Assigning more students to their top choices: A tiebreaking rule comparison

Itai Ashlagi Afshin Nikzad Assaf Romm *

June 2015

Abstract

School choice districts that implement stable matchings face various decisions that impact students' assignments to schools. We study properties of the rank distribution of students with random preferences, when schools use different tiebreaking rules to rank equivalent students. Under a single tiebreaking rule, in which all schools use the same ranking, a constant fraction of students are assigned to one of their top choices. In contrast, under a multiple tiebreaking rule, in which each school independently ranks students, a vanishing fraction of students match to one of their top choices. We show, however, that by restricting students to submit relatively short preference lists under a multiple tiebreaking rule, a constant fraction of students will match to one of their top choices while only a "small" fraction of students will remain unmatched.

1 Introduction

Two-sided matching markets have become an exciting research area following the work of 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 their work, little is known about to whom should agents expect to match 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 NYC (Abdulkadiroğlu et al. (2009) and Boston (Abdulkadiroğlu et al. (2005)), adopted centralized mechanisms to assign students based on Gale

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

and Shapley's Deferred Acceptance algorithm, which selects a stable matching with respect to schools' and students' preferences over each other (see also Abdulkadiroğlu and Sönmez (2003)). Schools, however, are often not strategic and only use artificial preferences to break ties among equivalent students. Another decision school districts make is how many schools should students rank. This paper provides some insights about the impact of such choices on the students' rank distribution.[1]

Two tiebreaking rules have been considered by school districts. Under the multiple tiebreaking rule (MTB) each school independently selects a random order over students for breaking ties, and under the single tiebreaking rule (STB) all schools use the same a priori order over students which is selected uniformly at random. The choice of which tiebreaking rule to use under the Deferred Acceptance mechanism was raised by Abdulkadiroğlu and Sönmez (2003), who suggested that MTB may result in unnecessary inefficiency.[2] School districts are often interested in the MTB rule, which naturally seems more equitable. Abdulkadiroğlu et al. (2009) compared empirically STB and MTB using NYC choice data and found that STB results in more students receiving top choices while more students receive "bad" choices or remain unassigned. We take here 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) be rejected from other schools. This intuition suggests that indeed STB yields more top choices. We provide, however, 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 tiebreaking rule affects the rank distribution for students and ask how many students are likely to be assigned to one of their "top" choices under each tiebreaking 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.

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

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

[2]See (Abdulkadiroğlu and Sönmez, 2003, Footnote 14).

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 an intuition that for the fact that 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 students' assignments. 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 in which 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 commonly known to be better than other). In an extreme case, when 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 by 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 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, under STB students who are temporarily admitted to their first choice under the Deferred Acceptance algorithm, are less likely to be rejected than under MTB since students are rejected from other schools are selected to have very bad "lotteries".

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 a similar analysis to one-to-one matching markets.[3] Our method allows 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

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

capacities. The main idea in our analysis is to simplify the random process that corresponds to the Deferred Acceptance algorithm by *coupling* it with a relatively simpler random process.

1.1 Related Work

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

Abdulkadiroğlu et al. (2009) raise the question of choosing STB or MTB in the context of the NYC school choice system. Independently and concurrently to our work, Arnosti (2015) studies the impact of STB and MTB on the number of matches formed when agents preferences are short and random, and also finds STB leads to many more students receiving their top choices than MTB. When the mechanism of choice is the Top Trading Cycles mechanism, both Pathak and Sethuraman (2011) and Carroll (2014) extend the results by Abdulkadiroğlu and Sönmez (1998) and show that there is no difference between single tiebreaking (i.e., using Random Serial Dictatorship) and multiple tiebreaking (TTC with random endowments). It is important to note that indifferences in schools' priorities over students have also induced other very interesting approaches, among which are the stable improvement cycles of Erdil and Ergin (2008), the efficiency-adjusted DA of Kesten (2011) and the choice-augmented DA of Abdulkadiroglu et al. (2008).

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 truncations and droppings strategies as suggestions to match agents isn't novel (see for example Roth and Rothblum (1999), Kojima and Pathak (2009)).[4]

We are not the first to study 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 the men-optimal stable matching, men's average rank is approximately $\ln n$ and every man is assigned to his 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,

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

which shows that only 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 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; so, the most preferred school for s has rank 1. Unless mentioned otherwise, students preferences over schools are drawn independently and uniformly at random.

Each school $c \in \mathcal{C}$ has a priority ranking over students which breaks 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 with lower priority than s. A matching is *stable* if it is not unstable.

Gale and Shapley (1962) have shown that a stable matching always exists and, under strict preferences, there is one 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 Tiebreaking rules

We consider two common tiebreaking rules for which school districts determine priority rankings of schools over students. Under a *multiple tiebreaking rule* (MTB) each school independently selects a priority ranking over all students. Under a *single tiebreaking rule* (STB) all schools use the same priority order, which is selected uniformly at random a priori. One way to implement STB is by assigning 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 tiebreaking rules.

2.2 No dominance in first choices

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

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 in expectation the fractions of students that obtain their first choice under STB and under MTB are 0.9951 and 0.9955, respectively.[5]

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 2,3,4 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. Under STB, these rejected students are less likely to be accepted to their second choice, c_2, 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 however, that STB leads to more students receiving their first choices in expectation under "nice distributions", such as when preferences are all drawn from a symmetric logit choice model. This example considered in this section motivates us to study large markets, in which each student has random preferences.

2.3 Examples with large capacities

In this section, we use a couple of examples to illustrate that the students' rank distribution under STB (almost) first order stochasticly dominates the rank distribution under MTB, when there is a

[5]Two different softwares on two different computers to obtain these exact results.

small number of schools with large capacities.[6] Then in the remainder of the paper we would study students' rank distribution when schools have small capacities relative to the size of the market.

We study the effect of STB and MTB on the distribution of ranks using two simple examples in which schools have large capacities. Consider a school choice problem with two schools, each of which has q seats. Also, suppose there are $n = 2q$ (or more) students.[7] 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 will be identical. This holds because only one school will be overdemanded by approximately $\sqrt{q}$ students and all rejected students will be assigned to their second choice which is underdemanded.

To make this example more interesting, suppose that there are only $0.8q$ seats in each of the two schools.[8] (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 will obtain their first choice under STB since rejected students are likely to have a very low lottery number and thus very unlikely to be cause a rejection of another student in their second choice who in turn has a very high lottery number.

When capacities are small obtaining such qualitative insights is more involved since the lottery numbers of rejected students does not provide as 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 we study in this section properties of the rank distribution in school choice problems with many schools, each of which has a "small" capacity and when 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.

3.1 Properties of the rank distribution under MTB

Ashlagi et al. (2013) have shown that there is a significant advantage to be on the short side in a two-sided matching market. In particular, students have a much better rank if there are sufficiently

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

[7]It may be convenient for the reader to think about a continuum model with a student mass of 2 and capacity of mass 1 in each school.

[8]In the continuum version the capacity is now a mass of 0.8.

many seats than in the case that 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.*

So when there is a shortage of seats students are not matching well on average in contrast to the case in which students are on the short side. Moreover, the larger the excess of seats the better choices students' obtain .

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.

Theorem 3.2. *Suppose $q \leq n + d$ for some constant d (either positive or negative).[9] Then for any constant k, $R_k(MTB)$ approaches 0 as n approaches infinity.[10]*

We remark here that Theorem 3.2 holds true when every agent has a ranking list of length $f(n)$ for any increasing function of n.

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 that list school c in one of their top k choices, denoted by $\psi(k)$, is very small (sub-logarithmic in n). Therefore, since each student that 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.

[9] In fact, our result also holds when there is a large sublinear excess. In particular, 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).

[10] 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 $\overline{q}$ grows slower than $\ln n$. If $\bar{q}$ is a constant, then the proof also holds if k grows slower than $\ln n$.

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 to get 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 for 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 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 success probability at least $\frac{i}{m-i}$ and summing up these random variables gives the result. Observe that Proposition 3.3 holds even if students are on the long side in contrast to the MTB setting.

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.*[11] *at least* $(1-\varepsilon)(t/2)$ *for any* $\varepsilon > 0$.

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. This provides matched students one of their listed choices and thus, if the length of the list is short enough, this can be

[11] 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 $\theta > 0$.

an opportunity to get around Theorem 3.2. Of course, shortening the lists also may increase the number of unassigned students. We next analyze this tradeoff.

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 $U_k \le \varepsilon q + \Delta$, where*
   - *(a)* $\mathbb{E}[\Delta] \le e^{-\varepsilon k} q$,
   - *(b)* *Δ is not much larger than $\mu = e^{-\varepsilon k} q$, in the following sense: (Chernoff bounds)*

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

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)* $\mathbb{E}[\Delta] \le 2q/k$ *for all* $k \le m$.
   - *(b) W.v.h.p.* [12] $\Delta \le (2+\varepsilon')q/k$, *for all* $\varepsilon, \varepsilon' > 0$ *and* $k \le q^{1/3-\varepsilon}$.

It is important to note here that when lists are short, the 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 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 they 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 they rank low (compared to other good schools) on their list. 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. $\Delta = 0$.*

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

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

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 in which there is an excess of seats, and observe that very similar bounds hold (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 tiebreaking lottery might be a better idea in school districts that adopt 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 tiebreaking will not result in additional efficiency loss if it is carried out through a single lottery (while that is likely to happen if the tiebreaking is independently carried out across different schools). (Abdulkadiroğlu and Sönmez, 2003, Footnote 14)

This paper provides a first analytical comparison between STB and MTB in a rigorous way in a model with random preferences.[13] Abdulkadiroğlu et al. (2009) show that STB and MTB cannot be compared in terms of stochastic dominance, i.e. neither stochastically dominates the other one, and that no strategy-proof and stable mechanism performs better than any specific tiebreaking 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 insight for the latter observation (Theorem 3.2). The intuition is roughly as follows. Consider what is the probability that student s that has been assigned to school c at the end of the first round of DA, will be rejected by the school at the second round. Under STB, since student s was assigned to her top choice, we learn that she probably had a good lottery number, whereas future contenders to her seat that were rejected from several other schools probably have a bad lottery number and are unlikely to cause s to be rejected. In 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 that has been rejected

[13] Our analysis ignores multiple priority classes.

from many schools. Under STB it is likely she will also be rejected from the next school she applies to, because she probably has a low lottery number, in contrast to MTB. 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 a at least her $\frac{n}{2}$ rank.Computer simulations suggest that in general there is a tipping point in the ranking distribution, such that STB dominates until that point (assigns more people to high-rank schools), but MTB dominates afterwards (assigns less people to low-rank schools). Formalizing and proving this claim is a very interesting direction for future research.

It is interesting also to gain insights about the distribution of rankings when students are on the long side of the market, as often some schools are better than other schools. In this case our model suggests that STB significantly outperforms MTB both in students' average rank of schools (Theorem 3.1 and Proposition 3.3) as well as students 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 excessive administrative burden.[14] 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 very small number of people with their top choices (this effect is even stronger when the students are on the long side).[15] 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 the implications of our results to the shape of the rank distribution. All figures show the cumulative rank distribution function that

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

[15] We do not formally compare the number of unassigned people under both STB and MTB, however, we believe (based on simulations) that MTB actually performs better in the sense of leaving less people unassigned.

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 number of students increases, the rank distribution becomes more flat. Figure 2 represents a similar plot when there is excess 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.

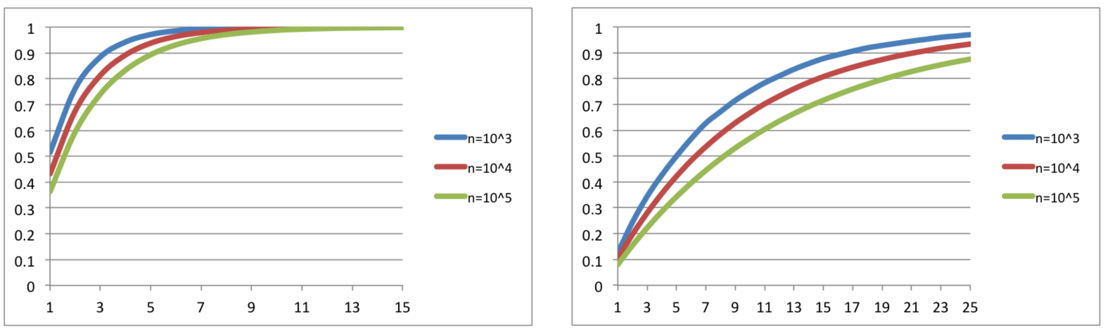

Figure 1: Rank distribution under MTB with $\overline{q} = 10$ (left) and $\overline{q} = 1$ (right)

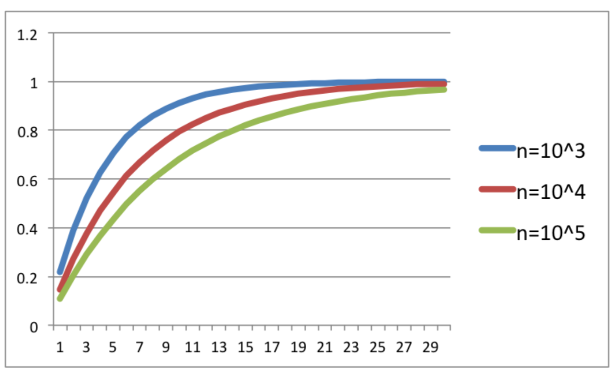

Figure 2: Rank distribution under MTB with $\overline{q} = 1$ and $m = n + 10$ (i.e. when there is excess of seats)

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 the 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

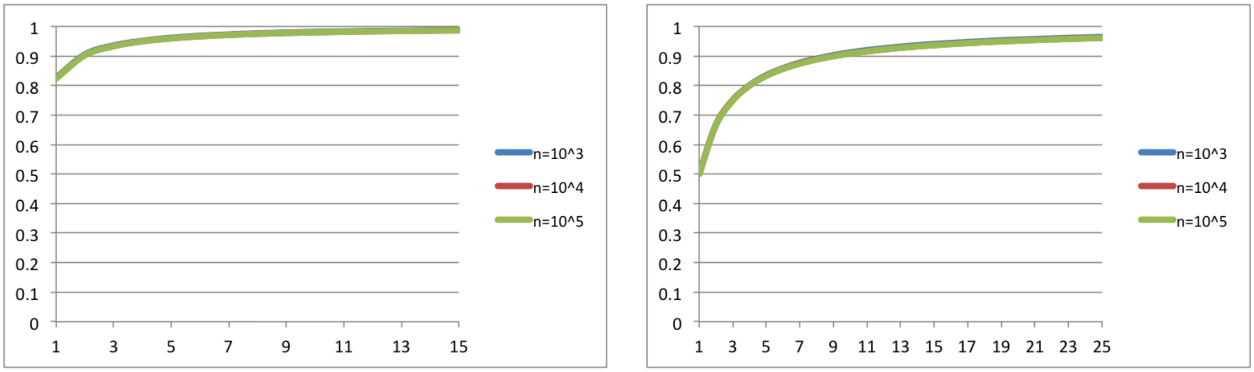

Figure 3: Rank distribution under STB with $\overline{q} = 10$ (left) and $\overline{q} = 1$ (right)

vertical axis). Observe the significant difference between STB and MTB in terms of the expected rank.

The next set of figures relate 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 market is balanced. Each of the three panels correspond to one of the lines in the left panel of Figure 1, where the brigther-colored line is exactly the same as the one in Figure 1, and the darker-colored line represents the rank distribution when students only submit their 5 top 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.

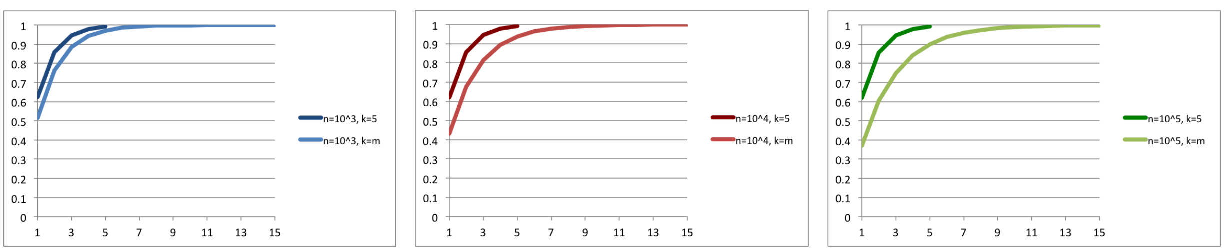

Figure 4: The effect of submitting short lists (balanced markets), $\overline{q} = 10$

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 preferences 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

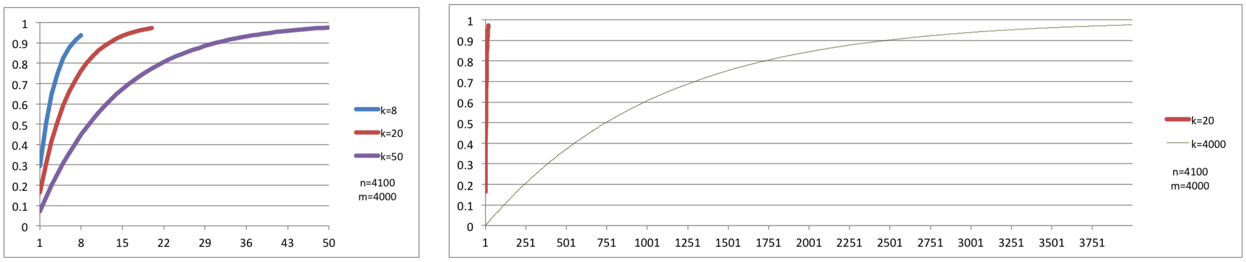

Figure 5: The effect of submitting short lists (unbalanced markets), $\overline{q} = 1$

capacities are relatively large. In particular, $\overline{q} = 100$, $m = 100$ and $n = 100^2 + 100$. Note that when students submit their full preferences, the number of people getting their 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.

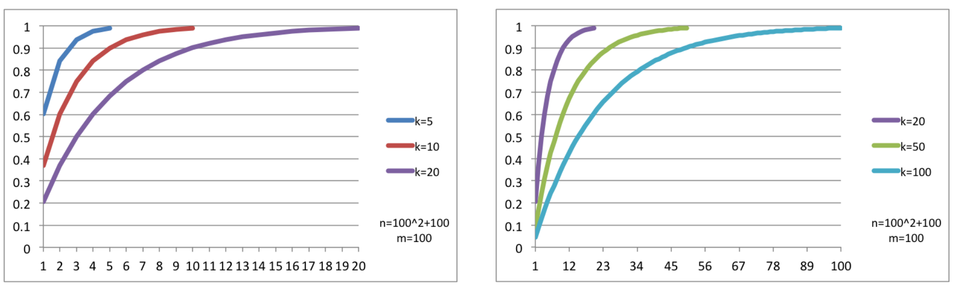

Figure 6: The effect of submitting short lists (unbalanced markets), $\overline{q} = 100$

7 Conclusion

In school choice environments, in which students are often coarsely grouped into priority classes, choosing a tiebreaking rule can have a significant effect on the resulting rank distribution. Information about this distribution is not only important 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 the deferred acceptance mechanism together with the MTB rule leads to much fewer students getting one of their top choices than when paired with the STB rule. Interestingly, shortening preference lists appropriately can overcome much of the drawback 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 in order to understand outcomes in typical markets. While qualitative results provide useful insights, quantitative results should be taken with caution. Our model assumes preferences are drawn independently and uniformly at random mainly for simplicity. This allows to overcome technical issues, but also 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 mention here two open questions. First, we conjecture that STB second order stochastic dominates MTB.[16] Second, we believe that if students preferences independently drawn from a symmetric logit choice model, STB maximizes the number of students that get their first choice.

References

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

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *The 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 Abdulkadiroglu, Yeon-Koo Che, and Yosuke Yasuda. Expanding" choice" in school choice. 2008.

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

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

Itai Ashlagi and Peng Shi. Optimal allocation without money: an engineering approach. Forthcoming in Management Science, 2014.

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

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

Caterina Calsamiglia, Guillaume Haeringer, and Flip Klijn. Constrained school choice: An experimental study. *The American Economic Review*, pages 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*, 2014. forthcoming.

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

D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The 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*, pages 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 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., $[n] = \{1, \ldots, n\}$. Let S be a sequence of integers in $[m]$, such that each integer in $[m]$ appears in S an infinite number of times. Let $S[j]$ denote the prefix of S up until 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 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 less 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 $\mathrm{DA}\,(= \mathrm{DA}\,(\infty))$.

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 $[m]$, while ignoring 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 $\lambda > 0$.

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
```

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 sktech:

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 $\mathbb{E}\left[X_c\right]$ 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. These two facts imply $\mathbb{E}\left[X_c\right] \leq\ O\left(\frac{\overline{q}\,\psi(k)}{\ln n}\right)$, 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 $\mathbb{E}\left[X_c\right]$. Since there are at most $\psi(k)$ such students, then $\mathbb{E}\left[X_c\right] \leq O\left(\frac{\overline{q}\,\psi(k)}{\ln n}\right)$. This proves the promised claim: $\mathbb{E}\left[X_c\right]$ 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 $\mathbb{E}\left[X_c\right] \approx \frac{\overline{q}\,\psi(k)}{\ln(n/\overline{q})}$. So, $\mathbb{E}\left[X_c\right]$ 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.

Now, we state two lemmas that are required to prove Theorem 3.2. In the next lemma, we show that with high probability, most of 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 would be 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. See 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 $\mathbb{E}[X_c]$ 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 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. 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.[17] 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 finish Step 1, we show that w.h.p. school c appears at least $\Omega(\ln m)$ times in $S[j]$. Using this fact we would 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 finish 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 which 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

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

$\mu = \ln m/5$. 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 $m^{-1/40}$.

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 would finish 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 finish 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 which is 1 iff s lists the school c on one of her top k positions. Let $\mu = \mathbb{E}\left[\sum_{s\in\mathcal{S}} Z_s\right]$. See that since the preferences are uniform, then $\mathbb{P}[Z_s = 1] = k/n$, 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 bound:

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

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.[18]

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

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 Step 1 and Step 2, we know that E^* holds w.h.p, i.e. probability at least $1 - m^{-1/40} - 4(4/\overline{q})^4 m^{-1/2} - f(\theta)$. See 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

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

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 in a school. Label each seat with a unique label, and let $\mathcal{L}$ denote the set of labels of all the q available seats. In each round of DA (k), when a student s proposes to a school, if the school has empty seats then 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 is defining a much simpler process, DA$''$ (k), which, roughly speaking, leaves more students unassigned by the end of round k. We would then 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 $P : \mathbb{Z}_+ \to \mathbb{R}_+$ 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 would 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 same identical bounds then would 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 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

is proposing to a school, if the school has 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 $\mathrm{DA}'(k)$ to $\mathrm{DA}''(k)$ while ensuring that P'' is stochastically dominated by P'. We start by defining $\mathrm{DA}''(k)$: 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 is proposing 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 will prove the bounds given in the theorem statement for $\mathrm{DA}''(k)$ (instead of $\mathrm{DA}(k)$). Stochastic dominance then would imply that the bounds hold for $\mathrm{DA}(k)$ as well. More formally, suppose that U''_k denotes the number of unassigned students by the end of $\mathrm{DA}''(k)$. We will 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)

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

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

This would prove the theorem.

Given the above definition for $\mathrm{DA}''(k)$, it is straight-forward 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 which is 1 iff seat l is empty by the end of $\mathrm{DA}''(k)$. By considering applications sent out by the εqk 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 $\mathbb{E}[\Delta''] = \sum_{l \in \mathcal{L}} \mathbb{E}[X_l]$, then 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. □

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

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

See that each empty seat remains empty (by the end of round $i+1$) with probability at most $\left(1 - \frac{1}{\bar{q}m}\right)^{U_i}$; 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, 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

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

Then, see that 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 would 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 show that $\mathbb{E}[U_k] \leq 2n/k$ for all $k \leq n$. This proves the theorem.

Proof for 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 trivially holds. Suppose induction hypothesis holds for $j = i$; for the induction step, we prove that w.h.p. we have $U_{i+1} \leq \theta n/(i + 1)$. In the end, a union bound over all steps would ensure that w.h.p. every step holds.

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

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

which proves 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 which is 1 iff l is empty by 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:*

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

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 by the end of $\mathcal{B}$ and suppose 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 the 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 is proposing to seat l_s from school c. Let $E(c), E'(c)$ denote the set of empty seats in c, respectively in the processes DA (k) $, \mathcal{B}$. We use the variable l'_s to denote the seat proposed to by s in $\mathcal{B}$, 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 straight-forward to see that in any sample path we have $U_{i+1} \leq U'_{i+1}$, 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 is 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 $U_{i+1} \leq U'_{i+1}$ in any sample path, then P stochastically dominates P'. Consequently, in order to prove the claim, we just need to show that:

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

To this end, let X'_l be a binary random variable which is 1 iff seat l is still empty by 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 $\mathcal{B}$), then we have

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

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

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

To complete the proof, we will compute an upper bound μ'_U and a lower bound μ'_L for μ'. We then will 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 will prove our claim.

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

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

To find μ'_L, see 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 finish 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 $\zeta(1+\delta) \leq \theta n/(i+1)$. 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 finish 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; we omit the proof.

Theorem A.6. *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:*
   (a) $\mathbb{E}[U_k] \leq e^{-\varepsilon k} n$,
   (b) *U_k is not much larger than $\mu = e^{-\varepsilon k} n$, in the following sense: (Chernoff bounds)*

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

2. *If $n = q - d$ for some $d = o(q)$, then:*
   (a) $\mathbb{E}[U_k] \leq 2n/k$ *for all* $k \leq m$.
   (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 for the Lemmas used in Theorem 3.2

of Lemma A.1. The proof contains two main steps. Fix any round of DA (r) and suppose there are more than m^{δ} empty schools by the beginning of this round. This also means that there should be at least $\overline{q}m^{\delta}$ 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 by the end of round r.

First, we show that if there is a subset E of empty schools with size at least m^{δ} in 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 by the end of this round and is set to 1 otherwise. Each unassigned student would propose to school s with probability at least $1/m$, so, the probability that s receives no proposals by the end of this round is at most $(1-1/m)^{\overline{q}m^{\delta}}$. Since this quantity is at most

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

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

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

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

So far, we have shown that if $|E| \geq m^\delta$ by the beginning of a round, then with probability at least $1 - e^{-\frac{\overline{q}m^{2\delta-1}}{16}}$, at least $\overline{q}m^{2\delta-1}/4$ of the schools in E will not be empty by 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. □

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 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 in 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 straight-forward 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 which is 1 iff c has received no proposals by the end of round 1. See 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 $\mathbb{E}[U_1] \geq ne^{-1.01}$. Using this lower bound and the upper bound given by (11) we can write the following bound which completes the proof:

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

$\square$

C The Couplings Required in 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 the 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 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 that are 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 the 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 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 that s'_j proposes 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 by the end of $\overline{\mathrm{DA}'}\,(k)$. It is straight-forward to see that in any sample path we have $U_k \leq \overline{U'}_k$, i.e. $\overline{\mathrm{DA}'}\,(k)$ will have more empty seats than $\mathrm{DA}\,(k)$. This holds simply because in any round i, $\overline{\mathrm{DA}'}\,(k)$ never allocates a seat that is not allocated in $\mathrm{DA}\,(k)$. 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 by the end of $\overline{\mathrm{DA}'}\,(k)$. To complete the proof, it is enough to observe that $\overline{\mathrm{DA}'}\,(k)$ is the same random process as $\mathrm{DA}'\,(k)$, 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 the 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 we are in round i of $\mathrm{DA}'\,(k)$; the coupling is then defined as follows. Let $q = \epsilon q$ and let $Q' = \{s'_1, \ldots, s'_q\}$ be the q new students arriving in round i 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 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 $\overline{\mathrm{DA}''}\,(k)$ is denoted by l''_j and is defined as follows:

1. If c_j is full in $\mathrm{DA}'\,(k)$, 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 $\mathrm{DA}'\,(k)$. With probability $1 - \frac{|E(c_j)|}{\overline{q}}$ let l''_j be a seat picked uniformly at random from the set of full seats in c_j. Otherwise (with probability $\frac{|E(c_j)|}{\overline{q}}$), let $l''_j = l_j$.

Define $\overline{U''}_k$ to be the number of empty seats by the end of $\overline{\mathrm{DA}''}(k)$. It is straight-forward to see that in any sample path we have $U'_k \leq \overline{U''}_k$, i.e. $\overline{\mathrm{DA}''}(k)$ will have more empty seats than $\mathrm{DA}'(k)$. This holds simply because in any round i, $\overline{\mathrm{DA}''}(k)$ never allocates a seat that is not allocated in $\mathrm{DA}'(k)$. 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 by the end of $\overline{\mathrm{DA}''}(k)$. 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 less 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 straight-forward to see that Z_i is a geometric random variable with mean $\frac{m}{m-i+1}$. We provide a (chernoff-type) concentration bound that the random variable $Z = \sum_{i=1}^{\zeta} Z_i$ is highly concentrated around its mean. To do this, first see that:

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

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

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

where (13) is due to Proposition D.2. Now, 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):

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

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

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

Now, 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 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 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 $\frac{m-i/\bar{q}}{m} = \frac{q-i}{q}$. 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 students are ordered from 1 (highest priority) to n (lower priority) in the tiebreaking. 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, see that in RSD, student i will get his top choice with probability at least $\frac{m-i/\bar{q}}{q}$, which is at least $\frac{q-i}{q}$. Seeing that $\sum_{i=1}^{t} \frac{q-i}{q} \geq t/2$ proves the first claim. The second claim can be proved using Chernoff bound. We omit the full proof here. □