

UNBALANCED RANDOM MATCHING MARKETS WITH PARTIAL PREFERENCES

ADITYA POTUKUCHI AND SHIKHA SINGH

ABSTRACT. Properties of stable matchings in the popular random-matching-market model have been studied for over 50 years. In a random matching market, each agent has complete preferences drawn uniformly and independently at random. Wilson (1972), Knuth (1976) and Pittel (1989) proved that in balanced random matching markets, the proposers are matched to their $\ln n$ th choice on average. In this paper, we consider markets where agents have partial (truncated) preferences, that is, the proposers only rank their top d partners. Despite the long history of the problem, the following fundamental question remained unanswered: *what is the smallest value of d that results in a perfect stable matching with high probability?* In this paper, we answer this question exactly—we prove that a degree of $\ln^2 n$ is necessary and sufficient. That is, we show that if $d < (1 - \varepsilon) \ln^2 n$ then no stable matching is perfect and if $d > (1 + \varepsilon) \ln^2 n$, then every stable matching is perfect with high probability. This settles a recent conjecture by Kanoria, Min and Qian (2021).

We generalize this threshold for unbalanced markets: we consider a matching market with n agents on the shorter side and $n(\alpha + 1)$ agents on the longer side. We show that for markets with $\alpha = o(1)$, the sharp threshold characterizing the existence of perfect stable matching occurs when d is $\ln n \cdot \ln \left(\frac{1+\alpha}{\alpha + (1/n(\alpha+1))} \right)$.

Finally, we extend the line of work studying the effect of imbalance on the expected rank of the proposers (termed the “stark effect of competition”). We establish the regime in unbalanced markets that forces this stark effect to take shape in markets with partial preferences.

1. INTRODUCTION

The stable matching problem is a classic problem in computer science, economics and mathematics. In the standard and perhaps antiquated description of the problem as the “stable marriage problem”, the two sides of the market are men and women, each with complete preferences over each other. A matching is considered to be **stable** if no pair of agents would prefer to break off from their current match and match with each other instead. Stability has proved to be an extremely robust notion in the success of centralized two-sided matching markets Roth (2002). The first stable matching algorithm proposed by Gale and Shapley (1962) has since been used in many real-life matching markets such as matching doctors to hospitals, students to schools, and applicants to jobs; see for example Abdulkadiroğlu et al. (2005a,b); Correa et al. (2019); Roth and Peranson (1999).

Initiated by Knuth (1976); Wilson (1972a), there has over 50 years of fascinating research to understand the properties of stable matchings in random-matching markets, where the preferences of the agents are generated uniformly and independently at random.

1.1. Our Results. In this paper, we study a fundamental and natural question about random matching markets with partial preference lists:

If each agent ranks only their top d choices (uniformly and independently at random), how big does d have to be to ensure that all agents are matched?

We consider a two-sided market, with potentially unequal number of agents on each side, and say that a stable matching is **perfect** if it matches all agents on the short side.

As our main result, we tightly characterize the existence of perfect stable matchings.

Theorem. *(Informal) Given a stable matching market with $n + k$ candidates, n jobs and uniform-random preference lists of length d , there exists a **threshold** d_0 such that for any (constant) $\varepsilon > 0$,*

- *if $d > (1 + \varepsilon)d_0$, then all stable matchings are perfect with high probability, and*
- *if $d < (1 - \varepsilon)d_0$, then no stable matching is perfect with high probability.*

Moreover, we determine d_0 asymptotically:

- (1) *for **balanced markets** with $k = 0$, $d_0 = \ln^2 n$*
- (2) *for **unbalanced markets** with $k = o(n)$, $d_0 = \ln n \ln \frac{1+(k/n)}{(k/n)+1/(n+k)}$*

For balanced markets, the above result settles a recent conjecture by Kanoria et al. (2021) who prove a weaker bound—they show $d = o(\ln^2 n)$ is not sufficient and $d = \Omega(\ln^2 n)$ is sufficient for a perfect stable matching to exist. Even for complete preference lists, the best-known bound on the maximum number of proposals any agent has to make is $\Theta(\log^2 n)$ by Pittel (2019).

For unbalanced markets, Theorem 2 provides a fundamentally new understanding of perfect stable matchings. Prior to this work, the best-known bounds are from Kanoria et al. (2021) who show that the threshold $d_0 = \Theta(\ln^2 n)$ continues to hold even in the presence of polynomially-small imbalance. In particular, for imbalance $k = \Omega(n^{1-\varepsilon})$, it was unknown whether (a) such a phenomenon of a sharp threshold continues to persist, and (b) how the existence of perfect stable matchings depends on the imbalance k and preference list d (even asymptotically).

Finally, we also analyze how the competition in the market affects the best-possible stable partner that the candidates can obtain.

Theorem. *(Informal) Given a stable matching market with $n + k$ candidates, n jobs and uniform-random preference lists of length d , the expected rank of the best-possible stable partner each candidate can obtain is at least $(1 - o(1))d / (\ln \frac{1+(k/n)}{(k/n)+1/(n+k)})$.*

Theorem 3 captures the intuitive property that the more the competition between the $n + k$ candidates for the n jobs, the worse their stable outcome (as a function of the imbalance k). Thus, it generalizes part of the results in the line of work studying the “stark effect of competition” in unbalanced matching markets Ashlagi et al. (2017); Cai and Thomas (2022); Kanoria et al. (2021).

1.2. Background on Stable Matchings. To provide some context for our results, we first review the classic **deferred-acceptance (DA) algorithm** by Gale and Shapley (1962); McVitie and Wilson (1970) to find a stable matching. Given the preference lists of each candidate and job, consider the following procedure in which the candidates “propose” to jobs on their list. At each point, a candidate without a potential match (chosen arbitrarily) proposes to the highest job on their list that they have not been rejected by yet. On receiving a proposal from a candidate c , a job j becomes potentially matched to c if c is the highest ranked among the candidates that have proposed to j so far, otherwise j rejects c . The process goes on until every candidate has either been rejected by every job on their list, or has a potential match. At the end, the potential matches are final.

The DA algorithm has some interesting properties that those unfamiliar may find surprising: (a) it always find a stable matching, and (b) this stable matching results in the best-possible stable partner for all candidates, and the worst-possible stable partner for the jobs. Beyond algorithmic application, these properties also make the DA algorithm an important tool in the study of random

instances. Notably, some of the early results by Knuth (1976); Pittel (1989); Wilson (1972a) on balanced matching markets with uniform-random and complete preference lists show that each candidate on average is matched to their $\ln n$ th ranked job and each job is matched to their $(n/\ln n)$ th ranked candidate.

The DA algorithm readily generalizes to partial preferences, which were first studied by Gusfield and Irving (1989). Notice that if the preference list length d is too small, some candidates/jobs may remain unmatched. Remarkably, the set of unmatched agents stays the same in every stable matching (this is referred to as the Lone Wolf Theorem McVitie and Wilson (1971)). Thus, if a perfect stable matching exists, then all stable matchings are perfect.

1.3. Model and Formal Statement of Results. Before we proceed, we formally define the model. Consider a two-sided matching instance $M_{n,d,\alpha}$ with $n(\alpha + 1)$ candidates and n jobs, where $\alpha \geq 0$. For balanced markets with $\alpha = 0$, we omit it from the instance and write $M_{n,d}$.

An edge (c, j) between a candidate c and job j occurs independently with probability d/n (thus the resulting graph is a bipartite Erdős-Rényi graph $G_{n(\alpha+1),n,d/n}$ graph). Each candidate and job have random ranking over their neighbors, chosen uniformly and independently. So, $M_{n,n}$ is just the random stable matching instance with complete preference lists.¹

We refer to d as the **degree** of the underlying random graph. One may easily deduce that for a perfect stable matching, d needs to be at least $(1 - o(1)) \ln n$, since otherwise, $M_{n,d}$ has no perfect matching with high probability. It turns out that the critical threshold for perfect stable matchings is a bit higher, and is $\ln^2 n$.

This is the formal statement of our first main theorem.

Theorem 1. *For each $\varepsilon > 0$,*

- (1) *if $d \geq (1 + \varepsilon) \log^2 n$, then w.h.p., all stable matchings in $M_{n,d}$ are perfect, and*
- (2) *if $d \leq (1 - \varepsilon) \log^2 n$, then w.h.p., no stable matching in $M_{n,d}$ is perfect.*

The DA algorithm also naturally motivates another distribution over random stable matching instances, where each candidate chooses a set of d (chosen in advance) jobs along with a uniform ranking independently, and each job ranks the candidates uniformly and independently. Let us use $M_{n,d}^C$ to describe this model. This model of *partial* preferences ($d < n$) was first considered by Immorlica (2005); Immorlica and Mahdian (2015), who assume that the proposing side has preference lists consisting of only $O(1)$ jobs (drawn independently at random using an arbitrary distribution). They show that the number of candidates with more than one stable partner makes up a vanishingly small fraction. The study of $M_{n,d}^C$ was revisited recently by Kanoria et al. (2021), who showed that if $d = o(\ln^2 n)$, then w.h.p. no stable matching is perfect and if $d = \Omega(\ln^2 n)$, then w.h.p. every stable matching is perfect. Empirically, they observe that the threshold is about $\ln^2 n$ for reasonable values of n .

It turns out that these different models are closely related to each other in the sense that some results (particularly, the type we are interested in) often translate between them when $d \gg \ln n$. Indeed, we prove that in this regime, a random instance from one of these models can be “sandwiched” using instances of another model that are close enough; see Lemma 7.

We now formally state our general theorem on threshold for perfect stable matching.

Theorem 2. *Consider a stable matching instance $M_{n,d,\alpha}$ where $0 \leq \alpha$ and $\alpha = o(1)$ and $1 \leq d \leq n$. Then, for each $\varepsilon > 0$,*

¹We abuse notation and use $M_{n,d}$ to refer both to the stable matching instance (graph and rankings) and the graph itself.

- (1) if $d > (1 + \varepsilon) \ln n \cdot \ln \left(\frac{1+\alpha}{\alpha + (1/n)(1+\alpha)} \right)$, then w.h.p, all stable matchings are perfect, and
- (2) if $d < (1 - \varepsilon) \ln n \cdot \ln \left(\frac{1+\alpha}{\alpha + (1/n)(1+\alpha)} \right)$, then w.h.p, no stable matching is perfect.

Competition in unbalanced matching markets. A particularly remarkable result by Ashlagi et al. (2017) reveals the striking effect of even a small imbalance on random stable matching markets. They consider random markets with $n + 1$ candidates and n jobs and complete preference lists—that is, a stable matching instance $M_{n,n,1/n}$. They show that on average each candidate is matched to their $\Omega(n/\ln n)$ th ranked job, a stark drop from their $O(\ln n)$ th ranked job in the balanced setting.² Likewise, each job gets its $O(\ln n)$ th ranked candidate, a stark rise from the $\Omega(n/\ln n)$ in the balanced case. Thus, the huge advantage of being on the proposing side (over the non-proposing side) is lost as soon as there is *any* competition. This intriguing discovery further motivates our study of stable matchings in a general $M_{n,d,\alpha}$ setting.

The relationship between imbalance α and degree d was recently studied by Kanoria et al. (2021). The focus of their work was to study the effect of competition as the degree d becomes smaller. They study stable matching instances $M_{n,d,\alpha}^C$ with *polynomially small imbalance*, that is, $1/n \leq \alpha \leq 1/n^\varepsilon$, and show the following dichotomy: (a) when $d = o(\ln^2 n)$, there is no perfect matching and both candidates and jobs are matched to the $\sqrt{d}(1 + o(1))$ -ranked partner on average and thus there is no effect of competition, and (b) when $d = \Omega(\ln^2 n)$, then all stable matchings are perfect and there is a stark effect of competition—in a candidate-proposing DA, the candidates are matched to their $(d/\ln n)$ th ranked job and the jobs to their $\ln n$ th ranked candidate.

Our results complements their findings. Theorem 2 determines the exact threshold for the existence of perfect stable matchings, and, as a function of imbalance α that can be much larger than $1/n^\varepsilon$. Finally, Theorem 3 below improves the lower bound of $(d/\ln n)$ on the expected rank of the candidates in a candidate-proposing DA, and shows how it degrades as the competition grows.

Theorem 3. *Consider a stable matching instance $M_{n,d,\alpha}^C$ where $0 \leq \alpha$ and $\alpha = o(1)$ and $1 \leq d \leq n$. Then, the expected rank of the candidates in the candidate-proposing DA is lower bounded by*

$$(1 - o(1)) \frac{d}{\ln \left(\frac{1+\alpha}{\alpha + 1/n(1+\alpha)} \right)}.$$

A brief, but rather curious, technical remark: the methods of Kanoria et al. (2021) prove a lower bound of $\sqrt{d}(1 + o(1))$ (with imbalance α being a lower-order term). This bound is better than the one in Theorem 3 when $\alpha \ll e^{-\sqrt{d}}$. Interestingly enough, the regime that they work in, $\alpha = n^{-\varepsilon}$ and $d = o(\log^2 n)$, reflects this and the competition in this regime plays a negligible role. It would be a natural guess that $\alpha \approx e^{-\sqrt{d}}$ is where the “stark effect of competition” phenomenon starts to take shape. So, we conjecture that the bound in Theorem 3 is tight when $\alpha \gg e^{-\sqrt{d}}$.

1.4. Technical Contributions. We develop the following technical tools in our analysis.

First, we compare the proposals in the DA algorithm to a balls-in-bins process. This comparison itself is not new: analyzing the DA as a balls-and-bins or coupon-collector game dates back to the original papers by Knuth (1976); Wilson (1972a). We add to this approach by providing some general bounds using simple methods that could simplify similar proofs in the future.

Second, to establish whether or not a stable matching exists, we focus on which of the two competing forces in the DA algorithm occurs first: each of the n jobs receiving a proposal or some candidate exhausting all d of their proposals. We analyze this competition by defining a

²Part of this result was later simplified by Cai and Thomas (2022).

simple probabilistic game on the “rejection chains” in the DA algorithm. Kanoria et al. (2021) also analyzes rejection chains, but at a conceptual level, understanding the competing forces seems to offer a fundamentally different approach. For instance, the analysis in Kanoria et al. (2021) requires a bound on the total number of proposals made in the DA algorithm. The aforementioned game lets us circumvent this and isolate the event of interest (the existence of a perfect stable matching).

Finally, we study three closely related random models: the symmetric case $M_{n,d,\alpha}$ which is arguably the most natural to state as it is independent of which side of the market proposes, and the asymmetric models $M_{n,d,\alpha}^C$ and $M_{n,d,\alpha}^J$ in which the candidates and jobs choose preference lists of length d respectively. Switching between these models helps analyze the problem from both sides: in particular, we analyze the candidate-proposing DA on $M_{n,d,\alpha}^C$ and the job-proposing DA on $M_{n,d,\alpha}^J$. We combine these results to obtain a bound for $M_{n,d,\alpha}$ using a “sandwiching lemma” (Lemma 7). We believe that this approach of analyzing the DA on different random stable-matching instances provides new insights as past work has only studied the instance $M_{n,d,\alpha}^C$.

1.5. Additional Related Work. The original stable matching algorithm by Gale and Shapley (1962) was extended to the imbalanced case by McVitie and Wilson (1970), and to the case of partial preferences by Gusfield and Irving (1989). Stable matchings have since been extensively studied over 60 years with several books on the topic; we refer to the readers to Manlove (2013).

Here we present closely related results on random-matching markets. Knuth (1976); Wilson (1972a) reduce the problem of running deferred acceptance in random balanced matching markets with complete lists, $M_{n,n}$, to the coupon-collector problem and use it to compute its average complexity. This technique is used in various works (including this paper), e.g. Ashlagi et al. (2021, 2017); Kanoria et al. (2021). Wilson (1972b) showed that the expected number of proposals in the candidate-proposing DA is at most nH_n . Knuth (1976) improved this upper bound to $(n-1)H_n + 1$ and proved a matching lower bound of $nH_n - O(\ln^4 n)$. Thus, the candidates side are matched to their $\ln n$ th ranked partner in expectation. In contrast, Pittel (1989) showed that the receiving side is a lot worse and is matched to their $(n/\ln n)$ th partner in expectation.

The algorithm to enumerate all stable partners of an agent (by breaking matches and continually running deferred acceptance) was suggested by McVittie McVitie and Wilson (1971) and Gusfield Gusfield (1987). In this paper, we use the simplified version by Knuth et al. (1990). This technique has also often used to show that each agent has a unique stable partner Ashlagi et al. (2017); Immorlica (2005), the phenomenon referred to as “core convergence”.

Non-uniform preference distributions have been studied, such as tiered preferences Ashlagi et al. (2021) or correlated preferences Gimbert et al. (2021). Partial preference lists based on public scores was recently studied by Agarwal and Cole (2023).

For a recent survey on the history and results on random matching markets, see Murras (2021).

1.6. Organization. We describe the model, the deferred acceptance algorithm, as well the main building blocks of our analyses—the probability game—in Section 2. We give the tight upper and lower bounds on the threshold in Section 3. Finally, we analyze the expected rank in Section 4.

2. PRELIMINARIES

In this section, we formally define our model, the stable matching algorithm used and other definition and notation used in the rest of the paper.

Stable matchings. Consider a stable matching instance M with $n(\alpha+1)$ candidates \mathcal{C} and n jobs \mathcal{J} . Without loss of generality assume that $\alpha \geq 0$. Each candidate $c \in \mathcal{C}$ has a (partial) preference

list of length d_c^M over a subset of jobs ranking them from the most to least preferred. Similarly, each job $j \in \mathcal{J}$ has a (partial) preference list of length d_j^M over a subset of candidates ranking them from most to least preferred. Consider the undirected bipartite graph G^M on $\mathcal{C} \cup \mathcal{J}$ where the edge (c, j) is present if c appears on j 's preference list and vice versa. Let $L_c^M = (j_1, \dots, j_{d_c^M})$ to denote the list of the candidate c in M . We define L_j^M for jobs analogously.

For stable matching instances M_1 and M_2 on $\mathcal{C} \cup \mathcal{J}$, we say that $M_1 \subseteq_{\mathcal{C}} M_2$ if for each candidate $c \in \mathcal{C}$, we have that $d_c^{M_1} \leq d_c^{M_2}$ and the first $d_c^{M_1}$ elements and their order in $L_c^{M_2}$ is the same as in $L_c^{M_1}$. Moreover, for every job j , the relative ranking of the candidates in $L_j^{M_1}$ remains the same. Similarly, we say that $M_1 \subseteq_{\mathcal{J}} M_2$ if the same holds for every $j \in \mathcal{J}$.

A **matching** μ is a set of vertex-disjoint edges in G^M . A matching μ is **perfect** if each job j is an endpoint of some edge in μ . For simplicity, we let $\mu(c)$ denote a candidate c 's match in μ (let $\mu(c) = \emptyset$ if c is unmatched). We define $\mu(j)$ similarly for a job j . A matching μ is **stable** if there exists no **blocking pair** with respect to μ . A pair (c, j) where $c \in \mathcal{C}$ and $j \in \mathcal{J}$ is a **blocking pair** with respect to the matching μ if any one of the following conditions hold: (a) c prefers j to $\mu(c)$ and j prefers c to $\mu(j)$, (b) c prefers \emptyset to $\mu(c)$, and (c) j prefers \emptyset to $\mu(j)$.

Deferred-acceptance algorithm. We describe the **candidate-proposing deferred-acceptance** (CPDA) algorithm in Algorithm 1. The job-proposing DA (JPDA) is analogous.

Let \mathcal{U} be the set of unmatched candidates. Let \mathcal{R} be the set of candidates who have **exhausted** their preference list, that is, have been rejected by all jobs on their preference list.

Algorithm 1: Candidate Proposing Deferred-Acceptance (CPDA) Algorithm

```

Fix an ordering over  $\mathcal{C}$ ; Initialize  $\mathcal{U} \leftarrow \mathcal{C}$  and  $\mathcal{R} \leftarrow \emptyset$  and  $\mu(j) \leftarrow \emptyset$  for all  $j \in \mathcal{J}$ 
while  $\mathcal{U} \setminus \mathcal{R} \neq \emptyset$  do
    Pick a candidate  $c \in \mathcal{U}$  with the lowest index
    if  $c \notin \mathcal{R}$  then
        Let  $j$  be the most preferred job on  $c$ 's list that  $c$  has not yet proposed to
        if  $j$  is the last job on  $c$ 's list then
            | add  $c$  to  $\mathcal{R}$ 
        end
         $c$  proposes to  $j$ ; Let  $c' = \mu(j)$ 
        if  $j$  prefers  $c$  to  $c'$  then
            |  $j$  rejects  $c'$ 
            | if  $c' \neq \emptyset$  then
            | | add  $c'$  to  $\mathcal{U}$ 
            | end
            |  $j$  accepts  $c$ ;
            | Set  $\mu(j) = c$  and remove  $c$  from  $\mathcal{U}$ 
        end
    end
end

```

We use the following properties of the CPDA algorithm.

Theorem 4. *Gale and Shapley (1962); McVitie and Wilson (1970)* The CPDA algorithm has the following properties:

- (1) It outputs a stable matching μ^* . The matching μ^* is **candidate optimal**, that is, each candidate is matched to their best-possible stable match. Moreover, μ^* is **job pessimal**, that is, each job is matched to their worst-possible stable match.

- (2) The matching μ^* is independent of the order of the proposals. Moreover, the same proposals are made during the algorithm, regardless of which candidate is chosen to propose at each step.
- (3) (**Lone Wolf Theorem**) The candidates and jobs that are unmatched in μ^* do not have any possible stable partner. That is, the set of unmatched agents is the same across all stable matchings.

Algorithm terminology. We define the current **time step** of the CPDA algorithm as the number of proposals made so far. Since the next proposal is always from a currently unmatched candidate with the lowest index, a candidate keeps proposing until either they are matched or exhaust their preference list. Moreover, if the t th proposal results in a rejection of a candidate c (either because c made the proposal, or because the t th proposal causes a previously-matched c to become unmatched) the next proposal is always from c . This sequence of proposals is termed as the **rejection chain** in the literature. We use rejection chains in our analysis and define it more formally below.

Definition 5. (*Rejection Chain*) Fix any time step t in CPDA (Algorithm 1) and let c be the next candidate in line to propose. The rejection chain $R(t)$ starting at time step t is empty if c has been rejected by all jobs on their preference list. Otherwise, let j the job c proposes to next.

- If j is currently unmatched, then $R(t) = (c, j, \text{accept})$.
- Otherwise, if j prefers $\mu(j)$ to c , then $R(t) = (c, j, \text{reject}) \circ R(t + 1)$.
- Otherwise, if j prefers c to $\mu(j)$, then $R(t) = ((c, j, \text{accept}), (c', j, \text{reject})) \circ R(t + 1)$.

We use the following two observations about rejection chains in our analysis.

Observation 6. Consider the rejection chain $R(t)$ at time step t defined in Definition 5.

- (1) $R(t)$ terminates when either previously unmatched job receives a proposal and thus will remain matched till the end, or a candidate is rejected by all jobs on their preference list and thus will remain unmatched till the end.
- (2) In a rejection chain, if d contiguous proposals result in a rejection, then some candidate must have been rejected d times and must remain unmatched.

Note that Part (2) of the above observation was also used by Kanoria et al. (2021).

Deferred Acceptance on Random Instances. We recall the random stable matching instance $M_{n,d,\alpha}$ defined in Section 1. In this instance, there are $m = n(\alpha + 1)$ candidates and n jobs. The underlying graph is a Erdős-Rényi graph $G_{n(\alpha+1),n,d/n}$. Thus, an edge (c, j) between a candidate c and job j occurs with independent probability d/n . Moreover, each candidate and job have a random ranking over their neighbors, chosen uniformly and independently. The stable matching instance $M_{n,d,\alpha}$ refers to the graph $G_{n(\alpha+1),n,d/n}$ and the preferences over the edges.

Let us define two other related models of random stable matching instances. Let us define $M_{n,\alpha,d}^C$ to denote a stable matching instance with n jobs, $n(1 + \alpha)$ candidates, and each candidates chooses a preference list of d jobs uniformly and independently. This is the model in Kanoria et al. (2021).

Similarly, we also define $M_{n,\alpha,d}^J$ where each job chooses a preference list of d candidates uniformly and independently.

We analyze the CPDA algorithm on $M_{n,\alpha,d}^C$ and the JPDA algorithm on $M_{n,\alpha,d}^J$. To obtain the final result on the symmetric stable matching instance $M_{n,d,\alpha}$, we relate all three random instances using the following “sandwiching” lemma.

Lemma 7. (*Sandwiching Lemma*) For every $\delta > 0$, there is a joint distribution $(M_{n,\alpha,d}^C, M_{n,\alpha,d(1+\delta)}^C)$ such that with probability at least $1 - O(n \exp\{-\delta^2 d/3\})$, we have

$$M_{n,\alpha,d(1-\delta)}^C \subseteq_C M_{n,\alpha,d} \subseteq_C M_{n,\alpha,d(1+\delta)}^C.$$

The same claim also holds for $(M_{n,\alpha,d(1-\delta)}^J, M_{n,\alpha,d(1+\delta)}^J)$.

Since we (eventually) analyze the DA algorithm on a random instance $M_{n,\alpha,d}^C$, we reveal the next job j in the list of candidate c when they are called up to propose, uniformly among all jobs that have not already rejected c . We also reveal c 's relative ranking among all the proposals received by j (including from c) when c proposes to j .

2.1. A Probability Game. Define a multi-round probabilistic game as follows. In each round, there are three events that can occur: G (Good), B (Bad), and N (Neutral). Given an input parameter d , for each round, conditioned on the events in previous rounds, define $\mathbb{P}(G) = p_G$, $\mathbb{P}(B) \geq p_B$ and $\mathbb{P}(N) = p_N$. The game is **won** if G occurs. The game is **lost** if B occurs δ times in a row.

Let $\mathbb{P}_{\text{win}}(p_G, p_B)$ denote the probability of winning this game. Then, we prove the following.

Lemma 8.

$$\mathbb{P}_{\text{win}}(p_G, p_B) \leq \frac{p_G}{p_B^d}$$

Proof. Suppose the game is played for i rounds, and let $S = (s_1, s_2, \dots, s_i)$ denote the random sequence of events that occur. Consider a new game that is the same as the original except now $\mathbb{P}(B) = p_B$, and let $S' = (s'_1, s'_2, \dots, s'_i)$ denote the random sequence of events that occur when the new game is played for i rounds. Let $\mathbb{P}_{\text{win}'}$ be the winning probability in the new game. Through a coupling argument, we first show that $\mathbb{P}_{\text{win}} \leq \mathbb{P}_{\text{win}'}$.

Given the sequence S' , we create a sequence S as follows. For $i \geq 0$, as long as the game does not end in a loss at i rounds, do:

- if $s'_i = G$, then $s_i = G$,
- if $s'_i = B$, then $s_i = B$,
- if $s'_i = N$, then $s_i = N$ with probability $\frac{1-p_G-p_B^i}{1-p_G-p_B}$ and $s_i = B$ otherwise. Here $p_B^i = \mathbb{P}(s_i = B \mid s_{i-1}, \dots, s_1) \geq p_B$.

Note that whenever S' ends in a loss (event B occurs d times in a row), then S also ends in a loss. Thus, $\mathbb{P}_{\text{win}} \leq \mathbb{P}_{\text{win}'}$.

Finally, we prove finish the proof of the lemma by using the following recurrence.

$$\begin{aligned} \mathbb{P}_{\text{win}'} &= p_G + p_N \cdot \mathbb{P}_{\text{win}'} + p_B \cdot p_N \cdot \mathbb{P}_{\text{win}'} + p_B^2 \cdot p_N \cdot \mathbb{P}_{\text{win}'} + \dots + p_B^{d-1} \cdot p_N \cdot \mathbb{P}_{\text{win}'} \\ &= p_G + p_N \cdot \mathbb{P}_{\text{win}'} (1 + p_B + p_B^2 + \dots + p_B^{d-1}) = p_G + p_N \cdot \mathbb{P}_{\text{win}'} \left(\frac{1 - p_B^d}{1 - p_B} \right) \\ \mathbb{P}_{\text{win}'} &= \frac{p_G(1 - p_B)}{1 - p_B - p_N + p_N \cdot p_B^d} = \frac{p_G(1 - p_B)}{p_G + p_N \cdot p_B^d} \\ &= \frac{p_G}{p_B^d} \cdot \frac{(1 - p_B)}{(p_G/p_B^d + p_N)} \leq \frac{p_G}{p_B^d} \end{aligned}$$

□

Notation. We say an event \mathcal{E} occurs **with high probability** if $\mathbb{P}(\mathcal{E}) = 1 - o(1)$.

3. PROOF OF THEOREM 2

In this section, we prove Theorem 2. That is, we prove tight upper and lower bounds on the degree d that determines whether or not a stable matching exists in a random matching market with imbalance α and average degree d . As alluded to earlier, we prove the lower bound on $M_{n,\alpha,d(1+o(1))}^C$, and upper bound on $M_{n,\alpha,d(1-o(1))}^J$.

Theorem 9. *For any $\gamma > 0$, if $d < (1 - \gamma) \log n \cdot \log \left(\frac{1+\alpha}{\alpha + (1/n)(1+\alpha)} \right)$, then*

$$\mathbb{P}(M_{n,\alpha,d}^C \text{ has no perfect stable matching}) \geq 1 - (\alpha + 1/n)^{-\Omega(\gamma)}.$$

Theorem 10. *For any $\gamma > 0$, if $d > (1 + \gamma) \log n \cdot \log \left(\frac{1+\alpha}{\alpha + (1/n)(1+\alpha)} \right)$, then*

$$\mathbb{P}(M_{n,\alpha,d}^J \text{ has a perfect stable matching}) \geq 1 - n^{-\Omega(\gamma)}$$

We prove Theorems 9 and 10 in Section 3.2 and Section 3.3 respectively. Using them, we finish the proof of Theorem 2 below.

Proof of Theorem 2. Let us denote $d_0 := \log n \cdot \log \left(\frac{1+\alpha}{\alpha + (1/n)(1+\alpha)} \right) = \omega(\log n)$. We have for $\alpha = o(1)$ that $d_0 = \omega(\log n)$.

Denote $M = M_{n,\alpha,d_0(1-\varepsilon)}$, and $M^C = M_{n,\alpha,d_0(1-\varepsilon/2)}^C$. For the proof of part 1, it suffices to prove the same statement for M^C . Indeed, since by Lemma 7, there is a joint distribution (M, M^C) such that $M \subseteq_C M^C$ with probability at least $1 - n \exp\{-\Omega(\varepsilon^2 d_0)\}$. By part 2 of Theorem 4, we have that if the CPDA algorithm on M^C does not give a perfect stable matching, then it does not give one even in M .

The proof of part 2 proceeds in a similar fashion, using $M_{n,\alpha,(1+\alpha)d_0(1+\varepsilon/2)}^J$. □

3.1. High-Level Overview. We first describe the high-level idea of the upper and lower bound proofs for the balanced case ($\alpha = 0$) to give some intuition.

Lower bound. To show that a perfect stable matching does not exist when the average degree $d < (1 - \gamma) \log^2 n$ in $M_{n,d}$, we analyze the proposals when the CPDA algorithm is run on the random instance $M_{n,d}^C$. Thus, each candidate has a preference list of length d . Notice that in the CPDA, as soon as all n jobs receive even a single proposal, the final matching must be perfect. This is because jobs only upgrade their potential matches as the algorithm proceeds. On the other hand, a candidate c remains unmatched if c has been rejected by all d jobs on their preference list. Thus, to show that a perfect stable matching does not exist, we show that some candidate exhausts all d of their proposals *before* all n jobs receive a proposal. We model this using the probability game from Section 2.1. Finally, we set the probabilities in the game using a careful coupling between CPDA and a balls-in-bins process.

This proof naturally extends to the unbalanced case: we simply analyze the probability that $\alpha n + 1$ candidates exhaust all d of their proposals before all n jobs receive a proposal.

Upper bound. To show that a perfect stable matching exists when the average degree $d > (1 + \gamma) \log^2 n$ in $M_{n,d}$, we analyze the proposals in the JPDA algorithm is run on the random instance $M_{n,d}^J$. Thus, each job has a preference list of length d . To prove that each job gets matched, we look at all the proposals made by the last job j in JPDA and compute the probability that j gets rejected from all d candidates on its list and show that it is small. For a single proposal

from j to a candidate c to be successful, not only must c accept it but the resulting rejection chain must not knock j off. We use the probability game to compute the probability of the latter.

3.2. Lower Bound: Proof of Theorem 9.

Proof of Theorem 9. Let \mathcal{P} denote the event that $M_{n,\alpha,d}^C$ has a perfect stable matching. Recall that in $M_{n,\alpha,d}^C$, each candidate has a preference list of length d . We consider the CPDA procedure in Algorithm 1 on $M_{n,\alpha,d}^C$. We define an **extended process** in the algorithm where there are $n(1+\alpha)$ real candidates and ∞ ‘fake’ candidates and n jobs. Recall that we assume that there is some predetermined order on the candidates and the unmatched candidate with the least index is next in line to propose.

Fix $\kappa = n \ln n - Cn$, for $C = O(\gamma \ln(1/(\alpha + 1/n)))$.

Suppose κ proposals have been made. We consider the rejection chain $R(\kappa)$ as defined in Definition 5. In particular, consider the proposal made by candidate \tilde{c} at step $\kappa + 1$. We have the following cases: (a) the proposal goes to an unmatched job and is accepted, (b) the proposal goes a matched job and is rejected, and (c) the proposal goes to a matched job, is accepted, which causes its previous match c' now to be free. By Observation 6, $R(\kappa)$ ends when either a proposal goes to an unmatched job, or when some candidate exhausts all their d proposals (and thus remains unmatched).

Let us define the following events and then bound their probabilities.

- Let \mathcal{M} be the event that CPDA finds a (potential) matching of size n before $\alpha n + 1$ candidates have exhausted all d of their proposals.³
- Let \mathcal{L} be the event that every job gets a proposal from a candidate in at most κ rounds.
- Let \mathcal{B} be the event that after κ rounds, the probability that the next proposal from a candidate is accepted is greater than $\frac{n}{n+\kappa}(1 + \frac{\gamma}{3})$.
- Let \mathcal{E} be the event that the number of unmatched jobs at κ rounds is at least e^{2C} .

Then,

$$(1) \quad \mathbb{P}(\mathcal{P}) \leq \mathbb{P}(\mathcal{M}) \leq \mathbb{P}(\mathcal{M} | \overline{\mathcal{B}\mathcal{E}\mathcal{L}}) + \mathbb{P}(\mathcal{B}) + \mathbb{P}(\mathcal{E}) + \mathbb{P}(\mathcal{L})$$

As there are $n(\alpha + 1)$ real candidates and n jobs, there is no perfect stable matching if more than αn candidates end up with all d of their proposals rejected before all n jobs receive even a single proposal. We have the following claim whose proof is postponed to Section 7.

Claim 11. *The following inequalities hold:*

- (1) $\mathbb{P}(\mathcal{B}) \leq O(1/\sqrt{n})$.
- (2) $\mathbb{P}(\mathcal{E}) \leq e^{-C}$.
- (3) $\mathbb{P}(\mathcal{L}) \leq 2e^{-C/2} + O(1/\sqrt{n})$

With Claim 11 in hand, we finish the proof by using the probability game defined in Section 2.1.

Let $p_B := 1 - \frac{n}{n+\kappa}(1 + \frac{\gamma}{2})$. Conditioned on $\overline{\mathcal{B}}$, the next proposal is rejected with probability at least p_B . Conditioned on $\overline{\mathcal{E}}$, let p_G be the probability of the next proposal goes to an unmatched job. By Observation 6 (1) the number of unmatched jobs stays the same during all proposals that occur within the rejection chain $R(\kappa)$. By Observation 6 (2), any time there are d rejects in a row in $R(\kappa)$, some candidate must have exhausted all d of their proposals.

³Note that a potential matching of size n is found as soon as n jobs have received a proposal from some candidate (not necessarily their final match).

Now, we use the probability game from Section 2.1 and set $p_G = \frac{e^{2C}}{n}$ and $p_B = p$. Using Lemma 8 and the fact that $\mathbb{P}_{\text{win}}(p_G, p_B)$ is the probability that every subsequent rejection chain ends in a new job getting matched, we get the following:

$$\begin{aligned} \mathbb{P}_{\text{win}}(p_G, p_B) &\leq p_G p_B^{-d} \\ &\leq e^{2C} n^{-1} \left(1 - \frac{n}{n + \kappa} (1 + \gamma/3)\right)^{-(1-\gamma) \ln n \ln \left(\frac{1+\alpha}{\alpha+1/n(1+\alpha)}\right)} \\ &\leq e^{2C} n^{-1} \left(\frac{\alpha + 1/n(1+\alpha)}{1+\alpha}\right)^{-1+\gamma/2}. \end{aligned}$$

In the last inequality, we used the fact that $1 - x \geq e^{-x-x^2}$, and that $\kappa = \Omega(n \ln n)$. So, with probability at least $1 - e^{2C} \left(\alpha + \frac{1}{n(1+\alpha)}\right)^{\gamma/4} (1+\alpha)^{1-\gamma/2}$ the next $n \cdot \left(\alpha + \frac{1}{n(1+\alpha)}\right)^{1-\gamma/4} > \alpha n$ rejection chains end in an additional unmatched candidate. Therefore,

$$(2) \quad \mathbb{P}(\mathcal{M}|\overline{\mathcal{BEL}}) \leq e^{2C} \left(\alpha + \frac{1}{n(1+\alpha)}\right)^{\gamma/4} (1+\alpha)^{1-\gamma/2}.$$

Finally, substituting $C = O(\gamma \ln(1/(\alpha + 1/n)))$, and plugging in the bounds from Claim 11 and Inequality (2) into the expression in Inequality (1), we get

$$\mathbb{P}(\mathcal{P}) \leq (\alpha + 1/n)^{-\Omega(\gamma)}.$$

□

3.3. Upper Bound: Proof of Theorem 10. In particular, we prove the following.

Proof of Theorem 10. For convenience, let $m = n(1 + \alpha)$. Suppose $d > (1 + \gamma) \log n \cdot \ln \left(\frac{1+\alpha}{\alpha+1/m}\right)$. Consider the JPDA algorithm (analogous to Algorithm 1). Fix a job j , and let us place it in the last place in our predetermined order for the JPDA algorithm—that is, j is the last job that has not yet proposed to any candidate on their list. Let κ be the number of proposals at the moment j starts proposing. Let (the random candidate) c be j 's next proposal and N_c be the random variable denoting the number of proposals c has received so far. We have that $\mathbf{E}[N_c|\kappa] = \kappa/m$. Observe that j 's proposal to c is **successful** if the following two events occur:

- (1) c accepts j 's proposal, and
- (2) the resulting rejection chain $R(\kappa)$ does not end with j getting unmatched.

The first event occurs with probability $1/(N_c + 1)$.

To bound the probability of the second event, we analyze the proposals in $R(\tau)$ using the probability game from Section 2.1. In the particular, consider the probability game with parameter $d = 1$. Let G be the event that j gets unmatched because another job j' proposes to c in $R(\tau)$. Then, $p_G \leq 1/(m(N_c + 2))$ because the probability that c is next on j' preference list is at most $1/m$ and c must prefer j' to the $N_c + 1$ proposals it already has received. Let B be the event that the rejection chain $R(\kappa)$ ends without j getting kicked off from c in $R(\kappa)$. This occurs if $R(\kappa)$ ends in a new candidate getting matched. Thus $p_B \geq \frac{\alpha+1/m}{1+\alpha}$.

Then, \mathbb{P}_{win} corresponds to the probability that j does not end up being matched to c at the end of $R(\kappa)$, and by Lemma 8 we have,

$$(3) \quad \mathbb{P}_{\text{win}} \leq p_G/p_B \leq \frac{1 + \alpha}{(\alpha + 1/m) \cdot m \cdot (N_c + 2)}$$

Thus, the probability that j 's proposal to c is successful is

$$\begin{aligned} &\geq \mathbb{E} \left[\frac{1}{N_c + 1} \left(1 - \frac{1 + \alpha}{(\alpha + 1/m) \cdot m \cdot (N_c + 2)} \right) \right] \\ &\geq \mathbb{E} \left[\frac{1}{N_c + 1} \left(1 - \frac{1}{N_c + 2} \right) \right] \\ &= \mathbb{E} \left[\frac{1}{N_c + 2} \right] \geq \frac{1}{\mathbb{E}[N_c + 2]} = \frac{m}{\mathbb{E}[\kappa] + 2m} \geq \frac{m}{\mathbb{E}[\tau] + 2m}. \end{aligned}$$

Here τ denotes the final number of proposals (at the end of JPDA) and κ is the number of proposals at the time j started proposing.

Finally, the probability that j remains unmatched at the end of JPDA is the probability that all d of j 's proposals fail. This occurs with probability

$$\begin{aligned} &\leq \left[1 - \frac{m}{\mathbb{E}[\tau] + 2m} \right]^d \leq \exp \left\{ -\frac{m \cdot d}{\mathbb{E}[\tau] + 2m} \right\} \\ (4) \quad &\leq \exp \left\{ -\frac{d}{2 + \ln \left(\frac{1 + \alpha}{\alpha + 1/m} \right)} \right\} \leq \frac{1}{n^{1 + \gamma/2}}. \end{aligned}$$

Here, the inequality in step 4 follows from the following claim that is postponed to Section 7.

Claim 12. $\mathbb{E}[\tau] \leq (1 + \gamma/2) \cdot m \cdot \log \left(\frac{1 + \alpha}{\alpha + 1/m} \right) + o(1)$

Theorem 4 part 3 gives us that this holds for any job j . Thus the expected number of unmatched jobs is at most $n^{-\gamma/2}$, which gives us the desired bound. □

4. EFFECT OF COMPETITION: PROOF OF THEOREM 3

In this section, we prove a lower bound on the expected ranks of the candidates and the number of proposals in CPDA in an unbalanced random matching market with imbalance α . In particular, we prove Theorem 3.

Proof of Theorem 3. Consider the JPDA algorithm on $M_{n,d,\alpha}^{\mathcal{J}}$. We use the following lemma from past works Cai and Thomas (2022); Immorlica (2005); Knuth (1976).

Lemma 13. (Cai and Thomas (2022)) *Fix a job j and for any preference list P_j of job j and index $i \in [n]$, let P_j^i denote the preference list P_j truncated at index i . Then, job j has a stable partner of rank better than i if and only if j is matched in CPDA even if j truncates their list after rank i .*

We fix a candidate c^* . Let X be the random variable that represents c^* 's highest-rank stable partner (among all stable matchings). Let Y be the highest rank offer c receives if in the JPDA algorithm c keeps rejecting all proposals (and we keep running JPDA). Lemma 13 gives us that $\mathbf{E}[X] = \mathbf{E}[Y]$. Let Z be the random variable representing the number of proposals received by c^* during this algorithm where c^* rejects all proposals.

Let us compare the JPDA process to a random balls-in-bins process. Assume that there are $m = n(1 + \alpha)$ bins (one for each candidate). Suppose balls are being thrown uniformly and independently in the bins. Fix a bin c^* and let \tilde{Z} be the number of balls in c^* at the time that exactly $n + 1$ bins are occupied. Let κ be the number of balls thrown at this point. We show that this process *stochastically dominates* JPDA as follows.

Lemma 14. *There is a coupling between the JPDA algorithm and the balls-in-bins process so that $Z \leq \tilde{Z}$, and $\tau \leq \kappa$.*

Proof. We abuse notation and use $[m]$ to denote both the set of bins and the set of candidates. Consider the following coupling between the processes of JPDA and balls-in-bins.

Suppose at a point in the JPDA algorithm, a job $j \in [n]$ is required to make a (random) proposal, and so far, \mathbf{C}_j is the set of candidates that j has already proposed to. We keep throwing balls-into-bins uniformly at random until we hit a bin $\mathbf{x} \notin \mathbf{C}_j$. We assign \mathbf{x} as the next proposal of j . Thus the distribution of the next proposal of j is uniform on $[m] \setminus \mathbf{C}_j$.

Let τ and κ denote the number of proposals made and number of balls thrown in the above algorithm respectively. Fix any $i \in [m]$, and let Z and \hat{Z} denote the number of proposals received by the man i and the number of balls in the bin i respectively. We finish the proof by noting that the set \mathcal{B}' of bins occupied is a superset of the set \mathcal{C}' of candidates who received proposals (the algorithm could end earlier than the balls-in-bins process, where we have unlimited balls). Also, for the above process, $\tau \leq \kappa$ and $Z \leq \hat{Z}$. \square

As an immediate consequence, using symmetry, we have

$$\mathbf{E}[Z] \leq \mathbf{E}[\tilde{Z}] = \mathbf{E}[\mathbf{E}[\tilde{Z}|\kappa]] = \mathbf{E}[\kappa/m].$$

Since m^* has d women on his list, the expected highest rank proposal he receives is

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y|Z]] = \mathbf{E}[d/(Z + 1)] \geq d/\mathbf{E}[Z + 1] \geq d \cdot m/(\mathbf{E}[\kappa] + m) =: r_m.$$

So the expected rank of the highest ranked stable partner of every man is at least r_m . Therefore, the expected number of proposals in the MPDA algorithm is at least $m \cdot r_m = d \cdot m^2/\mathbf{E}[\kappa]$. The following proposition finishes the proof.

Proposition 15. $\mathbf{E}[\kappa] \leq (1 + o(1))m \ln(\frac{1+\alpha}{\alpha+1/m})$.

The proof is identical to Claim 12 and so we omit it. \square

5. CONCLUSION

In this paper, we study a fundamental question about the existence of perfect stable matching in random matching markets with imbalance and partial preferences. We prove sharp upper and lower bounds on the length of the preference list d that results in a perfect stable matching.

Partial preferences and imbalance are common occurrences in real matching markets. We believe that the new insights from this paper about the interplay between them will serve as guiding principles in market design.

6. ACKNOWLEDGEMENTS

This paper was supported in part by NSF CCF 1947789.

We would like to thank Jackson Ehrenworth and Max Enis for their contributions to a related project which led us to work on this paper.

REFERENCES

- Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. 2005a. The new york city high school match. *American Economic Review* 95, 2 (2005), 364–367.
- Atila Abdulkadiroğlu, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez. 2005b. The Boston public school match. *American Economic Review* 95, 2 (2005), 368–371.
- Ishan Agarwal and Richard Cole. 2023. Stable Matching: Choosing Which Proposals to Make. arXiv:2204.04162 [cs.GT]
- Itai Ashlagi, Mark Braverman, Amin Saberi, Clayton Thomas, and Geng Zhao. 2021. Tiered Random Matching Markets: Rank Is Proportional to Popularity. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 185)*, James R. Lee (Ed.). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 46:1–46:16. <https://doi.org/10.4230/LIPIcs.ITCS.2021.46>
- Itai Ashlagi, Yash Kanoria, and Jacob D Leshno. 2017. Unbalanced random matching markets: The stark effect of competition. *Journal of Political Economy* 125, 1 (2017), 69–98.
- Linda Cai and Clayton Thomas. 2022. The short-side advantage in random matching markets. In *Symposium on Simplicity in Algorithms (SOSA)*. SIAM, 257–267.
- Jose Correa, Rafael Epstein, Juan Escobar, Ignacio Rios, Bastian Bahamondes, Carlos Bonet, Natalie Epstein, Nicolas Aramayo, Martin Castillo, Andres Cristi, et al. 2019. School choice in Chile. In *Proceedings of the 2019 ACM Conference on Economics and Computation*. 325–343.
- Devdatt P. Dubhashi and Alessandro Panconesi. 2009. Concentration of Measure for the Analysis of Randomized Algorithms. <https://api.semanticscholar.org/CorpusID:8135469>
- David Gale and Lloyd S Shapley. 1962. College admissions and the stability of marriage. *The American mathematical monthly* 69, 1 (1962), 9–15.
- Hugo Gimbert, Claire Mathieu, and Simon Mauras. 2021. Two-sided matching markets with strongly correlated preferences. In *Fundamentals of Computation Theory: 23rd International Symposium, FCT 2021, Athens, Greece, September 12–15, 2021, Proceedings 23*. Springer, 3–17.
- Dan Gusfield. 1987. Three fast algorithms for four problems in stable marriage. *SIAM J. Comput.* 16, 1 (1987), 111–128.
- Dan Gusfield and Robert W Irving. 1989. *The stable marriage problem: structure and algorithms*. MIT press.
- N Immorlica. 2005. Marriage, honesty, and stability. In *Proceedings of the 16th annual ACM-SIAM symposium on Discrete algorithms, 2005*. 53–62.
- Nicole Immorlica and Mohammad Mahdian. 2015. Incentives in large random two-sided markets. *ACM Transactions on Economics and Computation (TEAC)* 3, 3 (2015), 1–25.
- Yash Kanoria, Seungki Min, and Pengyu Qian. 2021. In which matching markets does the short side enjoy an advantage?. In *Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*. 1374–1386.
- Donald Ervin Knuth. 1976. Mariages stables et leurs relations avec d’autres problèmes combinatoires: introduction à l’analyse mathématique des algorithmes. *Les Presses de l’Université de Montreal* (1976).
- Donald E Knuth, Rajeev Motwani, and Boris Pittel. 1990. Stable husbands. *Random Structures & Algorithms* 1, 1 (1990), 1–14.
- David Manlove. 2013. *Algorithmics of matching under preferences*. Vol. 2. World Scientific.
- Simon Mauras. 2021. *Analysis of random models for stable matchings*. Ph.D. Dissertation. Université Paris Cité.
- David G McVitie and Leslie B Wilson. 1970. Stable marriage assignment for unequal sets. *BIT Numerical Mathematics* 10, 3 (1970), 295–309.
- David G McVitie and Leslie B Wilson. 1971. The stable marriage problem. *Commun. ACM* 14, 7 (1971), 486–490.

- Boris Pittel. 1989. The average number of stable matchings. *SIAM Journal on Discrete Mathematics* 2, 4 (1989), 530–549.
- Boris Pittel. 2019. On likely solutions of the stable matching problem with unequal numbers of men and women. *Mathematics of Operations Research* 44, 1 (2019), 122–146.
- Alvin E Roth. 2002. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica* 70, 4 (2002), 1341–1378.
- Alvin E Roth and Elliott Peranson. 1999. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American economic review* 89, 4 (1999), 748–780.
- LB Wilson. 1972a. An analysis of the stable marriage assignment algorithm. *BIT Numerical Mathematics* 12, 4 (1972), 569–575.
- Leslie B. Wilson. 1972b. An analysis of the stable marriage assignment algorithm. *BIT Numerical Mathematics* 12 (1972), 569–575.

7. APPENDIX: OMITTED PROOFS

In this section, we include the proofs omitted from the main paper of the paper.

Notation. We use $[n]$ to denote $\{1, \dots, n\}$ and $[a, b]$ to denote $\{a, a + 1, \dots, b\}$ for integers $b \geq a$.

7.1. Large deviation inequalities. We recall the Chernoff bound (Dubhashi and Panconesi (2009) Theorem 1.1). Let X be a $\text{Bin}(n, p)$ random variable. Then

$$\mathbb{P}(|X - \mathbf{E}[X]| \geq \delta \mathbf{E}[X]) \leq 2 \exp \{ \delta^2 \mathbf{E}[X] / 3 \}.$$

We say a function $f : \mathcal{X}^n \rightarrow \mathbb{R}$ is k -Lipschitz if for every $x_1, \dots, x_n \in \mathcal{X}$, we have

$$\max_{x'_i \in \mathbf{X}} |f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n)| \leq k.$$

McDiarmid’s inequality (Dubhashi and Panconesi (2009) Corollary 5.2) states that for randomly and independently chosen $X_1, \dots, X_n \in \mathcal{X}$, we have

$$\mathbb{P}(|f(X_1, \dots, X_n) - \mathbf{E}[f(X_1, \dots, X_n)]| \geq t) \leq 2 \exp \left\{ -\frac{2t^2}{k^2 n} \right\}$$

7.2. Sandwiching: Proof of Lemma 7.

Proof of Lemma 7. For $d^{(\ell)} < d < d^{(u)}$, let M_ℓ , M , and M_u be random stable matching instances sampled from $M_{n,\alpha,d^{(\ell)}}^C$, $M_{n,\alpha,d}$, and $M_{n,\alpha,d^{(u)}}^C$. For a candidate c , let ordered subsets L_ℓ , L , and L_u of $[n]$ denote their neighborhood in M_ℓ , M , and M_u respectively.

We sample M_ℓ , M , M_u as follows: Each job has a total order on all the candidates, which are the same in all three instances. The ranking of a job j is just the induced order on its neighbors. For each candidate c , sample d_c from $\text{Bin}(n, d/n)$, and set $d_{\max} := \max\{d_c, d^u\}$. Choose a tuple $(e_1, \dots, e_{d_{\max}})$ where each $e_i \in [n]$ uniformly without repetitions. Set $N_L := (e_1, \dots, e_{d_L})$, $N = (e_1, \dots, e_{d_c})$, and $N_U := (e_1, \dots, e_{d_U})$. Using Chernoff bound, we have that

$$\mathbb{P}(d_c \neq (1 \pm \delta)d) \leq \exp \{ -\delta^2 d / 3 \}.$$

So, with probability at least $1 - n(1 + \alpha)e^{-\delta^2 d / 3}$, we have that $M_L \subseteq M \subseteq M_U$.

The proof follows in a similar fashion for M^J . □

7.3. Probability Bounds for the Deferred Acceptance Algorithms: Proof of Claim 11.

Proof of Claim 11. We compare the random proposals in the CPDA algorithm with a random balls-in-bins process using the following coupling. In the CPDA algorithm, at each time step a new proposal is made by a candidate to a job that has not already rejected them and the time is measured as the number of such proposals.

Consider a balls-in-bins process with n bins, one for each job. Each time a random proposal for a candidate c is required for the DA algorithms, balls are placed in randomly chosen bins until it lands in a bin that corresponds to a job that has not yet rejected c . This job constitutes a new proposal made by c in the CPDA algorithm. At time t in the balls-in-bins process, t balls have been placed. Let b_1^t, \dots, b_n^t be the number of balls in the bins $1, \dots, n$ at time t . Let j_1, \dots, j_n be the number of proposals received by the jobs at time κ .

Let T be the total number of balls placed in bins in order to obtain κ proposals in the CPDA algorithm. Since, the expected number of balls placed to find one that can be used for the algorithm is at most $1 + d/n$, we have that $\mathbf{E}[T] \leq \kappa + O(d \ln n)$. Since $j_i \leq b_i$ by the coupling, Markov's inequality gives us that,

$$(5) \quad \mathbb{P}\left(T \geq \kappa + \tilde{O}(\sqrt{n})\right) \leq 1/\sqrt{n}.$$

To prove the part (1) of Claim 11, let $p_F^\kappa = 1 - \frac{1}{n} \sum_{i \in [n]} \frac{1}{b_i + 1}$ be the probability that the next proposal (at time $\kappa + 1$) is rejected. We will use

$$q_F^\kappa := \max_{t \in [\kappa, \kappa + \tilde{O}(\sqrt{n})]} \left(1 - \frac{1}{n} \sum_{i \in [n]} \frac{1}{b_i^t + 1}\right)$$

as a proxy for p_F^κ , since for every $t \geq \kappa$, we have that

$$0 \leq \frac{1}{n} \sum_{i \in [n]} \left(\frac{1}{j_i + 1} - \frac{1}{b_i^t + 1}\right) \leq \frac{1}{n} \sum_{i \in [n]} (b_i^t - j_i) = \frac{t - \kappa}{n}.$$

Thus it suffices to prove that for $t \in [\kappa, \kappa + \tilde{O}(\sqrt{n})]$, we have

$$(6) \quad \mathbb{P}\left(\left|q_F^t - \left(1 - \frac{n}{n + \kappa}\right)\right| \geq 2\frac{n^2}{\kappa^2}\right) = 2 \exp\left\{-\tilde{\Omega}(n)\right\}.$$

Indeed, since Union bound gives us

$$\begin{aligned} & \mathbb{P}\left(\left|p_F^\kappa - \left(1 - \frac{n}{n + \kappa}\right)\right| \geq 4\frac{n^2}{\kappa^2}\right) \\ & \leq \mathbb{P}\left(\left|q_F^T - \left(1 - \frac{n}{n + \kappa}\right)\right| \geq 3\frac{n^2}{\kappa^2} \mid T \leq \kappa + \tilde{O}(\sqrt{n})\right) + \mathbb{P}(T > \kappa + \tilde{O}(\sqrt{n})) \\ & \leq \sum_{t=\kappa}^{\kappa + \tilde{O}(\sqrt{n})} \mathbb{P}\left(\left|q_F^t - \left(1 - \frac{n}{n + \kappa}\right)\right| \geq 2\frac{n^2}{\kappa^2} \mid T \leq \kappa + \tilde{O}(\sqrt{n})\right) + \frac{1}{\sqrt{n}} \\ & \leq \frac{\sum_{t=\kappa}^{\kappa + \tilde{O}(\sqrt{n})} \mathbb{P}\left(\left|q_F^t - \left(1 - \frac{n}{n + \kappa}\right)\right| \geq 2\frac{n^2}{\kappa^2}\right)}{\mathbb{P}(T \leq \kappa + \tilde{O}(\sqrt{n}))} + \frac{1}{\sqrt{n}} \end{aligned}$$

$$= O\left(\frac{1}{\sqrt{n}}\right).$$

Next, we obtain a bound on $\mathbf{E}[q_F^T]$.

Proposition 16. *For $t \in [\kappa, \kappa + \tilde{O}\sqrt{n}]$, we have $\mathbf{E}[q_F] = 1 - \frac{n}{n+\kappa} \pm 2\left(\frac{n}{\kappa}\right)^2$.*

Proof. Let $\mu := \mathbf{E}[1 + b_i^t] = 1 + \frac{t}{n}$. By Jensen's inequality, we have

$$\mathbf{E}[1 - q_F^t] = \mathbf{E}\left[\frac{1}{1 + b_i^t}\right] \geq \frac{1}{\mathbf{E}[1 + b_i^t]} = \frac{n}{n + t}.$$

For the other direction, we have:

$$\begin{aligned} \mathbf{E}[1 - q_F^t] &= \mathbf{E}\left[\frac{1}{n} \sum_{i \in [n]} \frac{1}{b_i^t + 1}\right] \\ &= \frac{1}{n} \sum_{i \in [n]} \mathbf{E}\left[\frac{1}{b_i^t + 1}\right] \\ &= \frac{1}{\mu \cdot n} \sum_{i \in [n]} \mathbf{E}\left[\frac{1}{1 + \frac{b_i^t + 1 - \mu}{\mu}}\right] \\ &\leq \frac{1}{\mu \cdot n} \sum_{i \in [n]} \mathbf{E}\left[1 - \left(\frac{b_i^t + 1 - \mu}{\mu}\right) + \left(\frac{b_i^t + 1 - \mu}{\mu}\right)^2\right] \\ &= \frac{n}{n + t} + \frac{n}{n + t} \mathbf{E}\left[\left(\frac{b_1^t + 1 - \mu}{\mu}\right)^2\right] \\ &\leq \frac{n}{n + t} + \left(\frac{n}{t}\right)^2 \\ &\leq \frac{n}{n + \kappa} + 2\left(\frac{n}{\kappa}\right)^2. \end{aligned}$$

The second last inequality follows from the fact that b_i^t is a $\text{Bin}(t, \frac{1}{n})$ random variable, so $\text{var}(b_i^t) = \mathbf{E}[(b_i^t + 1 - \mu)^2] = \frac{t}{n} \left(1 - \frac{1}{n}\right)$. \square

With Proposition 16 in hand, (6) follows from McDiarmid's inequality. Indeed, consider a sequence $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_t)$ of t randomly chosen jobs, and define $b_i^t(\mathcal{X}) = |\{j \in [t] \mid \mathcal{X}_j = i\}|$. Let us denote

$$q(\mathcal{X}) = \frac{1}{n} \sum_{i \in [n]} \frac{1}{b_i^t(\mathcal{X}) + 1}.$$

Consider two sequences of jobs \mathcal{Y} and \mathcal{Z} that match in all but the i 'th proposal, i.e., $\mathcal{Y}_j = \mathcal{Z}_j$ for $j \neq i$. It is clear that $|q(\mathcal{Y}) - q(\mathcal{Z})| \leq \frac{2}{n}$. Thus $q_F(\cdot)$ is $\frac{2}{n}$ -Lipschitz, and so after a random sequence of t proposals, we have that for every $\alpha > 0$,

$$\mathbb{P}(|q_F^t - \mathbf{E}q_F^t| > \alpha) \leq 2 \exp\{-2\alpha^2 n^2 / \kappa\}.$$

It remains to plug in $\alpha = 3\frac{n^2}{\kappa^2}$, observe that $\kappa = \Theta(n \ln n)$, and union bound over all for $t \in [\kappa, \kappa + \tilde{O}\sqrt{n}]$ to finish the proof of (6).

To prove part (2) in Claim 11, we observe that in the coupling described above,

$$\{i \mid j_i = 0\} \subseteq \{i \mid b_i^\kappa = 0\}.$$

The expected number of empty bins at step κ is $n(1 - 1/n)^t \leq ne^{-\kappa/n} = e^C$. So by Markov's inequality,

$$\mathbb{P}(|\{i \mid j_i = 0\}| > e^{2C}) \leq \mathbb{P}(|\{i \mid b_i^T = 0\}| > e^{2C}) \leq e^{-C}.$$

To prove part (3) in Claim 11, we have

$$\begin{aligned} \mathbb{P}(\{i \mid j_i = 0\} = \emptyset) &= \mathbb{P}(\{i \mid b_i^T = 0\} = \emptyset) \\ &\leq \mathbb{P}(\{i \mid b_i^T = 0\} = \emptyset \mid T < \kappa + \tilde{O}(\sqrt{n})) + \mathbb{P}(T \geq \kappa + \tilde{O}(\sqrt{n})) \\ &\leq \mathbb{P}(\{i \mid b_i^{\kappa + \tilde{O}(\sqrt{n})} = 0\} = \emptyset \mid T < \kappa + \tilde{O}(\sqrt{n})) + \frac{1}{\sqrt{n}} \\ &\leq \frac{\mathbb{P}(\{i \mid b_i^{\kappa + \tilde{O}(\sqrt{n})} = 0\} = \emptyset)}{\mathbb{P}(T < \kappa + \tilde{O}(\sqrt{n}))} + \frac{1}{\sqrt{n}} \\ &\leq 2 \exp(-C/2) + \frac{1}{\sqrt{n}}. \end{aligned}$$

□

7.4. Expected number of rounds in JPDA: Proof of Claim 12.

Proof of Claim 12. We again do this by coupling with balls-in-bins process, one bin for each candidate. We abuse notation and use $[m]$ to denote both the set of bins and the set of candidates. Consider the following coupling between the processes of JPDA and balls-in-bins. Let us denote $\tau_0 = (1 + \gamma/2)m \cdot \ln\left(\frac{1+\alpha}{\alpha+1/m}\right)$.

Suppose at a point in the JPDA algorithm, a job $j \in [n]$ is required to make a (random) proposal, and so far, \mathbf{C}_j is the set of candidates that j has already proposed to. We keep throwing balls-into-bins uniformly at random until we hit a bin $\mathbf{x} \notin \mathbf{C}_j$. We assign \mathbf{x} as the next proposal of j . Thus the distribution of the next proposal of j is uniform on $[m] \setminus \mathbf{C}_j$.

Let τ and τ' denote the number of proposals made and the number of balls thrown in the above algorithm, respectively. Clearly $\tau \leq \tau'$, and so $\mathbf{E}[\tau] \leq \mathbf{E}[\tau']$. Moreover, let P and P' denote the candidates who have gotten at least one proposal and the bins that have gotten at least one ball just before the final throw. By the coupling, we have $P = P'$, and since there are at most n matched candidates by the end, $|P| = |P'| \leq n - 1$, since the JPDA algorithm terminates at the moment when n candidates have a match. We have

$$\begin{aligned} \mathbb{P}(\tau > \tau_0) &\leq \mathbb{P}(\tau' > \tau_0) \\ &\leq \mathbb{P}(|P'| < n \wedge \tau' > \tau_0) \\ &\leq \mathbb{P}(|P'| < n \mid \tau' > \tau_0). \end{aligned}$$

We upper bound this probability by

$$\binom{m}{\alpha n + 1} \left(1 - \frac{\alpha n}{m}\right)^{\tau'} \leq \left(\frac{e(1+\alpha)}{\alpha+1/m}\right)^{\alpha n} \left(\frac{1}{1+\alpha}\right)^{\tau'}.$$

Therefore,

$$\begin{aligned}
\mathbf{E}[\tau] - \tau_0 &\leq \sum_{t > \tau_0} \mathbb{P}(\tau > t) \\
&\leq \left(\frac{e(1+\alpha)}{\alpha + 1/m} \right)^{\alpha n} \sum_{t \geq \tau_0} \left(\frac{1}{1+\alpha} \right)^t \\
&\leq \left(\frac{e(1+\alpha)}{\alpha + 1/m} \right)^{\alpha n+1} \left(\frac{1}{1+\alpha} \right)^{\tau_0} \\
&\leq \left(\frac{e(1+\alpha)}{\alpha + 1/m} \right)^{\alpha n+1} e^{-\alpha \tau} \\
&\leq e^{\alpha n+1} \left(\frac{\alpha + 1/m}{1+\alpha} \right)^{\gamma/2\alpha n} \\
&= o(1).
\end{aligned}$$

□

(AP) DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE, YORK UNIVERSITY, TORONTO, ON M3J 1P3, CANADA

Email address: `apotu@yorku.ca`

(SS) DEPARTMENT OF COMPUTER SCIENCE, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267, USA

Email address: `shikha@cs.williams.edu`