemma 6.1.** *For any constant  $p_0 > 0$ , any realization probability  $p \geq p_0$ , and any graph  $G(V, E)$  the expected maximum matching size in the graph  $H$  computed by Algorithm 2 is at least  $\left(\frac{1}{2} + \frac{p^2}{4} - \frac{p_0^2}{10^4}\right) \cdot \text{OPT}(G)$ .*

Before proving Lemma 6.1, we show how to combine Algorithm 1 and Algorithm 2 to prove Part (ii) of Theorem 1.

*Proof of Theorem 1, Part (ii).* Let  $p_0$  be the constant such that Algorithm 1 achieves an approximation ratio of 0.52 for any  $p \leq p_0$ . The algorithm for Part (ii) is simply as follows. If the realization probability  $p \leq p_0$ , run Algorithm 1 and otherwise run Algorithm 2. By Lemma 6.1, the approximation ratio of this algorithm is  $\min \left\{ 0.52, \frac{1}{2} + \frac{p^2}{4} - \frac{p_0^2}{10^4} \right\} = 0.5 + \delta_0$  for some absolute constant  $\delta_0$  (since  $p_0$  is an absolute constant and  $p \geq p_0$ ).

We note that by optimizing the choice of  $p_0$  and a more careful analysis of Algorithm 1 (to account for many constants involved), one can bound the value of  $\delta_0 \approx 0.001$ . We omit the tedious details of this calculation as it is not the main contribution of this paper. ■

We now prove Lemma 6.1.

*Proof of Lemma 6.1.* Recall that  $\text{OPT}$  (resp.  $\text{ALG}$ ) is the expected maximum matching size in a realization  $G_p$  of  $G$  (resp. a realization  $H_p$  of  $H$ ).

Firstly, by Claim 3.4, with the parameters  $\varepsilon, \delta$ , and  $X = M$ , we have that if  $|M_R|$  in  $E_{MC}$  is smaller than  $(\frac{1}{2} - \delta) \cdot \text{OPT}$ , then the expected matching size in  $G(V, Q)$  is at least  $(\frac{1}{2} + \delta - \varepsilon) \cdot \text{OPT} = (\frac{1}{2} + \frac{p^2}{4} - \frac{p_0^2}{104}) \cdot \text{OPT}$ , which proves the lemma. We now consider the case where  $|M_R| \geq (\frac{1}{2} - \delta) \cdot \text{OPT}$ .

Let  $M'$  be the random variable denotes a maximum matching in a realization of  $E_{MC}$  (breaking tie arbitrarily). By Lemma 3.3, w.h.p.,  $|M'| \geq (1 - \varepsilon) |M_R| \geq (\frac{1}{2} - \delta - \varepsilon) \cdot \text{OPT}$ . For simplicity, in the following, we always assume this event happens<sup>7</sup> and further remove any extra edges in  $M'$  so that  $|M'| = (\frac{1}{2} - \delta - \varepsilon) \cdot \text{OPT}$ . We now use the matching  $M$  chosen in the first step of the algorithm (which is a maximum matching of  $G$ ) to augment the matching  $M'$ . We should point out that at this point,  $M'$  refers to a realized matching, while  $M$  is still a random variable (independent of  $M'$  since  $M$  and  $E_{MC}$  are edge-disjoint).

Let  $\alpha_1, \alpha_3$  and  $\alpha_{\geq 5}$  denote, respectively, the number of augmenting paths (w.r.t.  $M'$ ) of length 1, 3, and at least 5 in  $M \triangle M'$ . We have the following claim. The proof uses standard facts about the augmenting paths (see, e.g., [24]).

**Claim 6.2.** *For  $\alpha_1, \alpha_3$ , and  $\alpha_{\geq 5}$ , defined as above:*

$$\begin{aligned}\alpha_3 + 2\alpha_{\geq 5} &\leq |M'| \\ \alpha_1 + \alpha_3 + \alpha_{\geq 5} &= |M| - |M'|\end{aligned}$$

*Proof.* Any augmenting path of length 3 has one edge in  $M'$  and any augmenting path of length at least 5 has at least two edges in  $M'$ . Since the augmenting paths are edge disjoints, the first inequality follows. The second inequality follows from the fact that  $M$  is a maximum matching in  $G$  and each augmenting path in  $M \triangle M'$  increases the size of  $M'$  by 1. ■

As stated earlier, each edge in  $M$  is realized w.p.  $p$  (independent of the choice of  $M'$ ). Since an augmenting path of length 1 (resp. of length 3) realizes in  $M' \triangle M_p$  w.p.  $p$  (resp.  $p^2$ ), we have that the expected number of times that  $M'$  can be augmented using realized edges of  $M$  is at least  $\alpha_1 p + \alpha_3 p^2$ , implying that the final matching size is at least  $(\frac{1}{2} - \delta - \varepsilon) \cdot \text{OPT} + \alpha_1 p + \alpha_3 p^2$  in expectation. Combining this with Claim 6.2, the minimum size of the output matching we obtain can be formulated as the following linear program (denoted by LP-(6)):

$$\begin{array}{ll}\text{minimize} & \alpha_1 p + \alpha_3 p^2 \\ \text{subject to} & \alpha_3 + 2\alpha_{\geq 5} \leq (\frac{1}{2} - \delta) \text{OPT} - \varepsilon \cdot \text{OPT} \\ & \alpha_1 + \alpha_3 + \alpha_{\geq 5} \geq (\frac{1}{2} + \delta) \text{OPT} + \varepsilon \cdot \text{OPT} \\ & \alpha_1, \alpha_3, \alpha_{\geq 5} \geq 0\end{array} \tag{6}$$

where in the second constraint, we use the fact that  $M$  is a maximum matching in  $G$  and hence  $|M| \geq \text{OPT}$ . We have the following claim.

---

<sup>7</sup>This assumption can be removed while losing a negligible factor of  $o(1)$  in the size of final matching.

**Claim 6.3.** *The minimum value of LP-(6) is at least  $\frac{p^2}{2} \cdot \text{OPT}$ .*

*Proof.* The two constraints of LP-(6) imply that,

$$2\alpha_1 + \alpha_3 \geq \left(\frac{1}{2} + 3\delta + 3\varepsilon\right) \cdot \text{OPT} \quad (7)$$

Suppose we want to minimize  $\alpha_1 p + \alpha_3 p^2$  subject to the constraint in Eq (7) (this is clearly a lower bound for the value of LP-(6)). In this case, since the contribution of  $\alpha_3$  to the objective value is  $p$  times the contribution of  $\alpha_1$ , while its contribution to the constraint is  $\frac{1}{2}$  times the contribution of  $\alpha_1$ , it is straightforward to verify that for  $p \leq 1/2$ , there is an optimal solution with  $\alpha_1 = 0$ , and for  $p > 1/2$ , there is an optimal solution with  $\alpha_3 = 0$ . We can now compute the value of solution in each case:

**$p \leq \frac{1}{2}$  case.** In this case  $\alpha_1 = 0$  and  $\alpha_3 = (\frac{1}{2} + 3\delta + 3\varepsilon) \cdot \text{OPT}$  minimizes  $\alpha_1 p + \alpha_3 p^2$ . Hence, the objective value is

$$\alpha_3 \cdot p^2 = \left(\frac{1}{2} + 3\delta + 3\varepsilon\right) \cdot \text{OPT} \cdot p^2 \geq \frac{p^2}{2} \cdot \text{OPT}$$

**$p > \frac{1}{2}$  case.** In this case  $\alpha_1 = (\frac{1}{4} + \frac{3}{2}\delta + \frac{3}{2}\varepsilon) \cdot \text{OPT}$  and  $\alpha_3 = 0$  minimizes  $\alpha_1 p + \alpha_3 p^2$ . Hence, the objective value is

$$\begin{aligned} \alpha_1 \cdot p &= \left(\frac{1}{4} + \frac{3}{2}\delta + \frac{3}{2}\varepsilon\right) p \cdot \text{OPT} \geq \left(\frac{1}{4} + \frac{3}{2}\delta\right) p \cdot \text{OPT} \\ &= \left(\frac{1}{4} + \frac{3p^2}{8}\right) p \cdot \text{OPT} \geq \frac{p^2}{2} \cdot \text{OPT} \quad (\delta = \frac{p^2}{4} \text{ and } \frac{1}{4} + \frac{3p^2}{8} \geq 2\sqrt{\frac{1}{4} \cdot \frac{3p^2}{8}} \geq \frac{p}{2}) \end{aligned}$$

The claim now follows since in above calculation we *relaxed* constraints of LP-(6) to the constraint in Eq (7). ■

By plugging in the bound from Claim 6.3, we obtain that the final matching size is at least:

$$\begin{aligned} \frac{\text{OPT}}{2} - \delta \cdot \text{OPT} - \varepsilon \cdot \text{OPT} + \alpha_1 p + \alpha_3 p^2 &\geq \left(\frac{1}{2} - \delta - \varepsilon + \frac{p^2}{2}\right) \cdot \text{OPT} \\ &= \left(\frac{1}{2} + \frac{p^2}{4} - \frac{p_0^2}{10^4}\right) \cdot \text{OPT} \end{aligned}$$

by plugging in  $\delta = \frac{p^2}{4}$  and  $\varepsilon = \frac{p_0^2}{10^4}$ . ■

## 7 Concluding Remarks and Open Problems

We presented the first non-adaptive algorithm for stochastic matching with an approximation ratio that is strictly better than half. In particular, we showed that any graph  $G$  has a subgraph  $H$  with maximum degree  $O(\frac{\log(1/p)}{p})$  such that the ratio of expected size of a maximum matching in realizations of  $H$  and  $G$  is at least 0.52 when  $p$  is sufficiently small, i.e., case of vanishing probabilities, and  $0.5 + \delta_0$  (for an absolute constant  $\delta_0 > 0$ ) for any  $p \in (0, 1)$ .

A main open problem is to determine the best approximation ratio achievable by a non-adaptive algorithm. In particular, can non-adaptive algorithms qualitatively match the performance of adaptive algorithms by achieving a  $(1 - \varepsilon)$ -approximation for any  $\varepsilon > 0$  using a subgraph with maximum degree  $f(\varepsilon, p)$  for some function  $f$ ? In the following, we mention some potential directions towards resolving this problem.

**A barrier to obtaining a  $(1 - \varepsilon)$ -approximation.** We briefly explain here a barrier to a  $(1 - \varepsilon)$ -approximation algorithm that was noted in [8]. It was shown in [8] that any (non-adaptive)  $(1 - \varepsilon)$ -approximation algorithm for stochastic matching needs to solve the following problem.

**Problem ([8]).** Suppose you are given a bipartite graph  $G(L, R, E)$  ( $|L| = |R| = n$ ) with the property that the expected maximum matching size between two uniformly at random chosen subsets  $A \subseteq L$  and  $B \subseteq R$  with  $|A| = |B| = n/3$ , is  $n/3 - o(n)$ . The goal is to compute a subgraph  $H(L, R, Q)$  with max-degree of  $O(1)$ , such that the expected size of a maximum matching between two randomly chosen subsets  $A$  and  $B$  is  $\Omega(n)$ .

For the harder problem in which the two subsets  $A$  and  $B$  are chosen *adversarially*, it is known that there exist graphs (in particular, a Ruzsa-Szemerédi graph; see, e.g. [3, 22]) that admit no such sparse subgraph  $H$  (see [8] for more details). However, in the stochastic matching application, our interest is in *randomly* chosen subsets  $A$  and  $B$ , and it is not known if there are instances such that the random set version of the problem is hard.

**A direct application of  $b$ -matching lemma.** There is another possible way of utilizing the  $b$ -matching lemma. In Lemma 4.4, we showed that for any  $\frac{1}{p}$ -matching  $B$  of size  $\frac{\text{OPT}}{p}$ , the expected maximum matching size of a realization of  $B$  is at least  $\frac{\text{OPT}}{3}$ . In fact, using a more careful analysis, we can improve this bound to  $\approx 0.4 \cdot \text{OPT}$ . This, together with our  $b$ -matching lemma, immediately implies a simple 0.4-approximation algorithm for stochastic matching. However, it is not clear to us whether this bound can be significantly improved to get a matching of size strictly more than  $\frac{\text{OPT}}{2}$ . It is worth mentioning that using a result of Karp and Sipser [25] on *sparse random graphs* (see also [7], Theorem 4), one can show that if the  $\frac{1}{p}$ -matching itself is chosen *randomly*, then its realizations contain a matching of size  $\approx 0.56 \cdot \text{OPT}$  in expectation. However, this result relies heavily on the fact that the original graph (in our case a realization of a random  $\frac{1}{p}$ -matching) is chosen randomly, and it seems unlikely that a similar result holds for an *adversarially* chosen  $\frac{1}{p}$ -matching.

## References

- [1] M. Adamczyk. Improved analysis of the greedy algorithm for stochastic matching. *Inf. Process. Lett.*, 111(15):731–737, 2011.
- [2] M. Akbarpour, S. Li, and S. O. Gharan. Dynamic matching market design. In *ACM Conference on Economics and Computation, EC ’14, Stanford , CA, USA, June 8-12, 2014*, page 355, 2014.
- [3] N. Alon, A. Moitra, and B. Sudakov. Nearly complete graphs decomposable into large induced matchings and their applications. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 1079–1090, 2012.
- [4] R. Anderson, I. Ashlagi, D. Gamarnik, and Y. Kanoria. A dynamic model of barter exchange. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 1925–1933, 2015.
- [5] R. Anderson, I. Ashlagi, D. Gamarnik, and A. E. Roth. Finding long chains in kidney exchange using the traveling salesman problem. *Proceedings of the National Academy of Sciences*, 112(3):663–668, 2015.

- [6] J. Aronson, M. E. Dyer, A. M. Frieze, and S. Suen. Randomized greedy matching II. *Random Struct. Algorithms*, 6(1):55–74, 1995.
- [7] J. Aronson, A. M. Frieze, and B. Pittel. Maximum matchings in sparse random graphs: Karp-sipser revisited. *Random Struct. Algorithms*, 12(2):111–177, 1998.
- [8] S. Assadi, S. Khanna, and Y. Li. The stochastic matching problem with (very) few queries. *Proceedings of the 2016 ACM Conference on Economics and Computation, EC ’16, Maastricht, The Netherlands, July 24-28, 2016*, pages 43–60, 2016.
- [9] P. Awasthi and T. Sandholm. Online stochastic optimization in the large: Application to kidney exchange. In *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, 2009*, pages 405–411, 2009.
- [10] N. Bansal, A. Gupta, J. Li, J. Mestre, V. Nagarajan, and A. Rudra. When LP is the cure for your matching woes: Improved bounds for stochastic matchings. *Algorithmica*, 63(4):733–762, 2012.
- [11] A. Blum, J. P. Dickerson, N. Haghtalab, A. D. Procaccia, T. Sandholm, and A. Sharma. Ignorance is almost bliss: Near-optimal stochastic matching with few queries. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation, EC ’15, Portland, OR, USA, June 15-19, 2015*, pages 325–342, 2015.
- [12] A. Blum, A. Gupta, A. D. Procaccia, and A. Sharma. Harnessing the power of two crossmatches. In *ACM Conference on Electronic Commerce, EC ’13, Philadelphia, PA, USA, June 16-20, 2013*, pages 123–140, 2013.
- [13] B. Bollobás. *Random Graphs*. Number 73. Cambridge University Press, 2001.
- [14] T. H. Chan, F. Chen, X. Wu, and Z. Zhao. Ranking on arbitrary graphs: Rematch via continuous LP with monotone and boundary condition constraints. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1112–1122, 2014.
- [15] N. Chen, N. Immorlica, A. R. Karlin, M. Mahdian, and A. Rudra. Approximating matches made in heaven. In *Automata, Languages and Programming, 36th International Colloquium, ICALP 2009, Rhodes, Greece, July 5-12, 2009, Proceedings, Part I*, pages 266–278, 2009.
- [16] K. P. Costello, P. Tetali, and P. Tripathi. Stochastic matching with commitment. In *Automata, Languages, and Programming - 39th International Colloquium, ICALP 2012, Proceedings, Part I*, pages 822–833, 2012.
- [17] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Dynamic matching via weighted myopia with application to kidney exchange. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.*, 2012.
- [18] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Failure-aware kidney exchange. In *ACM Conference on Electronic Commerce, EC ’13.*, pages 323–340, 2013.
- [19] J. P. Dickerson and T. Sandholm. Futurematch: Combining human value judgments and machine learning to match in dynamic environments. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 622–628, 2015.

- [20] M. E. Dyer and A. M. Frieze. Randomized greedy matching. *Random Struct. Algorithms*, 2(1):29–46, 1991.
- [21] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang. On graph problems in a semi-streaming model. *Theor. Comput. Sci.*, 348(2-3):207–216, 2005.
- [22] E. Fischer, E. Lehman, I. Newman, S. Raskhodnikova, R. Rubinfeld, and A. Samorodnitsky. Monotonicity testing over general poset domains. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing.*, pages 474–483, 2002.
- [23] A. Gupta and V. Nagarajan. A stochastic probing problem with applications. In *Integer Programming and Combinatorial Optimization - 16th International Conference, IPCO 2013. Proceedings*, pages 205–216, 2013.
- [24] J. E. Hopcroft and R. M. Karp. An  $n^{5/2}$  algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.*, 2(4):225–231, 1973.
- [25] R. M. Karp and M. Sipser. Maximum matchings in sparse random graphs. In *22nd Annual Symposium on Foundations of Computer Science, Nashville, Tennessee, USA, 28-30 October 1981*, pages 364–375, 1981.
- [26] R. M. Karp, U. V. Vazirani, and V. V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing, May 13-17, 1990, Baltimore, Maryland, USA*, pages 352–358, 1990.
- [27] C. Konrad, F. Magniez, and C. Mathieu. Maximum matching in semi-streaming with few passes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 15th International Workshop, APPROX 2012, and 16th International Workshop, RANDOM 2012, Cambridge, MA, USA, August 15-17, 2012. Proceedings*, pages 231–242, 2012.
- [28] L. Lovász and D. Plummer. *Matching Theory*. AMS Chelsea Publishing Series. American Mathematical Soc., 2009.
- [29] D. F. Manlove and G. O’Malley. Paired and altruistic kidney donation in the UK: algorithms and experimentation. *ACM Journal of Experimental Algorithms*, 19(1), 2014.
- [30] A. Mehta and D. Panigrahi. Online matching with stochastic rewards. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 728–737, 2012.
- [31] A. Mehta, B. Waggoner, and M. Zadimoghaddam. Online stochastic matching with unequal probabilities. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 1388–1404, 2015.
- [32] M. Poloczek and M. Szegedy. Randomized greedy algorithms for the maximum matching problem with new analysis. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 708–717, 2012.
- [33] A. Schrijver. *Combinatorial optimization: polyhedra and efficiency*. Springer Science & Business Media, 2003.
- [34] U. Unver. Dynamic kidney exchange. *Review of Economic Studies*, 77(1):372–414, 2010.

- [35] V. G. Vizing. On an estimate of the chromatic class of a p-graph. *Diskret. Analiz*, 3(7):25–30, 1964.

## A Omitted Proofs from Section 3

### A.1 Proof of Proposition 3.1

*Proof.* We first have  $f(x)$  is *monotonically decreasing*, since,

$$\frac{df}{dx} = \frac{e^{-x} \cdot x - 1 + e^{-x}}{x^2} = \frac{(x+1) \cdot e^{-x} - 1}{x^2} \leq \frac{e^x \cdot e^{-x} - 1}{x^2} = 0$$

where we used the inequality  $(1+x) \leq e^x$ .

Consequently, since  $x \leq c$ ,

$$f(c) \leq f(x) = \frac{1 - e^{-x}}{x}$$

which implies  $e^{-x} \leq 1 - f(c) \cdot x$ . ■

### A.2 Proof of Proposition 3.2

*Proof.* We first exam the equivalent conditions for the target inequality.

$$\begin{aligned} (1-x)^{\frac{1}{x}} &\geq \frac{1-x}{e} \\ \iff (1-x)^{\frac{1}{x}-1} &\geq \frac{1}{e} \\ \iff \left(\frac{1}{x}-1\right) \ln(1-x) &\geq -1 \quad (\text{by taking natural log of both sides}) \end{aligned}$$

Now, since  $\ln(1-x) \geq -x - \frac{x^2}{2} - \frac{x^3}{2}$  when  $x \in (0, 0.43]$ . We have

$$\begin{aligned} \left(\frac{1}{x}-1\right) \ln(1-x) &\geq \left(\frac{1}{x}-1\right)\left(-x - \frac{x^2}{2} - \frac{x^3}{2}\right) \quad (\text{since } (\frac{1}{x}-1) > 0) \\ &= -1 - \frac{x}{2} - \frac{x^2}{2} + x + \frac{x^2}{2} + \frac{x^3}{2} \\ &= -1 + \frac{x}{2} + \frac{x^3}{2} \\ &\geq -1 \end{aligned}$$

which completes the proof. ■

## B The Optimality of the $b$ -Matching Lemma

In this section, we establish that our  $b$ -matching lemma is essentially optimal in the sense that it is impossible to find a  $b$ -matching with at least  $b \cdot \text{OPT}(G)$  edge for  $b$  much larger than  $1/p$ . In particular, we show that,

**Claim.** *For any constant  $0 < p < 1$ , there exist bipartite graphs  $G$  where  $G_p$  has a matching of size  $n - o(n)$  in expectation, but for any  $b \geq \frac{2}{p}$ , there is no  $b$ -matching in  $G$  with (at least)  $b \cdot 0.99n$  edges; here  $n$  is the number of vertices on each side of  $G$ .*

*Proof.* For any integer  $N$ , let  $\mathcal{G}_{N,\frac{1}{N}}$  be the family of bipartite random graphs with  $N$  vertices on each side and probability of picking each edge being  $1/N$ . Let  $c^* \in (0, 1)$  such that any bipartite graph sampled from  $\mathcal{G}_{N,\frac{1}{N}}$  has a matching of size at least  $c^* \cdot N$  w.p.  $1 - o(1)$ . By a result of Karp and Sipser [25] on *sparse random graphs* (see also [7], Theorem 4), we have  $c^* \approx 0.56$ .

Consider bipartite graphs  $G(L, R, E)$  where the vertices in  $L$  consists of two disjoint sets  $L_1$  and  $L_2$  with  $|L_1| = N$  and  $|L_2| = (1 - c^*) \cdot N$  for parameter  $N = \frac{n}{2 - c^*}$ . Similarly,  $R$  contains two sets  $R_1$  and  $R_2$  with  $|R_1| = N$  and  $|R_2| = (1 - c^*) \cdot N$ .

The set of edges in  $G$  can be partitioned into two parts. First, there is a complete bipartite graph between  $L_1$  and  $R_2$ , and a complete bipartite graph between  $L_2$  and  $R_1$ . Second, there is a sparse graph between  $L_1$  and  $R_1$  defined through the following random process: each edge between  $L_1$  and  $R_1$  is independently chosen w.p.  $\frac{1}{pN}$ .

In the following, we show that for a graph  $G$  created through the above process, w.p.  $1 - o(1)$ ,  $G_p$  has a matching of size  $n - o(n)$  in expectation, and w.p.  $1 - o(1)$ , there is no  $b$ -matching in  $G$  with  $b \cdot 0.99n$  edges, for  $b \geq \frac{2}{p}$ . Hence, by applying a union bound, the above process find a graph with both properties w.p.  $1 - o(1)$ , proving the claim.

To see that  $G_p$  has a matching of size  $n - o(n)$  in expectation, we realize the edges in  $G$  in two steps: first realize the edges between  $L_1$  and  $R_1$ , and then the other edges (i.e., the two complete graphs between  $L_1, R_2$  and between  $L_2, R_1$ , respectively). For the subgraph between  $L_1$  and  $R_1$ , notice that each edge between  $L_1$  and  $R_1$  is realized w.p.  $\frac{1}{pN} \cdot p = \frac{1}{N}$  (chosen w.p.  $\frac{1}{pN}$  in the above process and realize w.p.  $p$ ). Since  $|L_1| = |R_1| = N$ , the subgraph between  $L_1$  and  $R_1$  is sampled from  $\mathcal{G}_{N,\frac{1}{N}}$  and hence w.p.  $1 - o(1)$ , there is a matching of size  $c^*N$  between  $L_1$  and  $R_1$ . Now for the remaining  $(1 - c^*)N$  unmatched vertices in  $L_1$  (resp. in  $R_1$ ), since there is a complete graph between  $L_1$  and  $R_2$  (resp.  $R_1$  and  $L_2$ ), w.p.  $1 - o(1)$ , a perfect matching realizes between the unmatched vertices in  $L_1$  and vertices in  $R_2$  (resp. between  $R_1$  and  $L_2$ ). We conclude that any realization  $G_p$  has a perfect matching w.p.  $1 - o(1)$  and hence the expected maximum matching size in  $G_p$  is at least  $(1 - o(1))n + o(1) \cdot 0 = n - o(n)$ .

It remains to show that w.p.  $1 - o(1)$ ,  $G$  has no  $b$ -matching with  $b \cdot 0.99n$  edges for  $b \geq \frac{2}{p}$ . For any  $b$ -matching in  $G$ , the number of edges incident on  $L_2$  and  $R_2$  is at most  $b \cdot (|L_2| + |R_2|) = 2bN/(1 - c^*)$ . The remaining edges of this  $b$ -matching must be between  $L_1$  and  $R_1$ . Each edge between  $L_1$  and  $R_2$  is chosen w.p.  $\frac{1}{pN}$ , and there are  $N^2$  possible edges between  $L_1$  and  $R_1$ . By Chernoff bound, w.p.  $1 - o(1)$ , the number of realized edges between  $L_1$  and  $R_1$  is at most  $(1 + o(1))\frac{N}{p}$ . Therefore, the total number of edges of any  $b$ -matching in  $G$  is at most

$$\begin{aligned} 2b(1 - c^*)N + (1 + o(1))\frac{N}{p} &= b \cdot n - \left(c^*b \cdot N - \frac{N}{p}\right) + o(n) && ((2 - c^*) \cdot N = n) \\ &\leq b \cdot n - (0.56bN - 0.5bN) + o(n) \\ &&& (b \geq 2/p \text{ and hence } 1/p \leq b/2; c^* \approx 0.56) \\ &< b \cdot 0.99n \end{aligned}$$

■