Assigning more students to their top choices: A comparison of tie-breaking rules

Itai Ashlagi Afshin Nikzad Assaf Romm *

June 2015

Abstract

School districts that implement stable matchings face various decisions that impact assignments of students to schools. We study properties of the rank distribution of students with random preferences when schools use different tie-breaking rules to rank equivalent students. Under a single tie-breaking rule where all schools use the same ranking, a constant fraction of students are assigned to one of their top choices. By contrast, under a multiple tie-breaking rule where each school independently ranks students, a vanishing fraction of students are matched to one of their top choices. Even so, we show that by restricting the students to submitting relatively short preference lists under a multiple tie-breaking rule, a constant fraction of students will be matched to one of their top choices, while only a "small" fraction of students will remain unmatched.

Keywords. school choice; tie-breaking rule; deferred acceptance; stable matching

*Ashlagi: Sloan, MIT, iashlagi@mit.edu. Nikzad: MSE, Stanford , afshin.nikzad@gmail.com. Romm: Economics, Hebrew University of Jerusalem, assafr@gmail.com. We thank Sergiu Hart, Avinatan Hassidim, Déborah Marciano-Romm, Parag Pathak and Alvin Roth for helpful discussions. Ashlagi acknowledges the research support of National Science Foundation grant SES-1254768 and the support of BSF grant 2012344.

1 Introduction

Two-sided matching markets have been an exciting research area ever since they were introduced by Gale and Shapley (1962). In these markets, each agent has ordinal preferences over agents from the other side of the market. Despite the vast literature following Gale and Shapley's work, little is known about whom should agents should expect to be matched to. One exciting area, in which the rank distribution plays a crucial role is school choice, as school districts make policy decisions that have a direct impact on matchings.

A growing number of school districts, such as NYC and Boston (see Abdulkadiroğlu et al. (2009) Abdulkadiroğlu et al. (2005)), adopted centralized mechanisms to assign students based on Gale and Shapley's Deferred Acceptance algorithm, which selects a stable matching with respect to schools' and students' preferences over each other. Schools, however, are often not strategic and only use artificial preferences to break ties among equivalent students (see Abdulkadiroğlu and Sönmez (2003)). Another decision school districts make is how many schools students should rank. This paper provides some insights into the impact of such choices on the students' rank distribution.[1]

Two tie-breaking rules have been considered by school districts: single and multiple. Under the multiple tie-breaking rule (MTB), each school independently selects a random order over students for breaking ties, and under the single tie-breaking rule (STB) all schools use the same a priori order over students, which is selected uniformly at random. The choice of which tie-breaking rule to use under the deferred acceptance mechanism was raised by Abdulkadiroğlu and Sönmez (2003), who suggested that MTB may result in unnecessary inefficiency. Yet, school districts are often interested in the MTB rule, which naturally seems more equitable (for example, this tie-breaking rule is adopted in Amsterdam). Abdulkadiroğlu et al. (2009) empirically compared STB and MTB using NYC choice data and found that STB results in more students receiving top choices but at the same time more students receive "bad" choices or remain unassigned. In the present paper we take a first analytical step in explaining these qualitative insights.

Consider first whether more students receive their first choice under STB than under MTB. Intuitively, under STB the more schools a student is rejected from, the less likely she is to cause other students (who may be assigned to their top choices) to be rejected from

[1] See Ashlagi and Shi (2014), who study the design of school choice menus under public constraints.

other schools. This intuition suggests that indeed STB yields more top choices. However, we provide a counterexample of a market in which more students, in expectation, receive their top choice under MTB than under STB (Example 1).

This motivates the study of large two-sided matching markets, in which students have random preferences over schools. For simplicity, in our model all students belong to a single priority class, or have the equal right to access all schools. We study how the choice of the tie-breaking rule affects the rank distribution of students and ask how many students are likely to be assigned to one of their "top" choices under each tie-breaking rule. We further explore the effect of having short preference lists on the number of students who get assigned to their top choices and the number of students who remain unassigned.

Our Contributions. We first study markets in which students rank all schools, and each school has a constant capacity. We show that under the MTB rule students are *unlikely* to get one of their top choices. Formally, for any constant number k, with high probability a vanishing fraction of students are matched to one of their k most preferred schools as the market grows large. This is in sharp contrast to the STB rule, under which a constant fraction of students are matched to one of their top k preferred choices. While intuitively MTB may provide students more opportunities to get one of their top preferences, it actually creates harsher competition among students. To get some intuition that that the STB rule results in many students receiving top choices, observe that the deferred acceptance mechanism with STB and with one priority class is equivalent to the Random Serial Dictatorship, under which each student chooses a seat in a random order.

We next analyze the effect of shortening preference lists on assignments of students to schools Complementing the first result, we show that under the MTB rule it is possible to select the length of the preference lists such that not only a large fraction of students get one of their top choices, but also only a "small" fraction of students remain unmatched. One interesting case where shortening preference lists has a substantial impact on students' rankings is when there there is a shortage of seats (this is an interesting case as often some schools are known to be better than others). In the extreme case where there is a linear shortage of seats, we show that with high probability the rank distribution of students can be improved without increasing the number of unassigned students. Intuitively, under MTB, allowing long preference lists essentially creates a thick market, whereas shortening the lists reduces the competition and produces better outcomes for matched students. In Section 5 we provide a more detailed discussion of our results and a comparison to the empirical findings of Abdulkadiroğlu et al. (2009).

Our main results are derived for markets in which the capacity of each school remains constant as the market grows large. In Section 2.3, however, we discuss why in a market with a small number of schools, each with a large capacity, the rank distribution under STB will (almost) stochastically dominate the rank distribution under MTB. If there are enough seats overall for students, STB and MTB will generate very similar outcomes. Otherwise, students who are temporarily admitted to their first choice under the deferred acceptance algorithm, are less likely to be rejected under STB than under MTB, since other rejected students (from other schools), are are selected to have very bad lottery numbers.

From a technical point of view, one of the novelties of our proof method is the explicit use of (non trivial) capacities. Most papers that study random matching markets assume either explicitly or implicitly that there is a large imbalance in the market, which leads to an analysis similar to that of one-to-one matching markets.[2] Our method allows us to get a better grasp of the dependence of the parameters of interest on capacities and thus we are able to extend the arguments to non constant capacities. The main aim of our analysis is to simplify the random process that corresponds to the deferred acceptance algorithm by *coupling* it with a simpler random process.

1.1 Related Work

A number of papers have used randomly generated markets to study different properties of two-sided matching markets. A closely related work is that by Che and Tercieux (2014), who propose a deferred acceptance-like mechanism based on a "circuit breaker." Their algorithm does not allow students who are preferred by a certain school to push out other students who rank that school much higher than the preferred students, thus allowing to improve the rank distribution at the expense of few students' priorities. Their mechanism is also reminiscent of the Chinese Parallel, described by Chen and Kesten (2013).

[2]This is not to say that capacities do not significantly affect these papers' approach and analysis (see, for example, Kojima and Pathak (2009)).

Abdulkadiroğlu et al. (2009) raise the question of choosing between STB and MTB in the context of the NYC school choice system. Independently of this paper, Arnosti (2015) studies the impact of STB and MTB on the number of matches made when agents' preferences lists are short and random, and also finds that STB leads to many more students receiving their top choices than MTB does. When the mechanism of choice is the Top Trading Cycles mechanism, both Pathak and Sethuraman (2011) and Carroll (2014) extend the results of Abdulkadiroğlu and Sönmez (1998) and show that there is no difference between a single tie-breaking (i.e., using random serial dictatorship) and a multiple tie-breaking rule (TTC with random endowments). It is important to note that instances of indifference in schools' priorities over students have also yielded other very interesting approaches, among them are the stable improvement cycles of Erdil and Ergin (2008), the efficiency-adjusted DA of Kesten (2011), and the choice-augmented DA of **?**.

The effect of requiring students to submit short preferences has been described and studied by Haeringer and Klijn (2009), and experimentally tested by Calsamiglia et al. (2010). However, the idea of using truncation and dropping strategies as isn't novel; see for example Roth and Rothblum (1999) and Kojima and Pathak (2009).[3]

We are not the first to study the properties of the rank distribution in two-sided matching markets. Pittel (1989) studied the mens' average rank in a balanced stable marriage model with random preferences and showed that under men-optimal stable matching, the men's average rank is approximately $\ln n$ and every man is assigned to one of his top $ln^2 n$ preferred women with high probability. Ashlagi et al. (2013) showed that if there are fewer men than women, men are matched on average to at most their $\ln n$-th choice under any stable matching. Our main result, which shows that only a few students obtain one of their top preferences under MTB, is not implied by Ashlagi et al. (2013). Both of our results are robust to a small imbalance in the market.

2 The model and preliminary findings

In a school choice problem there are n students, each of whom can be assigned to one seat at one of m schools. We denote the set of students by $\mathcal{S}$ and the set of schools by $\mathcal{C}$, each with

[3]See also Coles and Shorrer (2014), who study truncation in random markets, and Gonczarowski (2014) for an algebraic approach.

some fixed capacity. Each student has a strict preference ranking over all schools. Denote the rank of a school c in the preference list of student s to be the number of schools that s weakly prefers to c. Thus the most preferred school for s has rank 1. Unless otherwise specified, students' preferences over schools are drawn independently and uniformly at random.

Each school $c \in \mathcal{C}$ has a priority ranking over students that is used to break ties between students in the same priority class. We assume each school has a single priority class containing all students, and thus each school uses a single ordering to break ties between students.

A *matching* of students to schools assigns each student to at most one seat and at most $\overline{q}$ students to each school. A matching is *unstable* if there is a student s and a school c such that s prefers to be assigned to c over his current assignment, and c either has a vacant seat or a student witofh lower priority than s. A matching is *stable* if it is not unstable.

Gale and Shapley (1962) showed that a stable matching always exists and, under strict preferences, there is one matching that is weakly preferred by all students, called the student-optimal stable matching. They further proposed the student-proposing deferred acceptance algorithm that computes this matching with respect to students' revealed preferences and school priority rankings.

2.1 Tie-breaking rules

We consider two common tie-breaking rules with which school districts determine priority rankings of schools over students. Under a *multiple tie-breaking rule* (MTB) each school independently selects a priority ranking over all students. Under a *single tie-breaking rule* (STB) all schools use the same priority order, which is selected uniformly at random a priori. One way of implementing STB is to assign to each student a lottery number drawn independently and uniformly at random from $[0, 1]$. Similarly, MTB can be implemented in a similar fashion by using a different draw (and thus different lottery numbers) for each school. MTB and STB naturally lead to different outcomes and we are interested in analyzing the properties of the rank distribution of students under these two tie-breaking rules.

2.2 No dominance in first choices

Consider a student who is rejected from some school under both MTB and STB. One would expect that this student would more likely be rejected from her next choice due under STB than under MTB owing to the information about her lottery number. This would suggest that, in expectation, more students are assigned to their top choice under the STB rule than under the MTB rule. We show here that with arbitrary preferences and capacities, somewhat surprisingly, this is not always the case.

Example 1. *Consider the following market. There are 5 schools with capacities* $q_1 = 40$*,* $q_2 = 10$*,* $q_3 = 500$*,* $q_4 = 5000$*, and* $q_5 = 20000$*. There are 4 types of students. There are 50 students of type 1 with preferences* $c_1 > c_2 > c_3 > c_4 > c_5$*, 10 students of type 2 with preferences* $c_2 > c_5$*, 500 students of type 3 with preferences* $c_3 > c_4 > c_5$*, and 5000 students of type 4 with preferences* $c_4 > c_5$*.*

Computer simulations show that expectation the fractions of students who obtain their first choice under STB and under MTB are 0.9951 *and* 0.9955*, respectively.*[4]

The main idea behind Example 1 is that under MTB rejected chains are shorter in expectation than under STB. To see this, first notice that after the first round of the student-proposing deferred acceptance algorithm, all students of types 2 and 3 and 4 are accepted to their first choice. Also observe that after the first round, 10 students of type 1 will be rejected from their first choice, c_1. These rejected students are less likely to be accepted to their second choice, c_2, under STB than under MTB. Furthermore, a student who is accepted to c_2 in the second round triggers a rejection chain that goes immediately to school c_5, which has enough capacity for all students. Finally, roughly speaking, a student of type 1 who is accepted to c_3 is likely to trigger a chain of length 2 till it reaches c_5.

Despite this example, we conjecture that STB leads to more students receiving their first choices in expectation under a general class of symmetric distributions, such as when preferences are all drawn from a symmetric logit choice model. This example motivates us to study large markets, in which each student has random preferences.

[4] Two different softwares on two different computers were used to obtain these exact results.

2.3 Examples with large capacities

In this section, we use a couple of simple examples to illustrate that the students' rank distribution under STB (almost) first-order stochastically dominates the rank distribution under MTB, when there is a small number of schools with large capacities.[5] In the remainder of the paper we will study students' rank distribution when schools have small capacities relative to the size of the market.

Consider a school choice problem with two schools, each of which has q seats. Also, suppose there are $n = 2q$ (or more) students. With uniform preferences, most students are matched to their first choice in the student-optimal stable matching and in fact the rank distribution under MTB and STB is identical. This holds because only one school is overdemanded (by approximately $\sqrt{q}$ students) and all rejected students are assigned to their second choice, which is necessarily underdemanded.

To make this example slightly more interesting, suppose now that there are only $0.8q$ seats in each of the two schools. (One may also think of a situation in which there are two schools that are more preferred by all students and students compete for seats in these two schools.) Roughly speaking, almost all admitted students obtain their first choice under STB. The reason is that rejected students are likely to have a very low lottery number and thus are very unlikely to cause the rejection of tentatively accepted students from their second choice (who are likely to have a high lottery number). Furthermore, even if a rejected student bumps out a tentatively accepted student, the latter will not cause further rejections. Under the MTB rule, however, a student rejected from her first choice gets a new lottery number for her second choice, and is therefore likely to bump out a tentatively accepted student, who in turn has a non-negligible probability (around 0.8) of bumping out another student and so on. This will cause a relatively short acceptance-rejection chain, of length which is roughly distributed geometrically with parameter 0.2.[6]

When capacities are small, obtaining such qualitative insights is more involved since the

[5] In general, when there is a (linearly) large surplus of seats, STB and MTB will generate very similar outcomes.

[6] It may be convenient for the reader to consider this example within the continuum framework of **?**. Under STB, both schools' cutoffs equal 0.2, and students either get their first choice, or remain unmatched. Under MTB, cutoffs equal $\sqrt{0.2} \approx 0.447$, meaning that out of the accepted students, 69% get their first choice and 31% get their second choice.

lottery numbers of rejected students do not provide much information. In what follows we analyze markets in which schools have smaller capacities.

3 Markets with small capacities: Average rankings and top choices

Motivated by Example 1 and the discussion about large capacities, in this section we study properties of the rank distribution in school-choice problems when there are many schools, each of which has a "small" capacity, and students have random preferences. For simplicity, each school has the same capacity $\overline{q}$ and we denote by $q = \overline{q}m$ the total number of seats. Section 3.1 provides properties of the rank distribution under MTB and Section 3.2 studies the rank distribution under STB.

Main Theorem. *Suppose $q \leq n + d$ for some constant d (either positive or negative). Then for any constant k, the expected fraction of students who are matched to their top k choice under the student-optimal stable matching approaches zero under MTB, but approaches a positive constant under STB.*

Before we proceed further, a few more words are in place regarding our strong assumption that preferences are drawn uniformly at random. Our analysis still holds when there is limited correlation in students' preferences, allowing for limited "degree of school popularity", and thus resulting in the same qualitative insights. When correlation is high, with the extreme that all students agree on the rankings, STB and MTB will lead to a similar rank distribution, as only few students are assigned to one their top choices (thus one can interpret our model as focusing on a sub market with schools of similar popularity).

3.1 Properties of the rank distribution under MTB

Ashlagi et al. (2013) show that there is a significant advantage to being on the short side in a two-sided matching market. In particular, students have a much better rank if there are sufficiently many seats than in the case where there is a shortage of seats. Denote by $Avg(\pi)$ the matched students' average rank of schools under the student-optimal stable matching when the tie-breaking rule π is used.

Proposition 3.1 (Ashlagi et al. (2013)). *Suppose $\overline{q} = 1$, $n = q + d$ for some (possibly negative) constant d, and fix any $\epsilon > 0$. If $d > 0$, the probability that $Avg(MTB)$ is at least $(1-\epsilon)\frac{n}{\ln n}$ converges to 1 as n grows large. If $d \leq 0$, the probability that $Avg(MTB)$ is at most $(1+\epsilon)\ln n$ converges to 1 as n grows large.*

When there is a shortage of seats, students are on average matched to bad choices and when then there is a surplus of seats students are on average much better off (and matched to good choices).

We next study the fraction of students who get one of their top choices. In the first main result, Theorem 3.2, we show that very few students receive one of their top choices under MTB as the market grows large even when students are on the short side of the market. Denote by $R_k(\pi)$ the expected fraction of students who get one of their top k choices under the student-optimal stable matching when the tie-breaking rule π is used. The combination of the results from these two section gives the following main theorem.

Theorem 3.2. *Suppose $q \leq n + d$ for some constant d (either positive or negative).*[7] *Then for any constant k, $R_k(MTB)$ approaches 0 as n approaches infinity.*[8]

We remark here that Theorem 3.2 still holds when every agent has a preference list of length $f(m)$ for any super-constant function f of m, i.e., a function that approaches infinity as m approaches infinity.

The proof, given in Section A, is based on the following steps. Fix some arbitrary school c. By symmetry and linearity of expectation observe that $R_k(MTB)$ equals the expected number of students who are admitted to c and c is among their top k preferences. Denote the latter amount by $\mathbb{E}[X_c]$. Next, we show that school c receives many more applications than the number of students who list c in one of their top k choices. Formally, c receives at least $\Omega(\ln n)$ proposals with high probability. Moreover, with high probability the number of students who list c in one of their top k choices, denoted by $\psi(k)$, is very

[7] In fact, our result also holds when there is a large sublinear surplus. In particular, it holds for $d = n^b$, where b is a constant less than 1. The same proof goes through, with minor modifications (merely different choices of constants).

[8] As long as the number of seats is larger than the number of students, the result holds for any stable matching. As the proof will show, as long as k is a constant, Theorem 3.2 holds when $\bar{q}$ grows more slowly than $\ln n$. If $\bar{q}$ is a constant, then the proof also holds if k grows more slowly than $\ln n$.

small (sub-logarithmic in n). Therefore, since each student who applies to c is admitted with probability at most $O(\overline{q}/\ln n)$, $\mathbb{E}[X_c] \leq O\left(\frac{\overline{q}\,\psi(k)}{\ln n}\right)$. This implies that $\mathbb{E}[X_c]$ approaches 0 as n approaches infinity.

In sharp contrast to MTB, we show that under STB at least half of the assigned students receive their first choice (Section 3.2). MTB fails to have this property because when a student gets rejected from multiple schools, she still has a "reasonable" chance of being accepted to a different school (where she has a new generated lottery number), which may trigger a rejection of some other student who may be assigned to one of her top choices.

3.2 Properties of the rank distribution under STB

In this subsection we discuss the properties of the rank distribution under STB. We provide two propositions, the proofs of which are deferred to Appendix E.

Proposition 3.3. *Under STB, the students' average rank of schools is* $O(\ln \min\{n, q\})$.

The intuition for this result follows from observing that the deferred acceptance algorithm under STB is equivalent to random serial dictatorship algorithm, which lets students pick in a random order their favorite school from among schools that still have vacancies. When it is the turn of the $(i+1)$-th student to choose a school, the rank she will obtain is, roughly speaking, a geometric random variable with a success probability of at least $\frac{i}{m-i}$ and summing up these random variables gives the result. Observe that, unlike in the MTB setting, Proposition 3.3 holds even if students are on the long side.

Since preferences are random, the intuition captured through the random serial dictatorship algorithm suggests that many students (especially among those who pick early in the random order) are likely to obtain their first choices.

Proposition 3.4. *Let* $t = \min\{n, q\}$ *be the total number of assigned students under STB. Then: (i) The expected number of students assigned to their first choice under STB is at least* $t/2$*. Moreover, (ii) the total number of students assigned to their first choice under STB is w.h.p.*[9] *at least* $(1-\varepsilon)(t/2)$ *for any* $\varepsilon > 0$.

[9] Given a sequence of events $\{E_n\}$, we say that this sequence occurs *with high probability (w.h.p.)* if $\lim_{n\to\infty} \frac{1-\mathbb{P}[E_n]}{n^{-\theta}} = 0$, for some constant $\theta > 0$.

It follows directly from Proposition 3.4 that in expectation at least half of the students receive their top choice when the number of students is not more than the number of seats.

4 Short preference lists

School districts often allow students to submit only a short preference list. Therefore since matched students are assigned to one of their listed choices, this can be a way to get around Theorem 3.2. Of course, shortening the lists also may increase the number of unassigned students. We analyze this tradeoff next.

Theorem 4.1. *Let U_k denote the number of unassigned students given that students submit only their k most preferred schools under the MTB rule.*

1. *If $n = (1+\varepsilon)q$ for some constant $\varepsilon \in \mathbb{R}_+$, then there exists a random variable Δ such that $U_k \leq \varepsilon q + \Delta$, where*
   - *(a) $\mathbb{E}[\Delta] \leq e^{-\varepsilon k} q$,*
   - *(b) Δ is not much larger than $\mu = e^{-\varepsilon k} q$, in the following sense (Chernoff bounds):*

$$\mathbb{P}[\Delta > \mu(1+\delta)] \leq e^{\frac{-\delta^2 \mu}{3}} \qquad \forall\, 0 < \delta < 1$$

$$\mathbb{P}[\Delta > \delta\mu] \leq \left(\frac{e^{\delta-1}}{\delta^\delta}\right)^{\mu} \qquad \forall\, \delta > 1$$

2. *If $n = q + d$ for some $d = o(q)$, then there exists a random variable Δ such that $U_k < d + \Delta$, where:*
   - *(a) $\mathbb{E}[\Delta] \leq 2q/k$ for all $k \leq m$.*
   - *(b) W.v.h.p. [10] $\Delta \leq (2+\varepsilon')q/k$, for all $\varepsilon, \varepsilon' > 0$ and $k \leq q^{1/3-\varepsilon}$.*

It is important to note here that when lists are short, deferred acceptance is not strategyproof. However, our model, in which preferences are drawn uniformly at random and each student's preference list is taken to be her private information, allows us to abstract away

[10] Given a sequence of events $\{E_n\}$, we say that this sequence occurs *with very high probability (w.v.h.p.)* if $\lim_{n\to\infty} \frac{1-\mathbb{P}[E_n]}{\exp\left\{-(\log n)^{0.4}\right\}} = 0$.

from strategic decisions as it is an equilibrium profile for each student to rank their top k choices.

The proof for Theorem 4.1 appears in Section A.2. When there are not enough seats, Ashlagi et al. (2013) prove that, on average, students are assigned to schools that rank low on their list. Even when there is a seat for every student, but not enough seats in "good schools," the same phenomenon occurs: students who are assigned to good schools are in fact assigned to schools that rank low on their list (compared to other good schools). Theorem 4.1 suggests a remedy by shortening the lists. Indeed, shortening the lists can significantly improve the average rank of assigned students, while ensuring that the number of assigned students does not decrease significantly or does not decrease at all, as shown in the following corollary.

Corollary 4.2. *Suppose $n = (1+\varepsilon)q$, and let $k = \frac{1+\delta}{\varepsilon} \cdot \ln q$, where ε, δ are positive constants. Then, w.h.p. $\Delta = 0$.*

Proof. By Theorem 4.1 we have $\mathbb{E}[\Delta] \leq e^{-(1+\delta)\ln q} q = q^{-\delta}$. So, $\mathbb{P}[\Delta > 1] \leq q^{-\delta}$. □

In Section 6 we use simulations to study settings with larger capacities (e.g., $\overline{q} = \sqrt{n}$) and observe similar trends. Finally, in Appendix A we provide a counterpart for Theorem 4.1 for the case where in which there is a surplus of seats, and we observe that very similar bounds obtain (Theorem A.6).

5 Discussion

5.1 STB vs. MTB

In a seminal paper comparing school choice mechanisms, Abdulkadiroğlu and Sönmez write:

> "Using a single tie-breaking lottery might be a better idea in school districts that adopt a Gale-Shapley student-optimal stable mechanism, since this practice eliminates part of the inefficiency: In this case, any inefficiency will be necessarily caused by a fundamental policy consideration and not by an unlucky lottery draw. In other words, the tie-breaking will not result in additional efficiency loss if it

is carried out through a single lottery (while that is likely to happen if the tie-breaking is independently carried out across different schools)." (Abdulkadiroğlu and Sönmez, 2003, Footnote 14).

This paper provides the first analytical comparison between STB and MTB in a model with random preferences.[11] Abdulkadiroğlu et al. (2009) show that (i) STB and MTB cannot be compared in terms of stochastic dominance, i.e., neither one stochastically dominates the other, and (ii) no strategy-proof and stable mechanism performs better than any specific tie-breaking rule. They further present empirical evidence suggesting that STB performs better at the higher end of the rank distribution, whereas MTB outperforms STB at the lower end. Our model provides support for the latter observation (Theorem 3.2). The intuition is roughly as follows. Consider what the probability is that student s who has been assigned to school c at the end of the first round of DA, will be rejected by the school in the second round. Since student s was assigned to her top choice under STB, we learn that she probably had a good lottery number, whereas future contenders for her seat who were rejected from several other schools probably had bad lottery numbers and are unlikely to cause s to be rejected. Under MTB, however, the fact that someone was rejected from a different school provides less information about her chances of taking s's place in school c.

A similar argument supports the frequently mentioned claim by school choice practitioners that MTB is in some sense "more fair" than STB. Indeed, consider a student who has been rejected from many schools. Under STB (but not under MTB) it is likely that she will also be rejected from the next school she applies to, because she probably has a low lottery number. When there are more seats than students, Pittel (1992) shows that every student gets assigned to at most her $O\left(\ln^2 n\right)$ rank, but under STB the last assigned student gets assigned to at least her $\frac{n}{2}$ rank. Computer simulations suggest that in general there is a tipping point in the rank distribution, such that STB dominates until that point (assigns more people to high-ranked schools), but MTB dominates afterwards (assigns fewer people to low-ranked schools). Formalizing and proving this claim is a very interesting direction for future research.

It is also interesting to gain insights into the rank distribution when students are on the long side of the market, as often some schools are better than other schools. In this

[11] Our analysis ignores multiple priority classes.

case our model suggests that STB significantly outperforms MTB both in the average rank of students (Theorem 3.1 and Proposition 3.3) and in assigning more students to their top choices (Theorem 3.2 and Proposition 3.4).

5.2 How many schools are students allowed to list?

School districts, such as Boston and NYC, often allow students to rank only a small number of schools. One concern is that many students will remain unassigned, which will result in an excessive administrative burden.[12] Theorem 4.1 provides a rationale for shortening lists in addition to some guidance for practitioners. It shows that by shortening lists (appropriately) under MTB, the social planner can bound the fraction of people who get assigned to a school they do not like, whereas allowing long lists may leave only a very small number of people with their top choices (this effect is even stronger when the students are on the long side).[13] This observation also resonates with the logic behind the "circuit-breaker" mechanism of Che and Tercieux (2014). Indeed, setting a small number of choices and then later administratively assigning unmatched students can be viewed as a rudimentary way to implement a two-stage mechanism.

6 Simulations

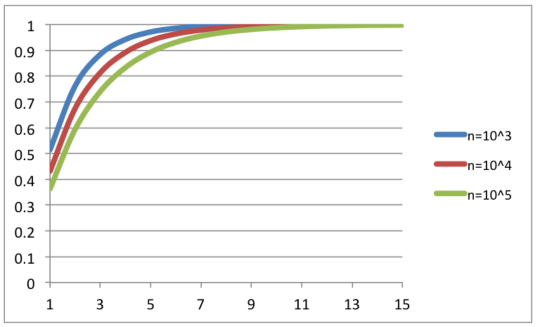
We present several simulation results in order to convey some of their implications for the rank distribution. All figures show the cumulative rank distribution function that students receive under different mechanisms.

The left panel in Figure 1 shows that as the number of students (and seats) grows large, the percentage of students who get their top choices under MTB becomes smaller. The fact that the difference between the percentage of those students, when $n = 10^4$ and $n = 10^3$, is roughly similar to the difference between $n = 10^5$ and $n = 10^4$ follows from the logarithmic decay rate proved in Theorem 3.2. The right panel in Figure 1, where $\overline{q} = 1$, shows more clearly the same intuition for positions 5 to 20. In this panel it is also more visible that as the

[12] Abdulkadiroğlu et al. (2009) provide empirical evidence that many students in NYC remain unmatched after the first round.

[13] We do not formally compare the number of unassigned people under both STB and MTB; however, we believe that MTB actually performs better in the sense of leaving fewer people unassigned.

number of students increases, the rank distribution becomes more flat. Figure 2 represents a similar plot when there is a surplus of seats. Observe that the percentage of students assigned to their top choices still becomes smaller as the number of students grows; this percentage in fact converges to 0 by Theorem 3.2.



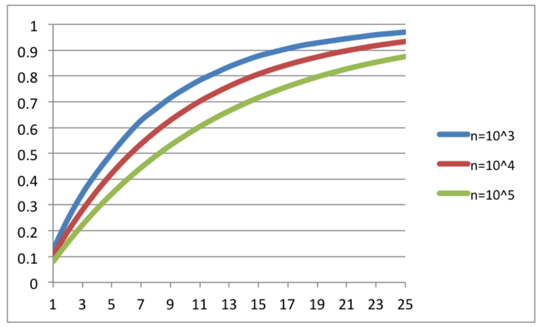


Figure 1: Rank distribution under MTB with $\overline{q} = 10$ (left) and $\overline{q} = 1$ (right)

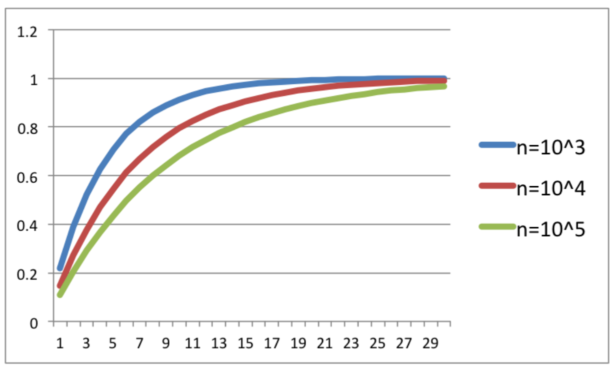


Figure 2: Rank distribution under MTB with $\overline{q} = 1$ and $m = n + 10$ (i.e., when there is a surplus of seats)

When we look at the corresponding graphs for STB (Figure 3), the situation is completely different. In this case, all three lines are on top of each other, with the rank distribution barely changing as n increases. Under this representation, the area of the region enclosed by the cumulative rank distribution function and the line $y = 1$ represents the expected rank (where y denotes the vertical axis). Observe the significant difference between STB and MTB in terms of the expected rank.

The next set of figures relates to the effect of allowing students to submit only a limited number of schools under MTB. Figure 4 shows the effect of submitting short lists when the

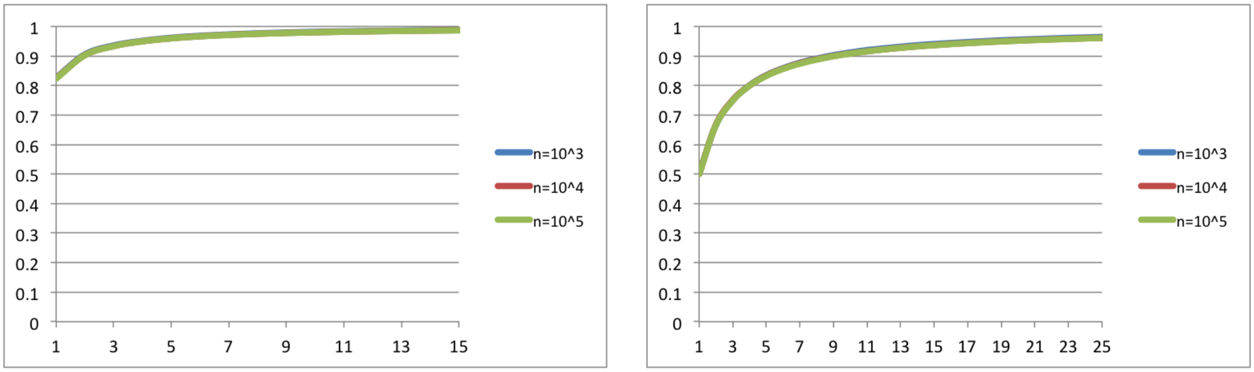


Figure 3: Rank distribution under STB with $\bar{q} = 10$ (left) and $\bar{q} = 1$ (right)

market is balanced. Each of the three panels corresponds to one of the lines in the left panel of Figure 1, where the brighter-colored line is exactly the same as the one in Figure 1, and the darker-colored line represents the rank distribution when students submit only their top 5 schools. The simulations show that the darker-colored lines do not go down as market size increases, and that the percentage of unassigned students does not increase either, as predicted in Theorem 4.1.

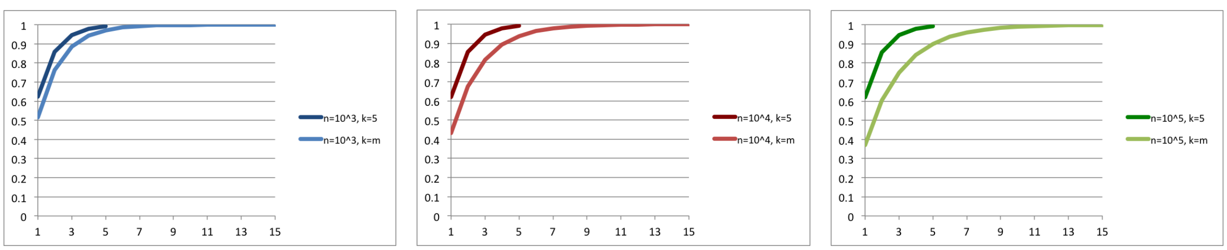


Figure 4: The effect of submitting short lists (balanced markets); $\bar{q} = 10$

Figure 5 illustrates what happens in an unbalanced market, when the students submit lists of different lengths. The left panel shows that as the length of preference lists increases, the percentage of unassigned students decreases, but this is accompanied by an increase in the expected rank of schools that students receive. The right panel shows this effect in full, when it compares the case of submitting only 20 schools to the case of submitting a list containing all the schools ($k = 4000$).

Figure 6 plots the effect of shortening preference lists when the market is unbalanced and the capacities are relatively large. In particular, $\bar{q} = 100$, $m = 100$, and $n = 100^2 + 100$. Note that when students submit their full preferences, the number of people getting their

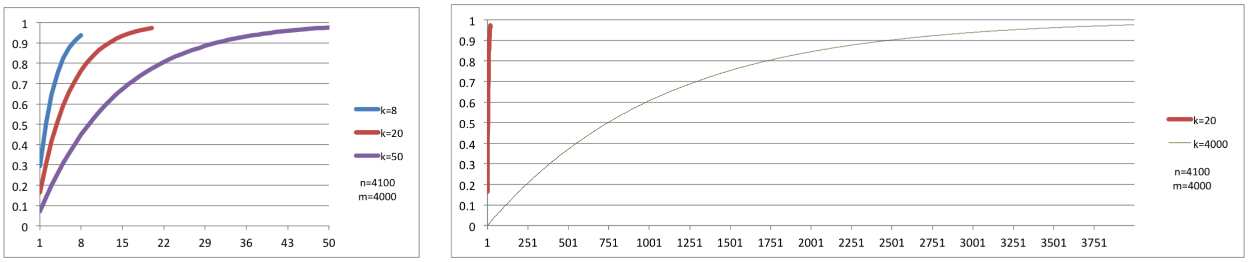


Figure 5: The effect of submitting short lists (unbalanced markets); $\overline{q} = 1$

first choice is very small (recall that under STB at least half of the population gets their first choice). However, as lists become shorter, the rank distribution becomes better, while the number of unassigned students does not increase notably.

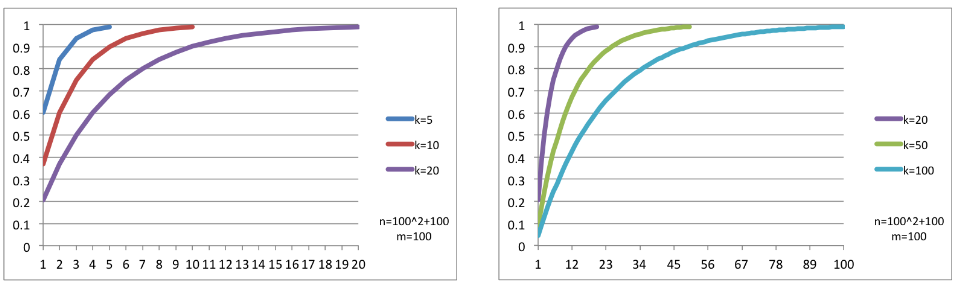


Figure 6: The effect of submitting short lists (unbalanced markets); $\overline{q} = 100$

7 Conclusion

In school choice environments, in which students are often coarsely grouped into priority classes, choosing a tie-breaking rule can have a significant effect on the resulting rank distribution. Information about this distribution is not only an important way to guide school choice design, but also beneficial for students who need to go through the arguably costly task of ranking schools.

Our results indicate that pairing the deferred acceptance mechanism with the MTB rule leads to far fewer students getting one of their top choices than pairing it with the STB rule. Interestingly, shortening preference lists appropriately can overcome the drawbacks of MTB, so that many students receive top choices while few remain unmatched.

This direction of research is important for other two-sided matching markets as well. For instance, providing better information to residents or hospitals regarding the rank distribution resulting in the NRMP can significantly impact and possibly optimize their interviewing process.

This paper contributes to a growing line of research on two-sided matching markets with random preferences that aims to understand outcomes in typical markets. While qualitative results provide useful insights, quantitative results should be taken with caution. For simplicity our model assumes that preferences are drawn independently and uniformly at random. This allows us to bypass technical issues, but also to abstract away from strategic issues when analyzing short preference lists. For example, when students' preferences are correlated one should analyze equilibrium outcomes as students may submit false preferences when their lists are short. Nevertheless, our results can be extended to also allow random preferences in which no school is significantly "better" than any other school (see, e.g., Immorlica and Mahdian (2005) and Kojima and Pathak (2009)).

Finally, we raise here two questions. First, we conjecture that STB second-order stochastically dominates MTB.[14] Second, we believe that if students' preferences are independently drawn from a symmetric logit choice model, STB maximizes the number of students who get their first choice.

References

Atila Abdulkadiroğlu and Tayfun Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–701, 1998.

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003. ISSN 00028282.

Atila Abdulkadiroğlu, Parag A Pathak, Alvin E. Roth, and Tayfun Sönmez. The boston public school match. *American Economic Review*, 95(2):368–371, 2005.

Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth. Strategy-proofness versus efficiency in matching with indifferences: Redesigning the nyc high school match. *American Economic Review*, 99(5):1954–1978, 2009.

[14]Second-order stochastic dominance is also known as dominance in increasing concave stochastic order.

Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda. Expanding choice in school choice. *American Economic Journal: Microeconomics*, 7(1):1–42, 2015.

Nick Arnosti. Short lists in centralized clearinghouses. *Available at SSRN*, 2015.

Itai Ashlagi and Peng Shi. Optimal allocation without money: an engineering approach. *Management Science*, 2015. Forthcoming.

Itai Ashlagi, Yashodhan Kanoria, and Jacob D Leshno. Unbalanced random matching markets. In *ACM Conference on Electronic Commerce*, pages 27–28, 2013.

Eduardo M Azevedo and Jacob D Leshno. A supply and demand framework for two-sided matching markets. *Available at SSRN 2260567*, 2014.

Caterina Calsamiglia, Guillaume Haeringer, and Flip Klijn. Constrained school choice: An experimental study. *American Economic Review*, 100(4):1860–1874, 2010.

Gabriel Carroll. A general equivalence theorem for allocation of indivisible objects. *Journal of Mathematical Economics*, 51:163–177, 2014.

Yeon-Koo Che and Olivier Tercieux. Efficiency and stability in large matching markets. Mimeo, 2014.

Yan Chen and Onur Kesten. From boston to chinese parallel to deferred acceptance: Theory and experiments on a family of school choice mechanisms. Discussion Paper, Social Science Research Center Berlin (WZB), Research Area Markets and Politics, Research Unit Market Behavior, 2013.

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

Aytek Erdil and Haluk Ergin. What's the matter with tie-breaking? improving efficiency in school choice. *American Economic Review*, 98(3):669–689, 2008.

D. Gale and L. S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15, 1962. ISSN 00029890.

Yannai A Gonczarowski. Manipulation of stable matchings using minimal blacklists. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 449–449. ACM, 2014.

Guillaume Haeringer and Flip Klijn. Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947, 2009.

Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 53–62. Society for Industrial and Applied Mathematics, 2005.

Onur Kesten. On two kinds of manipulation for school choice problems. *Economic Theory*, pages 1–17, 2011. ISSN 0938-2259. 10.1007/s00199-011-0618-6.

Fuhito Kojima and Parag A. Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–27, 2009. doi: 10.1257/aer.99.3.608.

Parag A Pathak and Jay Sethuraman. Lotteries in student assignment: An equivalence result. *Theoretical Economics*, 6(1):1–17, 2011.

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

Boris Pittel. On likely solutions of a stable marriage problem. *The Annals of Applied Probability*, 2(2):358–401, 1992.

Alvin E. Roth and Uriel G. Rothblum. Truncation strategies in matching markets-in search of advice for participants. *Econometrica*, 67(1):21–43, 1999. ISSN 00129682.

A Analysis

We fix some notation and then proceed to the proof. We use a version of the deferred acceptance algorithm that accepts an infinite stream of integers, representing a source of randomness from which students' preferences are drawn. We use square brackets to denote a range, i.e., $[n] = \{1, \dots, n\}$. Let S be a sequence of integers in $[m]$, such that each integer in

$[m]$ appears an infinite number of times in S. Let $S[j]$ denote the prefix of S up to the j-th place, and let S_h denote the h-th integer in S. The student-proposing deferred acceptance algorithm with k rounds based on S (denoted by $\mathrm{DA}\,(k)$) works as described in Algorithm 1 by letting all proposers in a specific round draw schools one by one, while skipping any school they already proposed to. Note that when a prefix of size j is read, it is possible that fewer than j proposals were made, because some integers were read by students who already proposed to those schools. The regular deferred acceptance algorithm (without limit on the number of rounds) is denoted by DA ($= \mathrm{DA}\,(\infty)$).

We almost always omit the dependency of $\mathrm{DA}\,(k)$ and DA on the schools' preferences, assuming they are drawn at random as described above. Whenever we refer to $\mathrm{DA}\,(k)$ or DA without specifying a stream S, it means that we refer to the operation of the deferred acceptance algorithm on a random stream in which every integer is drawn uniformly at random from $[m]$. At the same time we ignore the (measure 0) event of having a stream in which some integer appears only finitely many times. Some of our results are related to the expectation of the rank distribution and others are concentration results. For the latter we often use the expression *with high probability* (also denoted by w.h.p.) to mean with probability at least $1 - n^{\lambda}$, and *with very high probability* (w.v.h.p.) to mean with probability at least $1 - e^{-n^{\lambda}}$, for some $\lambda > 0$.

Finally, we say that a Probability Mass Function (PMF) $P : \mathbb{Z}_+ \to \mathbb{R}_+$ stochastically dominates PMF $Q : \mathbb{Z}_+ \to \mathbb{R}_+$ if for any $i \in \mathbb{Z}_+$ we have: $\sum_{j=0}^{i} P(j) \geq \sum_{j=0}^{i} Q(j)$.

A.1 Proof of Theorem 3.2

Let us first provide a proof sketch.

Proof Sketch. For each school c, we define a random variable X_c that takes values in $\{0, 1, \ldots, \overline{q}\}$ and denotes the number of students who are admitted to school c and c is among their top k choices. Note that $R_k = \mathbb{E}\left[\sum_{c \in \mathcal{C}} X_c\right]/n$. By linearity of expectation and symmetry, we have $R_k = \mathbb{E}\left[X_c\right]$ for any fixed $c \in \mathcal{C}$. So the theorem is proved if we show that $\mathbb{E}\left[X_c\right]$ approaches 0 as n approaches infinity.

To prove the latter fact, first we show that school c receives at least $\Omega(\ln n)$ proposals with high probability. Second, we show that with high probability, the number of students who have listed c in one of their top k positions, denoted by $\psi(k)$, is sub-logarithmic in n.

Algorithm 1: DA (k) based on S

```
Input: k, S, schools' preferences
Output: Matching μ
P ← [n]
h ← 0
∀c ∈ C : μ[c] ← ∅
for round ← 1 to k do
    for s in P do
        if s already proposed to all schools then
            continue
        end
        while s already proposed to S_h do
            h ← h + 1
        end
        μ[S_h] ← μ[S_h] ∪ {s}
        h ← h + 1
    end
    P ← ∅
    for c in C do
        T ← top q̄ students in μ[c] according to c's preference
        P ← P ∪ (μ[c] \ T)
        μ[c] ← T
    end
    sort P
end
```

These two facts imply $\mathbb{E}[X_c] \leq O\left(\frac{\overline{q}\,\psi(k)}{\ln n}\right)$, because each of the students who applied to c and have c in one of their top k positions would be admitted with probability at most $O(\overline{q}/\ln n)$. So, each of them contributes at most $O(\overline{q}/\ln n)$ to $\mathbb{E}[X_c]$. Since there are at most $\psi(k)$ such students, it follows that $\mathbb{E}[X_c] \leq O\left(\frac{\overline{q}\,\psi(k)}{\ln n}\right)$. This proves the promised claim: $\mathbb{E}[X_c]$ approaches 0 as n approaches infinity. □

Remark 1. *Given that k is a constant, Theorem 3.2 still holds when $\overline{q} = o(\ln n)$. In fact, the same proof works; here, we verify this by following the same proof sketch. Note that $\mathbb{E}[X_c] \approx \frac{\overline{q}\,\psi(k)}{\ln(n/\overline{q})}$. So, $\mathbb{E}[X_c]$ approaches 0 when $\overline{q} = o(\ln(n/\overline{q}))$, or equivalently, when $\overline{q} = o(\ln n)$.*

Remark 2. *Given that $\overline{q}$ is a constant, Theorem 3.2 holds for any $k = o(\ln n)$. To see why, it is enough to follow the proof of Theorem 3.2 and note that when $k = o(\ln n)$, the right-hand side of concentration bound (1) still approaches 0 as n approaches infinity, for a suitable choice of $\theta > 1$.*

Before proceeding to the proof, we introduce two parameters, r and δ, that are frequently used in the analysis. During the analysis, we typically find it helpful to run DA for only r rounds. We define $r = 4m^{1/2}/\overline{q}$. Another frequently used parameter in our analysis is δ, which is set to $3/4$ in this proof.

Next, we state two lemmas that are required to prove Theorem 3.2. In the first lemma, we show that with high probability, most of the schools are not empty by the end of DA (r).

Lemma A.1. *The probability of having more than m^δ empty schools by the end of DA (r) is at most $r \cdot e^{-\frac{\overline{q}m^{2\delta-1}}{16}}$*

The proof for Lemma A.1 is deferred to Section B. Next, we show that w.h.p. the total number of proposals sent in DA (r) is at least $\Omega(m \ln m)$; this is a consequence of the next lemma.

Lemma A.2. *Let $l = m(\ln m - t)$ for some $t > 0$. Then, with probability at least $1 - \frac{m^\delta+1}{e^t} - r \cdot e^{-\frac{\overline{q}m^{2\delta-1}}{16}}$, at least l proposals are sent in DA (r).*

A suitable choice of t, e.g. $t = 4/5 \ln m$, implies that w.h.p. the total number of proposals sent in DA (r) is at least $\Omega(m \ln m)$. Lemma A.2 is proved in Section B.

We are now ready to prove Theorem 3.2.

Proof of Theorem 3.2. We follow the proof sketch. Recall that we define a random variable X_c for each school c; X_c is the number of students who are admitted to c and c is among their top k choices. Notice that $R_k = \mathbb{E}\left[\sum_{c\in\mathcal{C}} X_c\right]/n$. By linearity of expectation and symmetry, we have $R_k = \mathbb{E}[X_c]$ for any fixed $c \in \mathcal{C}$. So the theorem is proved if we show that $\mathbb{E}[X_c]$ approaches 0 as n approaches infinity.

We now formally prove the above fact in three steps. In Step 1, we show that school c receives at least $\Omega(\ln n)$ proposals with high probability. In Step 2, we show that with high probability, the number of students who have listed c in one of their top k positions is a sub-logarithmic function of m. Step 3 simply uses what we proved in the first two steps to complete the proof.

Step 1 We use Lemma A.2 with $t = 4/5 \ln m$. This implies DA (r) makes at least $m \ln m/5$ proposals with probability at least

$$1 - \frac{m^\delta + 1}{e^t} - r \cdot e^{-\frac{\overline{q}m^{2\delta-1}}{16}},$$

which is at least $1 - 2m^{-1/20}$ for large enough m.[15] Now, given the school c, we want to show that this school receives at least $O(\ln m)$ proposals. We just showed that at least $m \ln m/5$ proposals will be sent with high probability. Thus, DA (r) reads a prefix of S with length at least $j = m \ln m/5$. To complete Step 1, we show that w.h.p. school c appears at least $\Omega(\ln m)$ times in $S[j]$. Using this fact we then show that at least $\Omega(\ln m)$ different schools propose to c.

Let $E(c)$ be the event in which c appears at least $\ln m/10$ times in $S[j]$. To complete Step 1, we first prove that $E(c)$ holds with high probability; this is done by a standard application of Chernoff bounds. For each index h, let Y_h be a binary random variable that is 1 iff $S_h = c$. Let $\mu = \mathbb{E}\left[\sum_{h=1}^{j} Y_h\right]$; note that since Y_h is a Bernoulli random variable with mean $1/m$, we have $\mu = \ln m/5$. By Chernoff bounds we have

$$\mathbb{P}\left[\sum_{h=1}^{j} Y_h < \mu(1-\epsilon)\right] \le e^{\frac{-\epsilon^2\mu}{2}} = m^{-1/40},$$

for $\epsilon = 1/2$. So, the probability that c appears less than $\ln m/10$ times in $S[j]$ is at most $m^{-1/40}$.

[15] E.g. $m > 5$ works if $\overline{q} \ge 16$; $m > 10^5$ suffices for $\overline{q} = 1$. We have not tried to optimize this constant.

Proposition A.3. *For any school c, $E(c)$ holds with probability at least* $1 - m^{-1/40}$.

Assuming that c appears at least $\ln m/10$ times in $S[j]$, we show that $\Omega(\ln m)$ of these appearances correspond to proposals made by disjoint students. This will complete the proof of Step 1. To this end, define E to be the event in which each student makes at most 4 (possibly redundant) proposals in any round of DA$\,(r)$. We show that E holds with high probability. See that by Lemma B.1, the probability that each student makes more than 4 offers in each round is at most $(r/m)^4 = (4/\overline{q})^4 m^{-2}$. A union bound over all students and all rounds implies that E holds with probability at least $1 - 4(4/\overline{q})^4 m^{-1/2}$.

Proposition A.4. *E holds with probability at least* $1 - 4(4/\overline{q})^4 m^{-1/2}$.

By Propositions A.3 and A.4, $E(c) \wedge E$ holds with probability at least $1 - m^{-1/40} - 4(4/\overline{q})^4 m^{-1/2}$. To complete the proof of Step 1, just note that when $E(c) \wedge E$ holds, c must have received proposals from at least $\ln m/40$ disjoint students.

Step 2 In this step, we will show that with high probability, the number of students who have listed c in one of their top k positions is a sub-logarithmic function of m. For any student s, let Z_s be a binary random variable that is 1 iff s lists the school c on one of her top k positions. Let $\mu = \mathbb{E}\left[\sum_{s\in\mathcal{S}} Z_s\right]$. Note that since the preferences are uniform, it follows that $\mathbb{P}\left[Z_s = 1\right] = k/n$, which means $\mu = k$. We prove that with high probability, $\sum_{s\in\mathcal{S}} Z_s$ is not much larger than its mean. This is done by applying the following version of Chernoff bounds:

$$\mathbb{P}\left[\sum_{s\in\mathcal{S}} Z_s > \theta\mu\right] < \left(\frac{e^{\theta-1}}{\theta^{\theta}}\right)^{\mu}, \tag{1}$$

which holds for any $\theta > 1$. Let the right-hand side of (1) be denoted by $f(\theta)$ and observe that by setting $\theta = \sqrt{\ln m}$, $f(\theta)$ approaches 0 as n approaches infinity.[16]

Step 3 Let E^* be the event in which school c receives at least $\ln m/40$ proposals during DA$\,(r)$ and at most θk students list c among their top k choices. As a consequence of Steps 1 and 2, we know that E^* holds w.h.p., i.e., with probability at least $1 - m^{-1/40} -$

[16]In fact, the proof works for any non-constant function that grows slower than $\ln m$, i.e. $\theta = o(\ln m)$.

$4(4/\overline{q})^4 m^{-1/2} - f(\theta)$. Notice that

$$\mathbb{E}\left[X_c | E^*\right] \leq \theta k \cdot \frac{\overline{q}}{\ln m/40} = 40k\overline{q}/\sqrt{\ln m}.$$

Using this, we can write

$$\begin{aligned}\mathbb{E}\left[X_c\right] \leq &\left(1 - m^{-1/40} - 4(4/\overline{q})^4 m^{-1/2} - f(\theta)\right) \cdot \left(40k\overline{q}/\sqrt{\ln m}\right) \\ &+ \left(m^{-1/40} + 4(4/\overline{q})^4 m^{-1/2} + f(\theta)\right) \cdot k. \qquad (2)\end{aligned}$$

Now, observe that the right-hand side of (2) approaches 0 as m approaches infinity. This completes the proof. □

A.2 Proof of Theorem 4.1

Before proving Theorem 4.1, we extend the definition of DA (k) by specifying which "seat" each student takes if accepted to a school. Label each seat with a unique label, and let $\mathcal{L}$ denote the set of labels of all q available seats. In each round of DA (k), when a student s proposes to a school that has empty seats, one of the empty seats is chosen uniformly at random (among all the empty seats in that school) and is assigned to that student. This is the only change that we make to DA (k); as this is not a structural change, we still denote this process by DA (k).

Proof for Part (1) of Theorem 4.1. The main idea behind the proof to define a much simpler process, DA$''$ (k), which, roughly speaking, leaves more students unassigned at the end of round k. We next analyze DA$''$ (k) and use the number of unassigned students in it as an upper bound on the number of unassigned students in DA (k). More precisely, suppose that $P : \mathbb{Z}_+ \to \mathbb{R}_+$ and $P'' : \mathbb{Z}_+ \to \mathbb{R}_+$ respectively denote the PMFs of *the number of unassigned students* in DA (k) and DA$''$ (k). Then, we show that P stochastically dominates P''. Because of this stochastic domination, we only need to prove the bounds (stated in the theorem statement) for DA$''$ (k); the identical bounds will hold for DA (k).

We define DA$''$ (k) in two steps. In Step 1, we convert DA (k) to DA$'$ (k) (a random process slightly more complicated than DA$''$ (k)). This conversion is done so that P stochastically dominates P', the PMF of the number of unassigned students in DA$'$ (k). In Step 2, we convert DA$'$ (k) to DA$''$ (k) while ensuring that P' stochastically dominates P''.

We start by defining $\text{DA}'(k)$: This process involves k rounds. In each round, εq new students show up and, one by one, they propose to a school picked uniformly at random. When a student s proposes to a school that as at least one empty seat, then one of the empty seats in that school is chosen uniformly at random and is assigned to s. Otherwise, if the school is full, s is rejected in that round. Lemma C.1 shows that P' is stochastically dominated by P.

Next, we convert $\text{DA}'(k)$ to $\text{DA}''(k)$ while ensuring that P'' is stochastically dominated by P'. We start by defining $\text{DA}''(k)$: this process involves k rounds. In each round, εq new students show up and, one by one, they propose to a school picked uniformly at random. When a student s proposes to a school, she picks one of the seats uniformly at random among all the available $\overline{q}$ seats; her proposal in that round is accepted iff the seat is empty. Lemma C.2 shows that P'' is stochastically dominated by P'.

To prove the theorem, we first prove the bounds given in the theorem statement for $\text{DA}''(k)$ (instead of $\text{DA}(k)$). Stochastic dominance then implies that the bounds hold for $\text{DA}(k)$ as well. More formally, suppose that U''_k denotes the number of unassigned students at the end of $\text{DA}''(k)$. We then show that there exists a random variable Δ'' such that $U''_k \leq \varepsilon q + \Delta''$, where $\mathbb{E}[\Delta''] \leq e^{-\varepsilon k} q$ and, moreover, w.h.p. Δ'' is not much larger than $\mu = e^{-\varepsilon k} q$, in the following sense (Chernoff concentration bounds):

$$\mathbb{P}[\Delta'' > \mu(1+\delta)] \leq e^{\frac{-\delta^2 \mu}{3}} \qquad \forall\, 0 < \delta < 1$$

$$\mathbb{P}[\Delta'' > \delta\mu] \leq \left(\frac{e^{\delta-1}}{\delta^\delta}\right)^\mu \qquad \forall\, \delta > 1$$

This will prove the theorem.

Given the above definition for $\text{DA}''(k)$, it is straightforward to verify that in each round, each seat will receive a proposal with probability $\frac{1}{\overline{q}m} = 1/q$. We use this fact to prove the theorem. For any seat $l \in \mathcal{L}$, let X_l be a binary random variable that is 1 iff seat l is empty at the end of $\text{DA}''(k)$. By considering applications sent out by the $\varepsilon q k$ students (arriving over k rounds) we have:

$$\mathbb{E}[X_l] \leq \left[\left(1 - \frac{1}{q}\right)^{\varepsilon q}\right]^k \leq e^{-\varepsilon k}. \tag{3}$$

Let $\Delta'' = \sum_{l \in \mathcal{L}} X_l$; since $\mathbb{E}[\Delta''] = \sum_{l \in \mathcal{L}} \mathbb{E}[X_l]$, we have $\mathbb{E}[\Delta''] \leq e^{-\varepsilon k} q$, which is the promised claim. For the concentration result, just note that the variables X_l are negatively correlated, which means Chernoff bounds are applicable. $\square$

Proof of Part (2) of Theorem 4.1. We assume that $d = 0$; almost the same proof works for the general case when $d = o(q)$. Thus, we present the proof assuming that $n = q$.

Proof of Part (a) We run the process from round 1 to round k. We prove that in each round j, the expected number of unassigned students is at most $\theta n/j$, where $\theta \geq 1$ is a constant that we will set later. The proof is by induction. The induction base is $j = 1$, which holds trivially. Suppose that the induction hypothesis holds for $j = i$; for the induction step we need to prove that $\mathbb{E}[U_{i+1}] \leq \theta n/(i+1)$.

Notice that each empty seat remains empty (at the end of round $i+1$) with probability at most $\left(1 - \frac{1}{qm}\right)^{U_i}$; since the number of empty seats is equal to the number of unassigned students, we then have:

$$\begin{aligned}\mathbb{E}[U_{i+1}|U_i] &\leq U_i\left(1 - \frac{1}{n}\right)^{U_i} \\ &\leq U_i e^{-U_i/n} \leq U_i\left(1 - \frac{U_i}{2n}\right),\end{aligned}$$

where the last inequality is due to $e^{-x} \leq 1 - x/2$, which holds for all $0 \leq x \leq 1$. Then, we take another expectation from both sides of the above inequality to imply that

$$\mathbb{E}[U_{i+1}] \leq \mathbb{E}\left[U_i - \frac{U_i^2}{2n}\right].$$

Now we can use linearity of expectation and then Jensen's inequality to write

$$\mathbb{E}[U_{i+1}] \leq \mathbb{E}[U_i] - \mathbb{E}\left[\frac{U_i^2}{2n}\right] \leq \mathbb{E}[U_i] - \frac{\mathbb{E}[U_i]^2}{2n}.$$

We are almost done. First note that if we have $\mathbb{E}[U_i] = \theta n/i$, then we get

$$\mathbb{E}[U_{i+1}] \leq \mathbb{E}[U_i] - \frac{\mathbb{E}[U_i]^2}{2n} \leq \theta n/i - \frac{(\theta n/i)^2}{2n}.$$

Since for any $\theta \geq 2i/(i+1)$, we have

$$\theta n/i - \frac{(\theta n/i)^2}{2n} \leq \theta n/(i+1),$$

and so by setting $\theta = 2$ we always get

$$\mathbb{E}[U_{i+1}] \leq \theta n/(i+1).$$

This proves the induction step if $\mathbb{E}[U_i] = \theta n/i$. But on the other hand, note that the function $g(x) = x - x^2/(2n)$ is an increasing function of n for all $x < n$. So, even when $\mathbb{E}[U_i] < \theta n/i$, we have

$$\begin{aligned}\mathbb{E}[U_{i+1}] &\leq \mathbb{E}[U_i] - \frac{\mathbb{E}[U_i]^2}{2n} \\ &\leq \theta n/i - \frac{(\theta n/i)^2}{2n} \leq \theta n/(i+1),\end{aligned}$$

as long as we have $\theta n/i \leq n$, which holds for all $i \geq 2$. This completes the induction step. Thus, we have shown that $\mathbb{E}[U_k] \leq 2n/k$ for all $k \leq n$. This proves the theorem.

Proof of Part (b) We prove that w.h.p. in any round j, we have $U_j \leq \theta n/j$, where $\theta = 2 + \varepsilon'$. The induction base is $j = 1$, which holds trivially. Suppose induction hypothesis holds for $j = i$; for the induction step, we prove that w.h.p. we have $U_{i+1} \leq \theta n/(i+1)$. Then a union bound over all steps ensures that w.h.p. every step holds.

Before proving the induction step, notice that we can safely assume that $U_i \geq n^{2/3+\varepsilon}$. Otherwise, we have

$$\theta n/(i+1) \geq n^{2/3+\varepsilon} > U_i \geq U_{i+1},$$

which would prove the induction step. Now, suppose E denotes the set of empty seats at the beginning of round i. For any $l \in E$, let X_l be a binary random variable that is 1 iff l is empty at the end of round i. Note that $U_{i+1} = \sum_{l \in E} X_l$. We prove that w.h.p. U_{i+1} is not too large, in the following sense.

Claim A.5. *Let $\zeta = U_i(1 - 1/n)^{U_i}$, then for any positive $\delta < 1$ we have:*

$$\mathbb{P}[U_{i+1} > \zeta(1+\delta)] \leq e^{-\frac{\delta^2 n^{2/3+\varepsilon}}{6}}. \tag{4}$$

We remark that this is not a direct corollary of Chernoff concentration bounds since $\{X_l\}_{l \in E}$ are neither independent nor negatively correlated. We prove (4) as follows.

Proof of Claim A.5. We define a random process $\mathcal{B}$, which is in correspondence to round i of DA(k). $\mathcal{B}$ is in fact a simple "balls and bins process," defined as follows. In $\mathcal{B}$, students propose to the same school as in (round i of) DA(k); however, when a student s proposes to a school c, student s picks one of the $\overline{q}$ seats in c uniformly at random. Then, s is accepted to c iff the seat she picks is empty.

Let U'_{i+1} denote the number of empty seats at the end of $\mathcal{B}$ and suppose that P, P' respectively denote the PDFs of U_{i+1}, U'_{i+1}. The proof is done is two steps. First, we prove that P stochastically dominates P'. Then, we prove the promised bound for the random variable $U'_{i+1} \sim P'$ (instead of $U_{i+1} \sim P$). The stochastic domination property then implies that the bound holds for U_{i+1} as well.

To prove stochastic domination, we use a simple coupling argument as follows. We start running round i of DA (k) and define $\mathcal{B}$ based on the evolution of DA (k). Unassigned students in $\mathcal{B}$ submit proposals in the same order as (in round i of) DA (k). Suppose that in DA (k), it is the turn of an unassigned student s, who proposes to seat l_s from school c. Let $E(c), E'(c)$ denote the set of empty seats in c in the processes DA (k), $\mathcal{B}$, respectively. We use the variable l'_s to denote the seat for which s in $\mathcal{B}$ proposes to, and define it as follows:

1. If $|E(c)| = \overline{q}$, then $l'_s = l_s$.

2. If $|E(c)| < \overline{q}$, then with probability $|E(c)|/\overline{q}$ let $l'_s = l_s$, and with probability $1 - |E(c)|/\overline{q}$ let $l'_s = l'$, where l' is a seat picked uniformly at random from the set of full seats in c.

It is straightforward to see that in any sample path we have $U_{i+1} \le U'_{i+1}$; i.e., the coupled process ($\mathcal{B}$) will have more unassigned students than DA (k). This holds simply because $\mathcal{B}$ never allocates a seat that was not allocated in round i of DA (k). In other words, our coupling guarantees that $E(c') \subseteq E'(c')$ always holds during the process, for any school $c' \in C$. Since $U_{i+1} \le U'_{i+1}$ in any sample path, P stochastically dominates P'. Consequently, in order to prove the claim, we just need to show that

$$\mathbb{P}\left[U'_{i+1} > \zeta(1+\delta)\right] \le e^{-\frac{\delta^2 n^{2/3+\varepsilon}}{6}}.$$

To this end, let X'_l be a binary random variable that is 1 iff seat l is still empty at the end of $\mathcal{B}$. Note that $U'_{i+1} = \sum_{l \in E} X'_l$. Let $\mu' = \mathbb{E}\left[U'_{i+1}\right]$. Since the random variables $\{X'_l\}_{l \in E}$ are negatively correlated (which holds by the definition of $\mathcal{B}$), we have

$$\mathbb{P}\left[U'_{i+1} > \mu'(1+\delta)\right] \le e^{-\frac{\delta^2 \mu'}{3}},$$

for any $\delta > 0$. Because of the stochastic dominance, we have

$$\mathbb{P}\left[U_{i+1} > \mu'(1+\delta)\right] \le e^{-\frac{\delta^2 \mu'}{3}}. \tag{5}$$

To complete the proof, we compute an upper bound μ'_U and a lower bound μ'_L for μ'. We then plug these values into (5) to get

$$\mathbb{P}\left[U_{i+1} > \mu'_U(1+\delta)\right] \leq e^{-\frac{\delta^2 \mu'_L}{3}}, \tag{6}$$

which proves our claim.

First, we compute μ'_U. Fix a seat l and an unassigned student s; notice that in $\mathcal{B}$, this seat receives a proposal from s with probability at least $1/(\overline{q}m) = 1/n$. Consequently, $\mathbb{P}\left[X'_l = 1\right] \leq (1-1/n)^{U_i}$, which implies that $\mu' \leq U_i(1-1/n)^{U_i}$. Thus, we set

$$\mu'_U = U_i(1-1/n)^{U_i}.$$

To find μ'_L, notice that in $\mathcal{B}$, s proposes to l with probability at most $\frac{1}{\overline{q}(m-i)}$. So we have

$$\begin{aligned}\mathbb{P}\left[X'_l = 1\right] \geq \left(1 - \frac{1}{\overline{q}(m-i)}\right)^{U_i} &\geq 1 - \frac{U_i}{n - \overline{q}i} \\ &\geq 1/2, \end{aligned} \tag{7}$$

where (7) holds with very high probability by Lemma B.2. So, we can set μ'_L to be any number not larger than $U_i/2$. Now, recall that we assumed $U_i \geq n^{2/3+\varepsilon}$. So, we can safely set $\mu'_L = n^{2/3+\varepsilon}/2$.

Note that $\zeta = \mu'_R$, and plug the values for μ'_L, μ'_R into (6); this completes the proof:

$$\mathbb{P}\left[U_{i+1} > \zeta(1+\delta)\right] \leq e^{-\frac{\delta^2 n^{2/3+\varepsilon}}{6}}.$$

□

We use Claim A.5 to prove the induction step; this is done by finding the δ for which $\zeta(1+\delta) \leq \theta n/(i+1)$; after finding such δ, we plug it into (4) and complete the proof. From the latter inequality, we should have:

$$\delta \leq \frac{\theta n}{\zeta(i+1)} - 1. \tag{8}$$

So, to find the right value for δ, we provide an upper bound on ζ and plug it into the right-hand side of (8). This is done as follows.

$$\zeta = \left(1 - \frac{1}{n}\right)^{U_i} \leq U_i e^{-U_i/n} \leq U_i\left(1 - \frac{U_i}{2n}\right),$$

where the last inequality is due to $e^{-x} \leq 1 - x/2$, which holds for all $0 \leq x \leq 1$. Using the induction hypothesis, we can rewrite the above inequality as:

$$\zeta \leq \theta n/i \left(1 - \frac{\theta n/i}{2n}\right), \tag{9}$$

where in writing (9) we have used the fact that $U_i \left(1 - \frac{U_i}{2n}\right)$ is an increasing function of U_i for all $U_i < n$.

By plugging (9) into (8) we get $\delta \leq \varepsilon'/i$. Since $i \leq n^{1/3-\varepsilon}$, we can set $\delta = \varepsilon' n^{\varepsilon-1/3}$. Now, we are ready to finish the proof using (4). Recall that our choice of δ guarantees that $\zeta(1+\delta) \leq \theta n/(i+1)$. So, we can use (4) to write

$$\mathbb{P}\left[U_{i+1} > \theta n/(i+1)\right] \leq e^{-\frac{\delta^2 n^{2/3+\varepsilon}}{6}}.$$

Since we have $\delta = \varepsilon' n^{\varepsilon-1/3}$, we can rewrite the above bound as

$$\mathbb{P}\left[U_{i+1} > \theta n/(i+1)\right] \leq e^{\frac{-(\varepsilon')^2 n^{3\varepsilon}}{6}}. \tag{10}$$

To complete the proof, we just need a union bound over all rounds: (10) holds for all i with probability at least $1 - ke^{-(\varepsilon')^2 n^{3\varepsilon}/6}$, i.e., with very high probability. □ □

Finally, we provide here a counterpart for Theorem 4.1 for the case in which students are on the short side. The proof is very similar to the proof of Theorem 4.1; accordingly, we omit the proof.

Theorem A.6. *Let U_k denote the number of unassigned students given that students submit proposals for only their k most preferred schools under the MTB rule.*

1. *If $n = (1-\varepsilon)q$ for some constant $\varepsilon \in \mathbb{R}_+$, then:*
   - *(a) $\mathbb{E}\left[U_k\right] \leq e^{-\varepsilon k} n$,*
   - *(b) U_k is not much larger than $\mu = e^{-\varepsilon k} n$, in the following sense: (Chernoff bounds)*

$$\mathbb{P}\left[\Delta > \mu(1+\delta)\right] \leq e^{\frac{-\delta^2 \mu}{3}} \qquad \forall\, 0 < \delta < 1$$

$$\mathbb{P}\left[\Delta > \delta\mu\right] \leq \left(\frac{e^{\delta-1}}{\delta^\delta}\right)^\mu \qquad \forall\, \delta > 1$$

2. *If $n = q - d$ for some $d = o(q)$, then:*
   - *(a) $\mathbb{E}\left[U_k\right] \leq 2n/k$ for all $k \leq m$, and*
   - *(b) W.v.h.p. $U_k \leq (2+\varepsilon')n/k$, for all $\varepsilon, \varepsilon' > 0$ and $k \leq q^{1/3-\varepsilon}$.*

B Proofs of Lemmas used in Theorem 3.2

Proof of Lemma A.1. The proof contains two main steps. Fix any round of DA (r) and suppose that there are more than m^δ empty schools at the beginning of this round. This also means that there should be at least $\bar{q}m^\delta$ unassigned students. In the first step of the proof, we will prove that with high probability, at least $O(\sqrt{m})$ of the unassigned students get assigned by the end of this round. Then, in the second step, we will use a union bound over all r rounds to show that with high probability, we have at most m^δ empty schools at the end of round r.

First, we show that if there is a subset E of empty schools with size at least m^δ at the beginning of a round, then with high probability $O(\sqrt{m})$ students must get assigned to these schools by the end of this round. For each $s \in E$, let X_s be a binary random variable that is set to 0 if school s is still empty at the end of this round and is set to 1 otherwise. Each unassigned student then proposes to school s with probability at least $1/m$, and the probability that s receives no proposals by the end of this round is at most $(1-1/m)^{\bar{q}m^\delta}$. Since this quantity is at most

$$(1-1/m)^{\bar{q}m^\delta} \leq e^{-\bar{q}m^{\delta-1}} \leq \frac{1}{1+\bar{q}m^{\delta-1}} \leq 1 - \frac{\bar{q}m^{\delta-1}}{2},$$

we have $X_s = 1$ with probability at least $\bar{q}m^{\delta-1}/2$. This means that $\mathbb{E}\left[\sum_{s\in E} X_s\right] \geq \bar{q}m^{2\delta-1}/2$. We show that with high probability, the sum is not too small relative to its mean.

To prove the latter fact, we use a Chernoff bound on the set of variables $\{X_s\}_{s\in E}$. A straightforward calculation shows that these variables are negatively correlated, and so Chernoff bounds are applicable. Let $\mu = \bar{q}m^{2\delta-1}/2$. By Chernoff bounds we have

$$\mathbb{P}\left[\sum_{s\in E} X_s \leq \mu(1-\epsilon)\right] \leq e^{\frac{-\epsilon^2\mu}{2}}.$$

Setting $\epsilon = 1/2$ implies

$$\mathbb{P}\left[\sum_{s\in E} X_s \leq \mu/2\right] \leq e^{-\frac{\bar{q}m^{2\delta-1}}{16}}.$$

So far, we have shown that if $|E| \geq m^\delta$ at the beginning of a round, then, with probability at least $1 - e^{-\frac{\bar{q}m^{2\delta-1}}{16}}$, at least $\bar{q}m^{2\delta-1}/4$ of the schools in E will not be empty at the end of that round. Using a union bound over all the r rounds implies that with probability $1 - r \cdot e^{-\frac{\overline{q}m^{2\delta-1}}{16}}$, we have at most $\max\{m^\delta, m - r \cdot \overline{q}m^{2\delta-1}/4\}$ empty schools by the end of round r. Observing that $r \cdot \overline{q}m^{2\delta-1}/4 \geq m$ proves the lemma. □

Proof of Lemma A.2. Let E_1 be the event that DA(r) observes at least $m - m^\delta$ different schools in the sequence S. By Lemma A.1, E_1 happens with high probability. Also, let E_2 be the event that $S[l]$ does not contain $m - m^\delta$ different schools. Lemma D.1 shows that E_2 happens with high probability. A union bound over the probabilistic bounds provided by Lemmas A.1 and D.1 implies that $E_1 \wedge E_2$ happens with probability at least $1 - \frac{m^\delta+1}{e^t} - r \cdot e^{-\frac{\overline{q}m^{2\delta-1}}{16}}$. When $E_1 \wedge E_2$ is true, DA(r) reads a prefix of length at least l from S, which means it makes at least l proposals. □

Lemma B.1. *Fix a student s. The probability that s makes more than 4 proposals in any round of DA(r) is at most $(r/m)^4$.*

Proof. Suppose that we are in round t. Since there are at most r rounds, s has made at most r proposals so far. The probability that s makes 4 redundant proposals is then at most $(r/m)^4$. □

Lemma B.2. *Suppose that we are running DA(k) with $n = q$. Denote the expected number of unassigned students at the end of round 1 by U_1. Then, the following holds:*

1. $\mathbb{E}[U_1] \leq q/e$.
2. *For any positive $\delta < 1$, U_1 is not larger than $(1+\delta)q/e$ with very high probability.*

Proof. We prove the lemma for when $\overline{q} = 1$. It is straightforward to use a coupling argument (similar to the coupling in the proof of Claim A.5) and show that the same bounds hold for $\overline{q} > 1$. For each school $c \in \mathcal{C}$, let X_c be a binary random variable that is 1 if and only if c has received no proposals by the end of round 1. Notice that

$$\mathbb{P}[X_c = 1] = (1 - 1/n)^n \leq e^{-1}. \tag{11}$$

So, we have $\mathbb{E}[U_1] = \sum_{c \in \mathcal{C}} \mathbb{E}[X_c] \leq n/e$, which proves Part 1. To prove Part 2, we use the negative correlation of random variables $\{X_c\}_{c \in \mathcal{C}}$ to apply Chernoff concentration bounds. To this end, first we need to give a lower bound on $\mathbb{E}[U_1]$. See that

$$\mathbb{P}[X_c = 1] = (1 - 1/n)^n \geq e^{-1.01},$$

which means that $\mathbb{E}[U_1] \geq ne^{-1.01}$. Using this lower bound and the upper bound given by (11) we can write the following bound, which completes the proof:

$$\mathbb{P}\left[U_1 > ne^{-1.01}(1+\delta)\right] \leq e^{\frac{-\delta^2 n}{3e}}, \qquad \forall\, 0 < \delta < 1.$$

□

C Couplings required for the proof of Theorem 4.1

Lemma C.1. *Let $P : \mathbb{Z}_+ \to \mathbb{R}_+$ and $P' : \mathbb{Z}_+ \to \mathbb{R}_+$ respectively denote the PMF of unassigned students in DA (k) and DA$'$ (k). Then, P stochastically dominates P'.*

Proof. To prove stochastic domination, we couple the random process $\mathrm{DA}'(k)$ with $\mathrm{DA}(k)$; i.e., we start running $\mathrm{DA}(k)$ and define another random process $\overline{\mathrm{DA}'}(k)$ based on the evolution of $\mathrm{DA}(k)$. We define this coupling so that the resulting process $\overline{\mathrm{DA}'}(k)$ becomes the same process as $\mathrm{DA}'(k)$.

Suppose that we are in round i of $\mathrm{DA}(k)$; the coupling is then defined as follows. Let $q = \epsilon q$ and let $Q = \{s_1, \ldots, s_q\}$ denote the first q unassigned students visited in $\mathrm{DA}(k)$. Also, let $Q' = \{s'_1, \ldots, s'_q\}$ be the q new students arriving in round i of $\overline{\mathrm{DA}'}(k)$. We define the proposals of the students in Q' based on the proposals made by the students in Q. Suppose that s_j has applied to the subset $H_j \subseteq \mathcal{C}$ of schools in previous rounds, and is applying to (a new) school c_j in this round. Suppose $E(c), E'(c)$ denote the set of empty seats in any school $c \in \mathcal{C}$ respectively in the processes $\mathrm{DA}(k), \overline{\mathrm{DA}'}(k)$. Our coupling would guarantee that $E(c) \subseteq E'(c)$ always holds during the process.

Moreover, suppose that l_j denotes the seat (from c_j) that s_j is assigned to; set $l_j = \emptyset$ if s_j is rejected from c_j. We now define the proposal made by s'_j in round i of $\overline{\mathrm{DA}'}(k)$. Let l'_j denote the seat for which s'_j submits a proposal to, and define it as follows:

1. With probability $|H_j|/m$, s'_j applies to a school c picked uniformly at random from H_j. If $E'(c) = \emptyset$, then let $l'_j = \emptyset$; otherwise, let l'_j be an empty seat selected uniformly at random from c.

2. Otherwise (with probability $1 - |H_j|/m$), s'_j applies to c_j. If $E'(c_j) = \emptyset$, then let $l'_j = \emptyset$. If $E'(c_j) \neq \emptyset$, then with probability $\frac{|E'(c_j) - E(c_j)|}{|E'(c_j)|}$, let l'_j be a seat picked uniformly at random from $E'(c_j) - E(c_j)$; and with probability $1 - \frac{|E'(c_j) - E(c_j)|}{|E'(c_j)|}$, let $l'_j = l_j$.

Define $\overline{U'}_k$ to be the number of empty seats at the end of $\overline{\text{DA}'}(k)$. It is straightforward to see that in any sample path we have $U_k \leq \overline{U'}_k$; i.e., $\overline{\text{DA}'}(k)$ will have more empty seats than $\text{DA}(k)$. This holds simply because in any round i, $\overline{\text{DA}'}(k)$ never allocates a seat that was not allocated in $\text{DA}(k)$. Since $U_k \leq \overline{U'}_k$ in any sample path, then P stochastically dominates $\overline{P'}$, the PMF of the number of unassigned seats at the end of $\overline{\text{DA}'}(k)$. To complete the proof, it is enough to observe that $\overline{\text{DA}'}(k)$ is the same random process as $\text{DA}'(k)$, by definition. □

Lemma C.2. *Let $P' : \mathbb{Z}_+ \to \mathbb{R}_+$ and $P'' : \mathbb{Z}_+ \to \mathbb{R}_+$ respectively denote the PMF of unassigned students in $DA'(k)$ and $DA''(k)$. Then, P' stochastically dominates P''.*

Proof. To prove stochastic domination, we couple the random process $\text{DA}''(k)$ with $\text{DA}'(k)$; i.e., we start running $\text{DA}'(k)$ and define another random process $\overline{\text{DA}''}(k)$ based on the evolution of $\text{DA}'(k)$. We define this coupling so that the resulting process $\overline{\text{DA}''}(k)$ becomes the same process as $\text{DA}''(k)$.

Suppose that we are in round i of $\text{DA}'(k)$; the coupling is then defined as follows. Let $q = \epsilon q$ and let $Q' = \{s'_1, \ldots, s'_q\}$ be the q new students arriving in round i in $\text{DA}'(k)$. Also, let $Q'' = \{s''_1, \ldots, s''_q\}$ be the q new students arriving in round i of $\overline{\text{DA}''}(k)$. We define the proposals of the students in Q'' based on the proposals made by the students in Q'. Suppose that s'_j applies to school c_j in this round; s''_j also applies to c_j. To complete the definition of our coupling, it remains to define the seat assigned to s''_j. Suppose l_j denotes the seat (from c_j) that s_j is assigned to; set $l_j = \emptyset$ if c_j is full (and so cannot accept s_j). The seat assigned to s''_j in the process $\overline{\text{DA}''}(k)$ is denoted by l''_j and is defined as follows:

1. If c_j is full in $\text{DA}'(k)$, then let l''_j be a seat picked uniformly at random from the set of seats in c_j.

2. Otherwise, let $E(c_j)$ denote the set of empty seats in school c_j in the process $\text{DA}'(k)$. With probability $1 - \frac{|E(c_j)|}{\bar{q}}$ let l''_j be a seat picked uniformly at random from the set of full seats in c_j. Otherwise (with probability $\frac{|E(c_j)|}{\bar{q}}$), let $l''_j = l_j$.

Define $\overline{U''}_k$ to be the number of empty seats at the end of $\overline{\text{DA}''}(k)$. It is straightforward to see that in any sample path we have $U'_k \leq \overline{U''}_k$; i.e., $\overline{\text{DA}''}(k)$ will have more empty seats than $\text{DA}'(k)$. This holds simply because in any round i, $\overline{\text{DA}''}(k)$ never allocates a seat that was not allocated in $\text{DA}'(k)$. Since $U'_k \leq \overline{U''}_k$ in any sample path, then P' stochastically dominates $\overline{P''}$, the PMF of the number of unassigned seats at the end of $\overline{\text{DA}''}(k)$. To complete the proof, it is enough to observe that $\overline{\mathrm{DA}''}(k)$ is identically the same random process as $\mathrm{DA}''(k)$, by definition. ☐

D Missing proofs

Lemma D.1. *Let $l < m(\ln m - t)$ for some $t > 0$. Then, with probability at least $1 - \frac{m^\delta+1}{e^t}$, $S[l]$ contains fewer than $m - m^\delta$ different schools.*

Proof. Let $\zeta = m - m^\delta$. For any $i \geq 1$, let X_i be a variable that denotes the smallest integer j such that $S[j]$ contains i different schools. Define $X_0 = 0$. Note that since S is a random variable, so is X_i. Also, define $Z_i = X_i - X_{i-1}$ for all positive i. It is straightforward to see that Z_i is a geometric random variable with mean $\frac{m}{m-i+1}$. We provide a (Chernoff-type) concentration bound such that the random variable $Z = \sum_{i=1}^{\zeta} Z_i$ is highly concentrated around its mean. To do so, first notice that

$$\mathbb{P}[Z < \beta] = \mathbb{P}\left[e^{-\theta Z} > e^{-\theta\beta}\right] \leq \mathbb{E}\left[e^{\theta\beta - \theta Z}\right]. \tag{12}$$

Now we use the independence of Z_i's to rewrite (the right-hand side of) (12):

$$\begin{aligned}\mathbb{P}[Z < \beta] \leq \mathbb{E}\left[e^{\theta\beta-\theta Z}\right] &= e^{\theta\beta} \cdot \prod_{i=1}^{\zeta} \mathbb{E}\left[e^{-\theta Z_i}\right]. \\ &= e^{\theta\beta} \cdot \prod_{i=1}^{\zeta} \frac{\frac{m-i+1}{m} \cdot e^{-\theta}}{1 - (1 - \frac{m-i+1}{m}) \cdot e^{-\theta}}, \end{aligned} \tag{13}$$

where (13) is due to Proposition D.2. Then, we choose $\theta = 1/m$ and use the fact that $e^{1/m} \geq 1 + 1/m$ to bound the right-hand side of (13)

$$\begin{aligned}\mathbb{P}[Z < \beta] &\leq e^{\theta\beta} \cdot \prod_{i=1}^{\zeta} \frac{\frac{m-i+1}{m} \cdot e^{-\theta}}{1 - (1 - \frac{m-i+1}{m}) \cdot e^{-\theta}} \\ &\leq e^{\theta\beta} \cdot \prod_{i=1}^{\zeta} \frac{\frac{m-i+1}{m}}{1 + 1/m - (1 - \frac{m-i+1}{m})} \\ &= e^{\theta\beta} \cdot \prod_{i=1}^{\zeta} \frac{m-i+1}{m-i+2} = e^{\theta\beta} \cdot \frac{m-\zeta+1}{m+1}. \end{aligned} \tag{14}$$

Plugging $\beta = m(\ln m - t)$ into (14) implies that

$$\mathbb{P}[X_\zeta < m(\ln m - t)] \leq e^{-t}(m - \zeta + 1) = \frac{m^\delta + 1}{e^t}.$$

Then, recall that $l < m(\ln m - t)$. Thus, this bound says the probability of seeing ζ different schools in $S[l]$ is at most $\frac{m^\delta+1}{e^t}$. The lemma is proved. □

Proposition D.2. *Suppose that X is a geometric random variable with mean $1/p$. Then for any $\theta > 0$ we have $\mathbb{E}\left[e^{-\theta X}\right] = \frac{pe^{-\theta}}{1-(1-p)e^{-\theta}}$.*

Proof.

$$\begin{aligned}\mathbb{E}\left[e^{-\theta X}\right] &= \sum_{i=1}^{\infty} p(1-p)^{i-1}e^{-i\theta} \\ &= \frac{p}{1-p} \cdot \sum_{i=1}^{\infty} \left((1-p)e^{-\theta}\right)^i = \frac{pe^{-\theta}}{1-(1-p)e^{-\theta}}\end{aligned}$$

□

E A few simple properties of STB

Proof of Proposition 3.3. Without loss of generality, assume that students are ordered from 1 to n. When it is the turn of student $i+1$ to select a school, there are already i assigned students. The probability of success at each attempt made by student $i+1$ is at least $\frac{m-i/\overline{q}}{m} = \frac{q-i}{q}$. So, the expected number of attempts made by student $i+1$ is upper bounded by $\frac{q}{q-i}$. Consequently, the expected total number of attempts made is upper bounded by

$$\sum_{i=1}^{n} \frac{q}{q-i} = O(q \cdot (\ln q - \ln \min\{q-n, 1\})).$$

Now consider the two cases of $q \geq n$ and $q < n$ separately and plugging them into the above expression proves the lemma. □

Proof of Proposition 3.4. Suppose that students are ordered from 1 (highest priority) to n (lower priority) in the tie-breaking. It is easy to see that if instead of running DA, we run Random Serial Dictatorship (RSD) in this order (student 1 choosing first), we will get exactly the same outcome. Now, notice that in RSD, student i gets his top choice with probability at least $\frac{m-i/\overline{q}}{q}$, which is at least $\frac{q-i}{q}$. Thus $\sum_{i=1}^{t} \frac{q-i}{q} \geq t/2$ proves the first claim. The second claim can be proved using a Chernoff bound. We omit the full proof here. □