# Tiered Random Matching Markets: Rank Is Proportional to Popularity

## Itai Ashlagi
Department of Management Science and Engineering, Stanford University, CA, USA
iashlagi@stanford.edu

## Mark Braverman
Department of Computer Science, Princeton University, NJ, USA
mbraverm@cs.princeton.edu

## Amin Saberi
Department of Management Science and Engineering, Stanford University, CA, USA
saberi@stanford.edu

## Clayton Thomas
Department of Computer Science, Princeton University, NJ, USA
claytont@cs.princeton.edu

## Geng Zhao
Department of Computer Science, Stanford University, CA, USA
gengz@stanford.edu

—————— **Abstract** ——————

We study the stable marriage problem in two-sided markets with randomly generated preferences. Agents on each side of the market are divided into a constant number of "soft" tiers, which capture agents' qualities. Specifically, every agent within a tier has the same public score, and agents on each side have preferences independently generated proportionally to the public scores of the other side.

We compute the expected average rank which agents in each tier have for their partners in the man-optimal stable matching, and prove concentration results for the average rank in asymptotically large markets. Furthermore, despite having a significant effect on ranks, public scores do not strongly influence the probability of an agent matching to a given tier of the other side. This generalizes the results by Pittel [20], which analyzed markets with uniform preferences. The results quantitatively demonstrate the effect of competition due to the heterogeneous attractiveness of agents in the market.

## 1 Introduction

The theory of stable matching, initiated by Gale and Shapley [9], has led to a deep understanding of two-sided matching markets and inspired successful real-world market designs. Examples of such markets include marriage markets, online dating, assigning students to

12th Innovations in Theoretical Computer Science Conference (ITCS 2021).
Editor: James R. Lee; Article No. 46; pp. 46:1–46:16

Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

schools, labor markets, and college admissions. In a market matching "men" to "women" (a commonly used analogy), a matching is stable if no man-woman pair prefer each other over their assigned partners.

A fundamental issue is characterizing stable outcomes of matching markets, i.e. the outcome agents should *expect* based on market characteristics. Such characterizations are not only useful for describing outcomes but also likely to be fruitful in market designs. Numerous papers so far have studied stable matchings in random markets, in which agents' preferences are generated uniformly at random [20, 16, 3, 22]. This paper contributes to the literature by expanding these results to a situation where preferences are drawn according to different tiers of "public scores", generalizing the uniform case. We ask how public scores, which correspond to the attractiveness of agents, impact the outcome in the market.

Formally, we study the following class of *tiered random markets*. There are $n$ men and $n$ women. Each side of the market is divided into a constant number of "soft tiers". There is a fraction of $\epsilon_i$ women in tier $i$, each of which has a public score $\alpha_i$. And there is a fraction of $\delta_j$ men in tier $j$, each of which has a public score $\beta_j$. For each agent we draw a complete preference list by sampling without replacement proportionally to the public scores of agents on the other side of the market.[1] So a man's preference list is generated by sampling women one at a time without replacement according to a distribution that is proportional to their public scores. Using $\boldsymbol{\alpha}, \boldsymbol{\epsilon}$ to denote the vector of scores and proportions of tiers on the women's side, we see that the marginal probability of drawing a woman in tier $i$ is $\alpha_i/(n\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha})$. An analogous statement holds for the tier configuration $\boldsymbol{\beta}, \boldsymbol{\delta}$ of the men. These preferences are a natural next-step beyond the uniform distribution over preference lists, and provide a priori heterogeneous quality of agents while still being tractable to theoretical analysis.

Our primary goal is to study the *average rank* of agents in each tier under the man-optimal stable matching, with a focus on the asymptotic behavior in large markets. The rank of an agent is defined to be the index of their partner on their full preference list, where lower is better. Additionally, we prove results on the *match type distribution*, i.e. the fraction of tier $i$ women matched to tier $j$ men (for each $i, j$).

We show that, for large enough markets, the following hold to within an arbitrarily small approximation factor:

**(i)** (Theorem 4.8.) With high probability, the average rank of men in tier $j$ is

$$\frac{\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha}}{\alpha_{\min}} \cdot \frac{1}{\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1}} \cdot \frac{\ln n}{\beta_j}.$$

**(ii)** (Theorem 5.1.) With high probability, the average rank of women in tier $i$ is

$$(\boldsymbol{\delta} \cdot \boldsymbol{\beta})(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1})\frac{\alpha_{\min}}{\alpha_i} \frac{n}{\ln n}.$$

**(iii)** (Theorem 5.2.) The probability that a woman in tier $i$ matches to a man in tier $j$ is $\delta_j$.

In the above, $\boldsymbol{\beta}^{-1} = \{1/\beta_j\}$ denotes the vector of the reciprocals of men's public scores, $\alpha_{\min}$ denotes the smallest public score on the women's side, and $\boldsymbol{x} \cdot \boldsymbol{y}$ denotes the dot product of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

---

[1] These are also termed popularity-based preferences [10, 13] and also equivalent to generating preferences according to a Multinomial-Logit (MNL) induced by the public scores.

**Intuition and Observations.**    As in the case of uniform preferences [20], in the man-optimal stable outcome, men get a much lower rank than women. Indeed, both men and women get the same order of rank as in the uniform case ($\ln n$ and $n/\ln n$, respectively). This in itself is an interesting consequence of this work – a constant tier structure affects the market only up to constants. This fact also highlights that determining these constants is an interesting area for investigation, as the constants capture how the outcome of the market changes with respect to the public scores. The first observation we make is that agents on each side get a rank inversely proportional to their public score.

Perhaps more interesting is the following observation: The rank of both sides depends on the tier structure of the other side, but *each tier is affected the same amount* by the tier parameters of the other side. This is closely related to the fact that the probability of a woman matching to a man in tier $j$ is proportional to only the number of men in tier $j$ (regardless of the tier the woman lies in). Moreover, both $\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha}/\alpha_{\min}$ and $(\boldsymbol{\delta} \cdot \boldsymbol{\beta})(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1})$ are always greater than or equal to one[2]. Thus, in these markets, any heterogeneity in the public scores of one side harms the average ranks of the other side (but does not significantly affect the likelihood that an agent matches to a certain tier on the other side).

Another interesting feature is the following: While the average ranks for men's tiers depend on public score distributions on both sides of the market, the average rank of women in tier $i$ depends only on the ratio between $\alpha_i$ and the public score $\alpha_{\min}$ of the bottom tier of women (and the distribution of public scores on the men's side). Intuitively, the rank of the men depends on the distribution of scores of the women because *men are competing to avoid being matched to the lowest tier of women.*

To elaborate on that last point, let us first consider the total number of proposals made during the man-proposing deferred acceptance process (DA). The algorithm will terminate when the last woman receives a proposal. Naturally one would expect that this woman will belong to the bottom tier. Therefore, using standard coupon collector arguments, the total number of proposals made to women *in the bottom tier* until they all receive a proposal is expected to be $(\epsilon_{\min}n)\ln(\epsilon_{\min}n)$, where $\epsilon_{\min}$ is the fraction of women in the bottom tier. These proposals are a $\epsilon_{\min}\alpha_{\min}/\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha}$ fraction of the total proposals, so one expects the number of total proposals to be

$$\frac{(\epsilon_{\min}n)\ln(\epsilon_{\min}n)}{\epsilon_{\min}\alpha_{\min}/\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha}} = \frac{\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha}}{\alpha_{\min}} \cdot n\ln n - O(n).$$

This introduces the factor of $\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha}/\alpha_{\min}$ in result (i) on the men's ranks (i.e. the number of proposals per man).

On the other hand, the probability that one of these proposals goes to a woman in tier $i$ is $\alpha_i/(n\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha})$, implying that such a woman should receive roughly $(\alpha_i/\alpha_{\min})\ln n$ proposals. Thus, for a given woman, the increase in the total number of proposals caused by the tier proportions $\boldsymbol{\epsilon}$ is exactly canceled out by the likelihood that a proposal goes to that woman, and the only thing that matters is the woman's score (relative to the bottom tier). If men are uniform, women should then expect rank roughly $(\alpha_{\min}/\alpha_i)(n/\ln n)$, which helps explain the corresponding factors in result (ii).

Consider now the public scores of the men, and for simplicity assume that the bottom tier of men has score 1. Suppose for the sake of demonstration that every time a man with public score $\beta_j$ proposes to a woman who is already matched, this man is $\beta_j$ times more likely to

---

[2]  To prove $(\boldsymbol{\delta} \cdot \boldsymbol{\beta})(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1}) \geq 1$, use Jensen's inequality to conclude that $\sum_j \delta_j\beta_j \geq \left(\sum_j \delta_j\beta_j^{-1}\right)^{-1}$.

be accepted than a man with than a man with public score 1.[3] We would expect that such a man makes a $1/\beta_j$ fraction fewer proposals before his next acceptance, and indeed $1/\beta_j$ fewer proposals overall. Let $S$ be the total number of proposals, let $r_j$ denote the rank of a man in tier $j$, and $r_{\min}$ the rank of the bottom tier of men. If every tier of size $\delta_j n$ each accounts for a share of proposals proportional to $1/\beta_j$, then we should have

$$S = \sum_j (n\delta_j)\beta_j^{-1} r_{\min} \quad \Longrightarrow \quad r_{\min} = \frac{S}{n\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1}}, \qquad r_j = \frac{S}{(n\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1})\beta_j},$$

which introduces the factor of $1/((\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1})\beta_j)$ in result (i) on the men's rank.

The final remaining factor in our results is $(\boldsymbol{\delta} \cdot \boldsymbol{\beta})(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1})$ in result (ii). Deriving this term requires reasoning about the number of proposals from each tier of men received by a fixed woman $w$. Building from the previous paragraph, we reason that each of the $\delta_j n$ men in tier $j$ makes a number of proposals proportional to $1/\beta_j$. Each such proposal has the same probability of going to $w$, regardless of the tier $j$. So the number of proposals $w$ receives from tier $j$ men is proportional to $\delta_j/\beta_j$. The factor $(\boldsymbol{\delta} \cdot \boldsymbol{\beta})(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1})$ then arises for somewhat technical reasons (described in Section 5) which have to do with the way women generate their preference lists.

We now describe how result (iii), which may seem somewhat more mysterious than the other results, emerges as a corollary of computing the ranks women receive. We argued above that a woman $w$ in tier $i$ receives approximately $(\delta_j/\beta_j)U_i$ proposals from men in tier $j$, for some value of $U_i$ independent of $j$. Recall that $w$ applies weight $\beta_j$ to each proposal she sees from a man in tier $j$. Moreover, the identity of $w$'s favorite proposal is independent of the order in which $w$ saw proposals. Thus, the probability that $w$'s favorite proposal (i.e. the proposal of the man she matches to) came from tier $j$ is approximately $(\delta_j U_i)/U_i = \delta_j$, which is *independent of* $\beta_j$, as well as independent of the tier $w$ is in. Thus, up to lower order terms, the distribution of match types is the same as it would be in a uniformly random matching market, and the match is not assortative.

Intuitively, result (iii) arises when men make enough proposals to offset any disadvantage (in the type of their match) they have due to public score. Due to the highly connected and relatively competitive nature of our markets, men in the lowest tier make more proposals, but they are not more likely to end up matched with lower tier agents. Put another way, men in lower tiers are less likely to attain matches they idiosyncratically like, but often settle for a high-quality agent which is low on their personal preference list. This indicates that public scores that differ by constant weight do not provide any significant a priori predictive power over the matches agents receive. In particular, agents with lower public scores can still hope to achieve high-tier matches if they consider enough options.

**Techniques.**  Our proofs require developing some technical tools that may be of independent interest, especially when we reason about the ranks achieved by the men. We build on the analysis of DA from [23, 20, 13, 3] to handle public scores rather than just uniform random preferences. As in these previous works, a key step in our proof is letting all men but one (call him $m$) first propose and match though DA, and then tracking the proposals of $m$ (this works because DA is independent of the order of proposals). For demonstration purposes,

---

[3]  As we discuss below, this approximation is only valid if the woman is already matched with a man she ranks highly. A major technical step in our proof is showing that, in certain situations, "enough" women are "matched well enough" for this approximation to be used.

let's call the proposals before man $m$ the "setup". A key fact in previous works is that the distribution of proposals made by $m$ is identical for every man, and moreover that the distribution of setups is identical as well. This fails to hold in tiered random markets, and thus we must develop new techniques.

We prove that, for "most" setups, the rank a man can achieve is approximately given by a certain geometric distribution, whose parameter $p$ is essentially the probability that a proposal by that man will be accepted. We then prove that, up to lower order terms, this success parameter scales up with the public score of the men. This gives the fact that the rank of men is inversely proportional to public score.

Characterizing the setups where our proof goes through requires a technical analysis, and we term the setups which work "smooth matching states". The most crucial thing we need for these setups is that *many women are matched to partners they rank highly*, which helps us prove that 1) men are likely to remain matched to their first acceptance (so our approximation with a geometric distribution is valid), and 2) a man with fitness $\beta$ is approximately $\beta$ times more likely to be accepted every time. For details, see Section 4.

Finally, to prove that the average rank of men within a tier *concentrates*, we need to show the correlation between the ranks of different men is not too large. Thus, we track the proposals of the last *two* men to propose, and find that the joint distribution of the ranks of these men can be approximated by a pair of independent geometric distributions. Intuitively, this is because men do not propose to very many women overall, and thus the last two men are unlikely to interfere with each other as they make proposals.

The crucial aspects of our model are that preferences of each agent are independent and identically distributed, that preference weights are constant, and that the market is roughly in balance. While our techniques are useful to reason about markets which do not have these properties, the results are not nearly as clean; indeed the tier structure simplifies our analysis, but most of it goes through if each agent has an individual, constant, bounded public score.

## 1.1 Related literature

Several papers have studied matching markets with complete preference lists that are generated uniformly at random. Coupon collector techniques are used in [23] to upper bound the men's average rank by $\ln n$. The papers [20, 16, 21] analyze further balanced markets with $n$ men and $n$ women. They find that in the man-optimal stable matching in balanced markets, men and women match on average to their $\ln n$ and $\frac{n}{\ln n}$ ranks, respectively. Our results generalize these findings to markets with preferences induced by public scores, thus incorporating much more heterogeneity in the market.

Several papers study markets with uniformly drawn preferences and an imbalance between men and women ([3, 22, 6]). These papers find that in any stable matching the average ranks of men and women are similar to the average ranks under the short-side-proposing DA. Additionally, [14] investigates the relation between the imbalance and the length of preference lists (though the model is still uniform for each agent). This paper does not consider imbalanced markets but we believe that similar techniques to those we develop will be useful to reason about unbalanced tiered random markets.

Several papers look at random matching markets in which preferences are generated based on public scores [13, 17, 1]. These papers restrict attention to the size of the core (a measure of the difference between the man-optimal and woman-optimal outcome) and strategic manipulation of agents under a stable matching mechanism. Key assumptions in

these papers generate outcomes which leave many agents unmatched. In particular, their models either assume that preference lists of men are of constant length, or, alternatively, one side has many more agents than the other.[4]

Closely related to this paper is [10], which primarily studies a special case of highly correlated popularity preferences which is termed "geometric preferences". While our work focuses on the rank agents achieve in the man-optimal outcome (a canonical stable matching), [10] focuses on the size of the core (more specifically, they study the number of stable partners that agents have in typical stable matchings) using techniques specialized to geometric preferences.

Other papers have addressed tiered matching markets, especially in market design settings. However, these papers mostly study "hard tiers", i.e. such that agents in higher tiers are deterministically ranked above lower tiers by every agent on the other side. Examples include [4, 2]. [18] also considers a certain restricted tiered model of cardinal utilities (which is incomparable with our model), focusing on which tier of agents match to which tier.

Our contribution to the literature is a detailed study of "soft tiers", a natural special case of the popularity preferences of [13, 17, 10]. In cases where each agent's utility for each match on the other side is independent and identically distributed, popularity preferences are the natural next step beyond uniform markets, as they model situations where agents on each side have significant but non-definitive variation in a priori quality. Our techniques build on the large body of work analyzing the "proposal dynamics" of deferred acceptance for random preferences, such as [23, 13, 3, 10]. Our results give insight into how constant-factor preference biases affect stable matching markets, including the first explicit calculations of expected rank beyond uniform markets.

The rest of the paper is organized as follows: Section 2 offers basic definitions and preliminaries for our discussion. Section 3 studies the tiered coupon collector process, which serves as an important coupling process for the deferred acceptance algorithm. Section 4 and 5 present the core results of this paper, namely the average rank among tiers of men and women. For missing proofs, see the full version of this paper.

## 2     Definitions and Preliminaries

A matching market consists of a finite set of men $M$ and a finite set of women $W$. Each man (woman) has a complete and strict preference list over women (men). A matching is a mapping $\mu : M \cup W \to M \cup W$ such that: for every $m \in M$, $\mu(m) \in W$ (or $\mu(m)$ is undefined), for every woman $w \in W$, $\mu(w) \in M$ (or $\mu(w)$ is undefined), and for every $m \in M$ and $w \in W$, $\mu(m) = w$ if and only if $\mu(m) = w$. A matching $\mu$ is *stable* if no man-woman pair who are not matched in $\mu$ prefer each other to their matched partners.

It is well-known that there is a unique man-optimal stable matching, which can be found using the man-proposing deferred acceptance algorithm (DA). While this algorithm does not fully specify an execution order, it is a classically known result that the order does not affect the final outcome.

▶ **Lemma 2.1** ([9, 19]). *The same proposals are made in every run of DA, regardless of which man is chosen to propose at each step.*

---

◼ **Algorithm 1** (Man-Proposing) Deferred Acceptance Algorithm (DA).

---
**1** Initialize matching $\mu$ to be empty (i.e. every agent's partner is undefined);
**2** Initialize $\mathcal{U} = M$ to be the set of all unmatched men;
**3 while** $|\mathcal{U}| > 0$ **do**
**4**     Choose any $m \in \mathcal{U}$;
**5**     Let $m$ propose to his most preferred woman $w$ to whom he has not made a
       proposal yet;
**6**     **if** $w$ *prefers* $m$ *to* $\mu(w)$ *(or if* $\mu(w)$ *is undefined)* **then**
**7**        **if** $\mu(w)$ *is defined* **then** Add $\mu(w)$ to $\mathcal{U}$;
**8**        Remove $m$ from $\mathcal{U}$;
**9**        Assign $\mu(w) = m$;
**10**    **end**
**11 end**

---

We study the man-optimal stable matching in a class of tiered random markets, which will be defined below. We will assume that $|M| = |W|$ and that no agent finds any other agent on the other side unacceptable. We will also assume that each side draws their preferences from an identical and independent underlying distribution, and moreover these preferences are generated by repeatedly sampling without replacement from a fixed distribution on the *agents* of each side. In [13, 10], this assumption is termed "popularity-based preferences", with the weight of an agent in the distribution intuitively indicating their popularity for agents on the other side.

Our main goal is to study randomized matching markets with *a constant number of constant weight tiers* of agents on each side. For this entire paper, we consider the tier structure to be defined by fixed proportions $\epsilon, \delta$ of agents in each tier and constant weights $\alpha, \beta$ for each tier, and we investigate the outcome of the man-proposing DA as $n \to \infty$.

▶ **Definition 2.2.** *Consider constant vectors* $\alpha, \epsilon \in \mathbb{R}_{>0}^{k_1}$ *and* $\beta, \delta \in \mathbb{R}_{>0}^{k_2}$, *where* $\|\epsilon\|_1, \|\delta\|_1 = 1$. *A tiered matching market of size* $n$ *with respect to* $\alpha, \epsilon, \beta, \delta$ *is defined by generating agents' preference lists as follows:*

- *The set of* $n$ *women* $W$ *is divided into tiers* $T_1, \ldots, T_{k_1}$, *of size* $|T_i| = \epsilon_i n$ *each[5]. Define a distribution* $\mathcal{W}$ *on women such that a woman in tier* $i$ *is selected with probability proportional to* $\alpha_i$. *That is, the weight of* $w \in T_i$ *in* $\mathcal{W}$ *is* $\alpha_i/(n\epsilon \cdot \alpha)$ *(which we often denote by* $\pi_i$*).*

- *The set of* $n$ *men* $M$ *is divided into tiers* $T_1, \ldots, T_{k_2}$, *of size* $|T_j| = \delta_j n$ *each. Define a distribution* $\mathcal{M}$ *on men such that a man in tier* $j$ *is selected with probability proportional to* $\beta_j$. *That is, the weight of* $m \in T_j$ *in* $\mathcal{M}$ *is* $\beta_j/(n\delta \cdot \beta)$.

*For each man* $m$ *independently, women are repeatedly sampled from* $\mathcal{W}$ *without replacement, and the order in which women are selected is* $m$*'s preference list. Preferences for the women are analogously drawn over the distribution* $\mathcal{M}$. *The rank that a man has for a woman* $w$ *is the index of* $w$ *on his preference list (where lower is better).*

---

[5] Note that, for most vectors $\epsilon, \delta$, many values of $n$ will produce tier sizes which are not integers. However, as all our results are *continuous* in $\epsilon, \delta$ this is not a problem – for any particular fixed $n$, each tier size can be rounded in a way that effectively just changes $\epsilon, \delta$ by a tiny amount, and our results will still hold as written as $n \to \infty$.

We refer to each $\alpha_i$ as the *weight* or *public score* of the women in tier $i$, and similarly for the men. For simplicity of certain arguments, we assume that each $\alpha_i \geq 1$ and each $\beta_j \geq 1$ (although for clarity of our results, we do not assume that the smallest weight is exactly 1). We write $\alpha_{\min}$ for the weight of the bottom tier of women, and $\epsilon_{\min}$ for the corresponding tier proportion.

Using a simple generalization of the "principle of deferred decisions" used in [15], we can arrive at a characterization of the random process of running DA with a tiered matching market.

▶ **Lemma 2.3.** *The distribution of runs of DA for a tiered matching market can be generated as follows: For the men, every time a man is chosen to propose, he samples a woman at random from $\mathcal{W}$, and repeats this until he samples a woman who he has not yet proposed to.*

*For the women, suppose $w$ has seen proposals from a set of men $p(w)$, and let $\Gamma_w = \sum_{m \in p(w)} \beta(m)$, where $\beta(m)$ denotes the public score of a man $m \in p(m)$. Then if a proposal from a man $m_*$ with public score $\beta_*$ arrives, $w$ accepts the proposal from $m_*$ with probability*

$$\frac{\beta_*}{\beta_* + \Gamma_w}.$$

**Proof.** The above formula gives the probability that $m_*$ is chosen as $w$'s favorite out of the set of men $p(w) \cup \{m_*\}$. The only additional observation we need to make is that the probability that $m_*$ is the new favorite is independent of the identity of the old favorite.  ◀

We often call $\Gamma_w$ the total "weight of proposals" woman $w$ has seen at some point during DA.

## 2.1 Deferred acceptance with re-proposals

With respect to any popularity-based model of preferences, we can define a procedure analogous to DA. In our case, we will show that the difference between DA and this procedure is indeed small.

▶ **Definition 2.4.** *Consider any random matching market with men's preferences determined by sampling from a distribution $\mathcal{W}$ over women. The* deferred acceptance with re-proposals *algorithm is defined as being identical to Algorithm 1, except*
- *Every time a man is chosen to propose to a woman, he draws a woman from $\mathcal{W}$ with replacement, and may propose more than once to a single woman.*
- *Women's preferences are consistent throughout proposals from the same man (so if a woman rejected a man before, she will reject him again).*

Since re-proposals are ignored, this process will always yield the same outcome as algorithm 1.

**Notation.** We write $x = (1 \pm \epsilon)y$ to mean $(1 - \epsilon)y \leq x \leq (1 + \epsilon)y$. We let $\epsilon$ denote an arbitrarily small constant greater than 0, while $\boldsymbol{\epsilon}$ and $\epsilon_i$ denote the tier parameters of the women. We let $\alpha_{\min}$ denote the smallest public score for the women's side, and $\epsilon_{\min}$ denotes the corresponding tier proportion. We let $\boldsymbol{v} \cdot \boldsymbol{w}$ denote the inner product of vectors $\boldsymbol{v}, \boldsymbol{w}$. We denote the exponential and geometric distributions by $\mathrm{Exp}(\lambda)$ and $\mathrm{Geo}(p)$, respectively. We denote the fact that a random variable $X$ is a draw from a distribution $D$ by $X \sim D$. We use $X \preceq Y$ to denote the fact that $X$ is statistically dominated by $Y$ (i.e. for all $t \in \mathbb{R}$, we have $\mathbb{P}[X \geq t] \leq \mathbb{P}[Y \geq t]$). We let $\mathrm{Cov}(X, Y)$ denote the covariance of $X$ and $Y$. We write $f(n) = \widetilde{O}(g(n))$ if there exists a constant $k$ such that $f(n) = O(g(n) \log^k(g(n)))$.

## 3   The Coupon Collector and the Total Number of Proposals

Fix a tier structure $\boldsymbol{\alpha}, \boldsymbol{\epsilon}$ corresponding to men's preferences over the women. Consider running deferred acceptance with re-proposals. Recall that each man samples a woman in tier $i$ with probability $\pi_i = \alpha_i/(n\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha})$ each draw. Define $\pi_{\min} = \alpha_{\min}/(n\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha})$ as the probability of drawing a woman in the lowest tier (and keep in mind that $\pi_{\min}$ scales like $O(1/n)$).

The core tool we use to reason about the total number of proposals in DA is the classically studied coupon collector process. In particular, we study this process when coupons from different tiers are drawn with a constant-factor difference in probability.

▶ **Definition 3.1.** *Given a probability distribution $(p_i)_{i \in [n]}$, we define the* coupon collector with unequal probabilities *as follows: once every time step, an integer from $[n]$ is drawn independently and with replacement according to distribution $(p_i)_{i \in [n]}$. The coupon collector random variable with respect to $(p_i)_{i \in [n]}$ is defined as the number of total draws required before every integer in $[n]$ has appeared at least once.*

*The coupon collector $T$ which we are interested in is defined by taking the distribution $\mathcal{W}$ of men's preferences.*

We will show in Section 3.1 that, in our case, this random process is also very close to that of DA (without re-proposals). For now, we simply bound the expectation of the coupon collector (with the proof deferred to the full version). Note that similar probabilistic problems have been considered before (see e.g. [5, 8]) but we include our own full proofs in the full version for completeness.

▶ **Theorem 3.2.** *Let $T$ denote the number of draws in a coupon collector process with weights proportional to $\mathcal{W}$. We have*

$$\mathbb{E}[T] = \left(1 \pm O(1/\ln n)\right) \frac{\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha}}{\alpha_{\min}} n \ln n.$$

▶ Remark 3.3. While we are mostly interested in the asymptotic performance of these matching markets, we make one comment here that the above big-$O$ notation hides a constant factor of order $\ln(1/\epsilon_{\min})$. For small values of $\epsilon_{\min}$, this can be much larger than $\ln n$ for most realistic market sizes. Note that this error term already showed up in the intuition given in Section 1, where our estimate for the total number of proposals had an additive term of $O(\ln(\epsilon_{\min})n)$. For more information, see the discussion of coupon collector lower bound in the full version of this paper.

### 3.1   The Total Number of Proposals in Deferred Acceptance

Let $S = S_n$ denote the total number of proposals made a run of DA with random preferences given by our tiered market. As before, let $T = T_n$ denote the distribution of a coupon collector with distribution $\mathcal{W}$. As in many prior studies of randomized deferred acceptance, our starting point is the fact that $S$ is statistically dominated by $T$:

The connection to stable matchings is the following very simple observation, which has been used in many previous works [16, 20]:

▶ **Proposition 3.4.** *The coupon collector random variable $T$ is distributed identically to the total number of proposals made in deferred acceptance with re-proposals (regardless of the preferences that women have for men).*

*Moreover, if $S$ is the number of proposals in DA, then $S \preceq T$ (i.e. $S$ is statistically dominated by $T$).*

**Proof.** First, recall that DA terminates as soon as every man is matched. Observe that women never return to being unmatched once they receive a single proposal. Because the market is balanced (i.e. $|W| = |M|$), this means DA will terminate as soon as every woman has been proposed to. Moreover, because re-proposals are allowed, every proposal is distributed exactly according to $\mathcal{W}$. Thus, ignoring the identity of the man doing the proposing, $T$ is distributed exactly according to the coupon collector random process.

Furthermore, we can recover the exact distribution $S$ of proposal in DA simply by ignoring each repeated proposal in $T$. Thus, $S \leq T$ for each run of deferred acceptance with re-proposals, so $S \preceq T$. ◀

We proceed to show that the upper bound provided by $T$ is essentially tight, i.e. there is not a big difference between $T$ and $S$. The key step will be to upper bound maximum number of distinct women any man proposes to in $S$, and thus upper bound the probability that any proposal in $T$ is a repeat for the man making the proposal. Crucially, this lemma will have to account for the preferences of the women (which up until this point have been ignored, but which play a significant role in the distribution of proposals in DA). Recall that we denote the sizes of the tiers of the men by the vector $\boldsymbol{\delta}$, and the public scores of the men in each tier by $\boldsymbol{\beta}$.

▶ **Lemma 3.5.** *Consider running DA with all men except $m_*$, and suppose that at most $O(n \ln n)$ proposals are made during this process. Afterwards, consider $m_*$ joining and run DA until the end. Then for any $C \geq 0$, with probability $1 - 1/n^C$, the number of proposals made by $m_*$ is at most $O(C \ln^2 n)$.*

**Proof.** This proof follows a similar logic as the proof of Lemma B.4 (ii) in [3]. Suppose $m_*$ has public score $\beta_*$, and that he proposes at the end (and $O(n \ln n)$ prior proposals have been made). We proceed as follows:

1. When $m_*$ makes a proposal, he will choose a woman who he has not yet proposed to. For some fixed proposal index $i$ of $m_*$, let's denote the set of all women $m_*$ has not proposed to by $W_*$, and denote by $\mathcal{W}_*$ the distribution of $m_*$'s next proposal, i.e. a sample over $W_*$ weighted by the public scores $\alpha_i$. For a women $w$ denote her sample weight by $\alpha(w)$ and the set of proposals she has received by $p(w)$. Further denote by $\Gamma_w = \sum_{m \in p(w)} \beta(m)$ the sum of the public scores of men who have proposed to $w$.

   Suppose that $|W_*| \geq n/2$, i.e. that $m_*$ has not yet proposed to over half the women. Using the assumption that the total number of proposals made is at most $O(n \ln n)$, we can bound the expected total weight of proposals women have seen by

$$\mathbb{E}_{w \sim \mathcal{W}_*} [\Gamma_w] = \frac{\sum_{w \in W_*} \alpha(w) \Gamma_w}{\sum_{w \in W_*} \alpha(w)} \leq \frac{\alpha_{\max} \sum_{w \in W} \Gamma_w}{|W_*| \alpha_{\min}} \leq \frac{\alpha_{\max} \beta_{\max} \cdot O(n \ln n)}{|W_*| \alpha_{\min}} \leq O(\ln n).$$

   Thus, by lemma 2.3, the probability that the proposal by $m_*$ will be accepted is

$$p_1 := \mathbb{E}_{w \sim \mathcal{W}_*} \left[ \frac{\beta_*}{\beta_* + \Gamma_w} \right] \geq \frac{\beta_*}{\beta_* + \mathbb{E}_{w \sim \mathcal{W}_*} [\Gamma_w]} \geq \Omega(1/\ln n).$$

   where the first inequality is due to Jensen's inequality.

2. If $m_*$ proposes to $w$ and is accepted, then the subsequent rejection chain can either end at the last woman without proposals, $w_{\text{last}}$, or cycles back to $w$ who this time rejects $m_*$. Notice that for each subsequent proposal, the ratio between the probability that it goes to $w_{\text{last}}$ (in which case the process will be terminated) and the probability that it returns to $w$ is at most $\alpha_{\max} : \alpha_{\min}$ (and possibly less if the proposing man has already proposed to $w$). Hence, the probability that the chain ends at the last women $w_{\text{last}}$ is bounded below by

$$p_2 := \frac{\alpha_{\min}}{\alpha_{\max} + \alpha_{\min}} \geq \Omega(1).$$

Note that this is ignoring the chance that a new proposal by $w$ is rejected, but it still suffices for a lower bound.

3. The probability that $m_*$ makes more than $K \ln^2 n$ proposals is thus bounded above by

$$(1 - p_1 p_2)^{K \ln^2 n} \leq \exp(-p_1 p_2 K \ln^2 n) = \exp(-\Omega(K \ln n)) \leq n^{-C}$$

as long as we choose $K = \Omega(C)$ large enough. ◄

▶ **Corollary 3.6.** *For any constant $C \geq 1$, with probability $1 - 1/n^C$, the maximum number of proposals made by any man in DA is $O(C \ln^2 n)$.*

**Proof.** By 3.4 and the upper bound for coupon collector (see the full version of this paper), the total number of proposals made in DA is $O(Cn \ln n)$ with probability $1 - 1/n^C$. In particular, if we consider any $m_*$ and let all other agents propose, this will be true. Recall that by lemma 2.1, DA is independent of the order in which men are chosen to propose. Thus, for each man $m_*$ we can apply lemma 3.5 to get that, with probability $1 - 1/n^{C+1}$, $m_*$ makes fewer than $O((C+1) \ln^2 n) = O(C \ln^2 n)$ proposals. Taking a union bound over the $n$ men gets the desired result. ◄

▶ Remark 3.7. Both of the above results hold for deferred acceptance with re-proposals as well as deferred acceptance. Indeed, even with re-proposals, deferred acceptance will be independent of the order of proposals (as re-proposals are ignored by the women). Moreover, the logic required to prove points 1. and 2. of the proof of lemma 3.5 is only easier to prove when men sample over all of $W$ as opposed to just the set $W_*$.

The above result is enough to show that proposition 3.2 holds for DA as well for the coupon collector, because repeated proposals are at most a $O(\ln^2 n/n) = o(1)$ fraction of total proposals in deferred acceptance with re-proposals. We defer the proof to the full version.

▶ **Theorem 3.8.** *Let $S$ be the total number of proposals made in DA with tiers of women $\boldsymbol{\epsilon}, \boldsymbol{\alpha}$, and arbitrary constant tiers on the men. We have*

$$\mathbb{E}[S] = \left(1 - O(\ln^2 n/n)\right) \mathbb{E}[T] = \left(1 \pm O(1/\ln n)\right) \frac{\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha}}{\alpha_{\min}} n \ln n.$$

## 4 Rank Achieved by the Men

Up until this point, our arguments have only crudely considered the preferences women have for men. Due to the asymmetry across the different tiers, this means we cannot yet calculate the expected rank men get.

Consider a man $m$ in tier $j$. Our main goal is to prove that the rank of $m$ is inversely proportional to $\beta_j$. As in 3.5, the core tool of our proof will be the fact that deferred acceptance is independent of execution order (by 2.1), and thus we can wait until all other men have finished proposing and found a match before letting $m$ propose. Once this is done, the major ideas are

1. Suppose $m$ has public score 1, and define

$$p = \mathop{\mathbb{E}}_{w \sim \mathcal{W}} \left[ \mathbb{P}\left[ w \text{ accepts a proposal from } m \right] \right].$$

   Note that, if $m$ were able to propose to a woman independently multiple times, the number of proposals until $m$ gets his first acceptance would be distributed exactly according to $\text{Geo}(p)$, and the expected value would be $1/p$. We show that (because men make much less than $n$ proposals) the difference due to re-proposals is not large.

2. Because $m$ is the last man to propose, most women have already seen many proposals and arrived at a decent match. When $m$ gets his first acceptance, he should thus be likely to stay where he is. We show that, while the probability of $m$ proposing to more women is non-negligible, it still contributes only $O(1)$ in expectation. So $m$'s expected rank is $1/p$ up to lower-order terms.

3. Another consequence of a woman $w$ receiving a large number of proposals is the following:

$$\mathbb{P}\left[ w \text{ accepts a proposal from } m' \text{ with weight } \beta \right]$$
$$\approx \beta \cdot \mathbb{P}\left[ w \text{ accepts a proposal from } m \text{ with weight } 1 \right].$$

   simply by 2.3 and the fact that $\beta/(\beta + \Gamma_w) \approx \beta \cdot 1/(1 + \Gamma_w)$ for $\Gamma_w$ (the sum of public scores of men who proposed to $w$) large. Thus, if $m$ had public score $\beta$, the effective value of $p$ would be approximately $\beta p$, and the expected rank of $m$ would become approximately $1/(\beta p)$. In other words, while we are not able to calculate $p$ directly, we show that $p$ scales properly with $m$'s score.

4. Finally, we prove that the above holds for most sequences of proposals of men before $m$, and thus holds in expectation over the entire execution of DA. Note that the distribution of proposals before $m$ changes slightly depending on which tier $m$ is chosen from, but in a large market, we do not expect this to make a big difference.

The biggest difference between the above proof sketch and its implementation is that we focus on *two* men proposing at the end of DA. This serves to address point 4 above – we are able to show that, for the vast majority of sequences of proposals before the last two men, their expected ranks are proportional to the ratio of their scores. Thus, this ratio holds in expectation over all of DA. Focusing on two men also allows us to bound the *correlation* between the two men's ranks, which is crucial for our concentration results.

In our proof, we also formalize what it means for all men other than two to propose, with the notion of a "partial matching state". Moreover, we give the term *smooth* to those states in which the proof sketch above goes through. Most crucially, in smooth matching states, "most women have received a lot of proposals", so that the reasoning in points 2 and 3 are valid. Additionally, to address certain technicalities (such as being able to bound the magnitude of the expected number of proposals) we define smooth matching states to not have too many proposals in total.

## 4.1   Smooth matching states

▶ **Definition 4.1.** *Given a set of men $L$, we define the* partial matching state excluding $L$, *denoted $\mu_{-L}$, as follows: Run DA with men in $M \setminus L$ proposing to $W$, and keep track of which proposals were made. More specifically, if $\mu$ is the (partial) matching resulting from running DA with a set of men $M \setminus L$ and set of women $W$, and $P = \{(m_{i_\ell}, w_{j_\ell})\}_\ell$ is the set of all tuples $(m_i, w_j)$ where $m_i$ proposed to $w_j$ during this process, then $\mu_{-L} = (\mu, P)$.*

*In a random matching market, we consider this state as a random variable. In a tiered random matching market, to specify this random variable, it suffices to give a multiset of tiers which the men in L belong to. For a fixed $\mu_{-L}$, denote by $\Gamma_w$ the total sum of weights which woman w received in P.*

Note that the state $\mu_{-L}$ keeps track of which proposals have been made (in addition to which current matches are formed) before the men in $L$ propose.

▶ **Definition 4.2.** *We call a partial matching state $\mu_{-L}$ smooth if the following hold for some constants $C_1, C_2, C_3 > 0$:*
1. *At most $C_1 n \ln n$ proposals were made to women overall.*
2. *At most $n^{1-C_2}$ women have received fewer than $C_3 \ln n$ proposals.*

The constants $C_1, C_2, C_3$ in the above depend on the tier structure, and can simply be chosen such that the following proposition holds. Our arguments will go through if smoothness holds with respect to any $C_1, C_2, C_3$ which are held constant as $n \to \infty$.

▶ **Proposition 4.3.** *Let $L = \{m_1, m_2\}$ be any pair of men. After running deferred acceptance, $\mu_{-L}$ is smooth with probability $1 - n^{-\Omega(1)}$.*

Once we know that $\mu_{-L}$ is smooth, our two main tasks are to show that men's ranks scale inverse-proportionally to their score, and that the ranks of different men do not correlate too highly. These are the main technical novelties of the paper. The exact details are given in the full version.

▶ **Proposition 4.4.** *Suppose $\mu_{-L}$ is smooth, and let $r_1$ and $r_2$ be the ranks of $m_1$ and $m_2$ after running DA with $m_1$ and $m_2$ starting from $\mu_{-L}$. We have*

$$\mathbb{E}_L[r_1] = \left(1 \pm O(1/\ln n)\right)\frac{\beta_2}{\beta_1}\mathbb{E}_L[r_2].$$

*where we use $\mathbb{E}_L[\,]$ to denote taking an expectation over the random process of $m_1, m_2$ proposing in DA after starting from state $\mu_{-L}$.*

▶ **Proposition 4.5.** *Suppose $\mu_{-L}$ is smooth, and let $r_1$ and $r_2$ be the ranks of $m_1$ and $m_2$ after running DA with $m_1$ and $m_2$ starting from $\mu_{-L}$. Then we have $\mathrm{Cov}(r_i, r_j) = O(\ln^{3/2} n)$.*

## 4.2   Expected rank of the men

In this subsection, we show that overall, expected rank scales proportionally to fitness (in addition to under smooth matching states). This allows us to compute the expected rank of the men. The proofs (deferred to the full version) follow by carefully keeping track of the (limited) effect of non-smooth matching states on the expectation.

▶ **Proposition 4.6.** *Let $r_i$ and $r_j$ denote the rank of a man in tiers $i$ and $j$. Then we have*

$$\mathbb{E}[r_i] = \left(1 \pm O(1/\ln n)\right)\frac{\beta_j}{\beta_i}\mathbb{E}[r_j].$$

▶ **Theorem 4.7.** *Let $\boldsymbol{\beta}^{-1}$ denote the vector $(1/\beta_i)_i$. For each tier $j$, the rank $r_j$ of men in tier $j$ has expectation*

$$\mathbb{E}[r_j] = \left(1 \pm O(1/\ln n)\right)\frac{\mathbb{E}[S]}{(n\boldsymbol{\delta}\cdot\boldsymbol{\beta})\beta_j} = \left(1 \pm O(1/\ln n)\right)\frac{\boldsymbol{\epsilon}\cdot\boldsymbol{\alpha}}{\alpha_{\min}}\cdot\frac{1}{(\boldsymbol{\delta}\cdot\boldsymbol{\beta}^{-1})}\cdot\frac{\ln n}{\beta_j}.$$

Finally, we also use our results on the covariance of men's ranks to prove concentration. We defer the proof to the full version. At a high level, the proof follows simply because the weak correlation implied by 4.5 means that the variance of the average of the ranks is lower-order (compared to its expectation), so Chebyshev's inequality can be used.

▶ **Theorem 4.8.** *For any tier $j$, let $\overline{R}_j^M = (\delta_j n)^{-1} \sum_m r_m$ denote the average rank of men in tier $j$. Then, for any $\epsilon > 0$,*

$$\overline{R}_j^M = (1 \pm \epsilon) \frac{\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha}}{\alpha_{\min}} \cdot \frac{1}{(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1})} \cdot \frac{\ln n}{\beta_j}$$

*with probability approaching 1 as $n \to \infty$.*

## 5   Expected rank of the women and the distribution of match types

### 5.1   Expected rank of women

We saw in Section 4.2 that men achieve ranks proportional to the inverse of their public scores. In this section, we turn to the women.

To study the rank the women achieve, we need to reason about the number of proposals women receive on average. By theorem 4.8, we expect that for each tier $j$ of men, the $\delta_j n$ men make a total number of proposals approximately

$$\frac{\delta_j \beta_j^{-1}}{\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1}} \cdot \frac{\boldsymbol{\alpha} \cdot \boldsymbol{\epsilon}}{\alpha_{\min}} n \ln n.$$

Each of these proposals goes to a woman in tier $i$ with probability $\pi_i = \alpha_i/(n\boldsymbol{\epsilon} \cdot \boldsymbol{\alpha})$, so we expect such a woman to receive approximately $(\delta_j \beta_j^{-1})/(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1}) \cdot (\alpha_i/\alpha_{\min}) \ln n$ proposals from men in tier $j$. Each of these men has public score $\beta_j$, so we expect $\Gamma_w$, the total sum of public scores of men proposing to $w$, to be roughly

$$\Gamma_w \approx \sum_j \beta_j \frac{\delta_j \beta_j^{-1}}{\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1}} \cdot \frac{\alpha_i}{\alpha_{\min}} \ln n = \frac{\alpha_i \ln n}{\alpha_{\min}(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1})}.$$

It is not immediately clear how the above value of $\Gamma_w$ should translate to the *rank* that $w$ gets. Unlike in the case where men are uniform, we cannot simply divide $n$ by the number of proposals which $w$ receives.

Indeed, suppose a woman $w$ receives exactly the total sum of weight $\Gamma_w$ predicted above. What should her rank be? This is essentially the following: across all tiers of $\delta_j n$ men each, how many do we expect to beat her best proposal so far? The probability that $w$ ranks a man $m$ higher than her match, when viewed according to 2.3, is a function only of the weight $\beta(m)$ of $m$ and the weight of proposals $\Gamma_w$ which $w$ received. Specifically, this probability is $\beta(m)/(\beta(m) + \Gamma_w) \approx \beta_j/\Gamma_w$. Summing this across all the men, we get

$$\mathbb{E}\left[r_w\right] \approx \sum_m \frac{\beta(m)}{\beta(m) + \Gamma_w} \approx \frac{n\boldsymbol{\delta} \cdot \boldsymbol{\beta}}{\Gamma_w} \approx (\boldsymbol{\delta} \cdot \boldsymbol{\beta})(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1})\frac{\alpha_{\min}}{\alpha_i} \cdot \frac{n}{\ln n}.$$

Note that this ignores the fact that a woman will never rank $m$ higher than her match if that $m$ already proposed to her during DA. But since $w$ only likely receives $\ln n \ll n/\ln n$ proposals, the difference is not noticeable.

It turns out that, with a detailed probabilistic analysis, the above proof sketch goes through. The details are given in the full version of this paper online.

▶ **Theorem 5.1.** *Let $\overline{R}_i^W = (\epsilon_i n)^{-1} \sum_{w \in T_i} r_w$ denote the average rank of women in tier $i$. For all $\epsilon > 0$, we have*

$$\overline{R}_i^W = (1 \pm \epsilon)(\boldsymbol{\delta} \cdot \boldsymbol{\beta})(\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1}) \frac{\alpha_{\min}}{\alpha_i} \frac{n}{\ln n}$$

*with probability approaching $1$ as $n \to \infty$.*

## 5.2 The distribution of match types

Fix a woman $w$ in tier $i$. We now study the probability that $w$ is matched to a man from some tier $j$. In the previous section, we argued that with high probability $w$ receives approximately a total of

$$\frac{\delta_j \beta_j^{-1}}{\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1}} \cdot \frac{\boldsymbol{\alpha} \cdot \boldsymbol{\epsilon}}{\alpha_{\min}} n \ln n$$

proposals from men in tier $j$. Thus, the contribution to $\Gamma_w$ (the total weight of proposals $w$ received) from men in tier $j$ is

$$\Gamma_{j \to w} \approx \frac{\delta_j}{\boldsymbol{\delta} \cdot \boldsymbol{\beta}^{-1}} \cdot \frac{\boldsymbol{\alpha} \cdot \boldsymbol{\epsilon}}{\alpha_{\min}} n \ln n \approx \delta_j \Gamma_w.$$

Moreover, it turns out that, with high probability, the above holds up to $(1 \pm \epsilon)$ for all tiers $j$ simultaneously. Regardless of the order in which $w$ saw proposals, the probability that $w$'s favorite proposal came from a man in tier $j$ is $\Gamma_{j \to w} / \Gamma_w$. Thus, this probability is approximately $\delta_j$. See the full version for a formal implement of the proof.

▶ **Theorem 5.2.** *Consider an arbitrary tier $i$ of women and $j$ of men. For all $\epsilon > 0$, there is an $n$ large enough such that the probability that a woman in tier $i$ matches to a man in tier $j$ is $(1 \pm \epsilon)\delta_j$.*

## 6 Summary

The model and findings in this paper contribute to the understanding of random stable matching markets. Indeed, the results quantify the effect of competition that arises from heterogeneous quality in agents, specifically, when the agents fall into different constant-factor tiers of quality. Novel technical tools are developed in order to reason about the proposal dynamics of deferred acceptance.

Relaxing some of the modeling assumptions raises interesting questions that cannot be trivially answered. This includes having non-constant (size, or public score) tiers, personalized private scores which give agents different distributions of preferences, and imbalance in the number of agents on each side of the market. Moreover, it is natural to ask when one should expect the matching to be sorted, i.e., higher tiers will be more likely to match with higher tiers (e.g., [12] demonstrates the presence of sorting in dating markets).

──── **References** ────

1    Itai Ashlagi, Mark Braverman, and Avinatan Hassidim. Stability in large matching markets with complementarities. *Operations Research*, 62(4):713–732, 2014.
2    Itai Ashlagi, Mark Braverman, Yash Kanoria, and Peng Shi. Communication requirements and informative signaling in matching markets. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, EC '17, page 263, New York, NY, USA, 2017. Association for Computing Machinery. `doi:10.1145/3033274.3084093`.

**3**    Itai Ashlagi, Yash Kanoria, and Jacob D Leshno. Unbalanced random matching markets: The stark effect of competition. *Journal of Political Economy*, 125(1):69–98, 2017.

**4**    Hedyeh Beyhaghi, Daniela Sabán, and Éva Tardos. Effect of selfish choices in deferred acceptance with short lists. *CoRR*, abs/1701.00849, 2017. `arXiv:1701.00849`.

**5**    Robert King Brayton. On the asymptotic behavior of the number of trials necessary to complete a set with random selection. *Journal of Mathematical Analysis and Applications*, 7(1):31–61, 1963.

**6**    Linda Cai and Clayton Thomas. The short-side advantage in random matching markets. *arXiv preprint arXiv:1910.04406*, 2019.

**7**    Peter Coles and Ran Shorrer. Optimal truncation in matching markets. *Games and Economic Behavior*, 87:591–615, 2014.

**8**    Aristides Doumas and Vassilis Papanicolaou. The coupon collector's problem revisited: Asymptotics of the variance. *Advances in Applied Probability - ADVAN APPL PROBAB*, 44, March 2012. `doi:10.1239/aap/1331216649`.

**9**    David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

**10**    Hugo Gimbert, Claire Mathieu, and Simon Mauras. Two-sided matching markets with correlated random preferences have few stable pairs. *arXiv preprint arXiv:1904.03890*, 2019.

**11**    Yannai A. Gonczarowski. Manipulation of stable matchings using minimal blacklists. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, page 449, New York, NY, USA, 2014. Association for Computing Machinery. `doi:10.1145/2600057.2602840`.

**12**    Gunter J Hitsch, Ali Hortaçsu, and Dan Ariely. Matching and sorting in online dating. *American Economic Review*, 100(1):130–63, 2010.

**13**    Nicole Immorlica and Mohammad Mahdian. Incentives in large random two-sided markets. *ACM Transactions on Economics and Computation (TEAC)*, 3(3):1–25, 2015.

**14**    Yash Kanoria, Seungki Min, and Pengyu Qian. Which random matching markets exhibit a stark effect of competition? *arXiv preprint arXiv:2006.14653*, 2020.

**15**    D. E. Knuth. *Mariages Stable*. Université de Montréal Press, 1976. Translated as "Stable Marriage and Its Relation to Other Combinatorial Problems, CRM Proceedings and Lecture Notes, 1997.

**16**    Donald E Knuth, Rajeev Motwani, and Boris Pittel. Stable husbands. *Random Structures & Algorithms*, 1(1):1–14, 1990.

**17**    Fuhito Kojima and Parag A Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–27, 2009.

**18**    SangMok Lee. Incentive Compatibility of Large Centralized Matching Markets. *The Review of Economic Studies*, 84(1):444–463, September 2016. `doi:10.1093/restud/rdw041`.

**19**    David G McVitie and Leslie B Wilson. Stable marriage assignment for unequal sets. *BIT Numerical Mathematics*, 10(3):295–309, 1970.

**20**    Boris Pittel. The average number of stable matchings. *SIAM Journal on Discrete Mathematics*, 2(4):530–549, 1989.

**21**    Boris Pittel. On likely solutions of a stable marriage problem. *The Annals of Applied Probability*, pages 358–401, 1992.

**22**    Boris Pittel. On likely solutions of the stable matching problem with unequal numbers of men and women. *Mathematics of Operations Research*, 44(1):122–146, 2019.

**23**    LB Wilson. An analysis of the stable marriage assignment algorithm. *BIT Numerical Mathematics*, 12(4):569–575, 1972.