Online Appendix to Unbalanced Random Matching Markets: The Stark Effect of Competition

Itai Ashlagi Yash Kanoria Jacob D. Leshno

In this Online Appendix we prove the correctness of our new matching algorithm in Appendix A, prove our main results from the paper in Appendix B, and discuss how our results may be extended to many-to-one random matching markets in Appendix C.

A Matching algorithm

The goal of this section is to present Algorithm 2 in the paper, which is the basis of our analysis. This algorithm allows us to calculate the WOSM by a process of successive proposals by men. It first finds the men-optimal stable matching using the men-proposing DA algorithm (Algorithm 1 in the paper, or MPDA), and then progresses to the women-optimal stable matching through a series of divorces of matched women followed by proposals by men. At the end of this section we show how the run of the algorithm on a random matching market is equivalent to a randomized algorithm. In Appendix B we analyze the randomized algorithm to prove Theorem 1 in the paper. Throughout our analysis and this section we assume that there are strictly more women than men, that is, $|\mathcal{W}| > |\mathcal{M}|$.

Before presenting Algorithm 2 we first give a simplified version. The following algorithm, adapted from McVitie and Wilson (1971) and Immorlica and Mahdian (2005), produces the WOSM from the MOSM by finding each woman's most preferred stable match. It uses $\tilde{\mu}$ to maintain the most recent stable matching. Each phase instigates a rejection chain to check whether there is a more women-preferred stable matching, using m to hold the proposing man and μ to hold the temporary assignment. It also maintains a set S of women whose most preferred stable match has been found. Denote the set of women who are unmatched under the MOSM by[1] $\bar{\mathcal{W}}$. We initialize S to be $S = \bar{\mathcal{W}}$, as by the rural hospital theorem these women are unmatched under any stable matching. We use the notation $x \leftarrow y$ for the operation of copying the value of variable y to variable x.

Algorithm 1. MOSM to WOSM (simplified)

- Input: *A matching market with n men and $n + k$ women.*
- Initialization: *Run the men-proposing deferred acceptance to get the men-optimal stable matching μ. Set $S = \bar{\mathcal{W}}$ to be the set of women unmatched under μ. Select any $\hat{w} \in \mathcal{W} \backslash S$.*
- New phase:

  1. *Set $\tilde{\mu} \leftarrow \mu$.*
  2. Divorce: *Set $m \leftarrow \mu(\hat{w})$ and have $\hat{w}$ reject m.*
  3. Proposal: *Man m proposes to his most preferred woman w to whom he has not yet proposed.*[2]
  4. w's Decision:
     (a) *If $w \neq \hat{w}$ prefers her current match $\mu(w)$ over m, or if $w = \hat{w}$ and she prefers $\tilde{\mu}(\hat{w})$ over m, she rejects m. Go to step 3.*
     (b) *If $w \notin \{\hat{w}\} \cup \bar{\mathcal{W}}$, and w prefers m over $\mu(w)$, then w rejects her current partner and accepts*[3] *m. Go to step 3.*
     (c) New stable matching: *If $w = \hat{w}$ and $\hat{w}$ prefers m over all her previous proposals, w accepts m and a new stable matching is found. Select $\hat{w} \in \mathcal{W} \backslash S$ and start a new phase from step 1.*
     (d) End of terminal phase: *If $w \in \bar{\mathcal{W}}$, restore $\mu \leftarrow \tilde{\mu}$, erase rejections accordingly, and add $\hat{w}$ to S. If $S = \mathcal{W}$, terminate and output $\tilde{\mu}$. Otherwise, select $\hat{w} \in \mathcal{W} \backslash S$ and begin a new phase from step 1.*

[1]Since there are more women than men $\bar{\mathcal{W}} \neq \phi$.

[2]See footnote 22 in the paper.

[3]More precisely, we use a temporary variable m' as follows: $m' \leftarrow \mu(w)$, $\mu(w) \leftarrow m$ and $m \leftarrow m'$.

It will be convenient to use the following terminology. A **phase** is a sequence of proposals made by the algorithm between visits to step 1. An **improvement phase** is a phase that terminates at step 4(c) (a new stable matching is found). A **terminal phase** is a phase that terminates at step 4(d) (there is no better stable husband for $\hat{w}$). We refer to the sequence of women who reject their husbands in a phase as the **rejection chain**.

In each phase the algorithm tries to find a more preferred husband for $\hat{w}$, which requires divorcing $m = \tilde{\mu}(\hat{w})$ and assigning him to another woman. In improvement phases the algorithm finds a more preferred stable husband for $\hat{w}$, and updates $\tilde{\mu}$ to the new stable matching. In terminal phases the algorithm finds that $\hat{w}$ cannot be assigned a man she prefers over m without creating a blocking pair, and therefore m is $\hat{w}$'s most preferred stable parter.

Proposition A.1. *Algorithm 1 outputs the women-optimal stable matching.*

Proof. The algorithm terminates as each man can make only a finite number of proposals, and there is at most one terminal phase (that is rolled back) per woman. Consider a phase that begins with a stable $\tilde{\mu}$. Immorlica and Mahdian (2005) show that if the phase ends at step 4(c), the matching $\tilde{\mu}$ at the end of the phase is stable as well. By induction, every $\tilde{\mu}$ is a stable matching. Immorlica and Mahdian (2005) also show that if a phase ends at step 4(d), then $\tilde{\mu}(\hat{w})$ is $\hat{w}$'s most preferred stable man. Any subsequent matching $\tilde{\mu}$ is a stable matching in which $\hat{w}$ is weakly better off, and therefore $\tilde{\mu}$ also matches $\hat{w}$ with her most preferred stable man. Finally, any woman in $\bar{\mathcal{W}}$ is unmatched under the WOSM by the rural hospital theorem (Roth, 1986). Thus the algorithm terminates with $\tilde{\mu}$ being the WOSM that matches all women with their most preferred stable husband. □

We now refine the algorithm to prune repetitions. Since proposals made during terminal phases are rolled back, we can end a phase (and roll back to $\tilde{\mu}$) as soon as we learn that the phase is a terminal phase. During the run of Algorithm 1 each woman in S is matched with her most preferred stable husband. Therefore we can terminate the phase (and roll back) if a woman in S accepts a proposal, as this can happen only in terminal phases. Furthermore, suppose that in a terminal phase the rejection chain includes each woman at most once. We show that every woman in the rejection chain is matched under $\tilde{\mu}$ with her most preferred stable husband, and can therefore be added to S. When the rejection chain includes a woman more than once there are improvement cycles in the chain. We can identify these improvement cycles and implement them as an **Internal Improvement Cycle** (IIC). Specifically, whenever a woman in the chain receives a new proposal we check whether she prefers the proposing man over the best stable partner she has found so far. If she prefers the proposing man, the part of the rejection chain between this proposal and her best stable partner so far forms an improvement cycle. We implement the IIC by recording the stable matching we found in $\tilde{\mu}$ and removing the cycle from the rejection chain. By removing these cycles from the rejection chain we are left with a rejection chain that includes each woman at most once, and when the phase is terminal the entire chain can be added to S. See Algorithm 2, step 4(c) below for a precise definition of IICs.

Applying these modifications to Algorithm 1 gives us Algorithm 2. It keeps track of the women in the current rejection chain as an ordered set $V = (v_1, v_2, .., v_J)$ of women, and adds all of them to S if the phase is terminal. If a woman in S accepts a proposal the phase ends as a terminal phase. As in Algorithm 1, the variable $\tilde{\mu}$ saves the most recent stable matching and μ stores a candidate matching which evolves through the phase. This version of the algorithm also keeps track of $\nu(w)$, the current number of proposals received by woman w, and $R(m)$, the set of women who rejected m so far. These counters were omitted in the version given in Section 3, and are added here to allow us to refer to them later in the proof (The two versions of Algorithm 2 are otherwise identical).

Algorithm 2. MOSM to WOSM

- Input: *A matching market with n men and $n + k$ women.*
- Initialization: *Run the men-proposing deferred acceptance algorithm to get the men-optimal stable matching μ, and set $R(m)$ and $\nu(w)$ accordingly. Set $t = 0$ since no proposals have occurred yet. Initialize S to be the set of women unmatched under μ. Select any $\hat{w} \in \mathcal{W} \backslash S$.*
- New phase:
  1. *Set $\tilde{\mu} \leftarrow \mu$. Set $v_1 \leftarrow \hat{w}$ and $V \leftarrow (\hat{w})$.*
  2. Divorce: *Set $m \leftarrow \mu(\hat{w})$ and have $\hat{w}$ reject m (add $\hat{w}$ to $R(m)$).*

3. Proposal: *Man m proposes to his most preferred woman*[4] *in $\mathcal{W}\backslash R(m)$. Increment $\nu(w)$ and proposal number t by one each.*

4. w's Decision:

(a) If $w \notin V$ and w prefers $\mu(w)$ over m, or if $w \in V$ and w prefers $\tilde{\mu}(w)$ over m, then w rejects m (add w to $R(m)$). Go to step 3.

(b) If $w \notin S \cup V$ and w prefers m over $\mu(w)$, then w rejects her current partner. Set $m' \leftarrow \mu(w)$, $\mu(w) \leftarrow m$. Add w to $R(m')$ and append w to the end of V. Set $m \leftarrow m'$ and go to step 3.

(c) New stable matching: *If $w \in V$ and w prefers m over $\tilde{\mu}(w)$, then we have found a stable matching. If $w = \hat{w} = v_1$, set $\mu(\hat{w}) \leftarrow m$. Select $\hat{w} \in \mathcal{W}\backslash S$ and start a new phase from step 1.*

If $w = v_\ell$ for $\ell > 1$, record her current husband as $m' \leftarrow \mu(w)$. Call the set of all proposals made after the proposal of m' to w an internal improvement cycle *(IIC). Set $\mu(w) \leftarrow m$ and update $\tilde{\mu}$ for the women in the loop by setting $\tilde{\mu}(v_j) \leftarrow \mu(v_j)$ for $j = \ell, \ell+1, ..., J$. Remove $v_\ell, .., v_J$ from V, set the proposer $m \leftarrow m'$, decrement $\nu(w)$, decrement t, and return to step 3, in which m (earlier m') will again propose to w.*

(d) End of terminal phase: *If $w \in S$ and w prefers m over $\mu(w)$, then restore $\mu \leftarrow \tilde{\mu}$ and add all the women in V to S. If $S = \mathcal{W}$, terminate and output $\tilde{\mu}$. Otherwise, select $\hat{w} \in \mathcal{W}\backslash S$ and begin a new phase from step 1.*

In step 4(c) we found a new stable matching. If the rejection chain cycles back to the original woman, we have an improvement phase. If the rejection chain cycles back to a woman v_ℓ in the middle of the chain we implement the IIC (implementing the IIC is equivalent to an improvement phase that begins with v_ℓ). Update the best stable matching $\tilde{\mu}$ for all women in the cycle, and make it the current assignment. Then take m' and make him propose again to v_ℓ, as we changed $\mu(v_\ell)$. Decrement t and $\nu(v_\ell)$ in order not to count this proposal twice.

Proposition A.2. *Algorithm 2 outputs the women-optimal stable matching.*

Proof. Consider the run of Algorithm 2 for a given sequence of selections of $\hat{w}$. We will find a sequence of selections for Algorithm 1 such that Algorithm 2 is equivalent to Algorithm 1.

[4]See footnote 22.

Since Algorithm 1 outputs the WOSM for every sequence of selections, this will prove that Algorithm 2 finds the WOSM as well.

We construct the sequence of selections of $\hat{w}$ for Algorithm 1 by following the run of Algorithm 2 and making the state of Algorithm 2 at the time of proposal t identical to the state of Algorithm 1 in some proposal (note that the two algorithms may have a different proposal count). After the initialization step both algorithms are in an identical state, and we start them both with the same selection of $\hat{w}$. Assume that under our sequence of selections the algorithms are identical up to proposal $t-1$, and consider the t-th proposal made by Algorithm 2. Both algorithms perform the same actions when Algorithm 1 performs steps 4(a), 4(b), and 4(c) when $w = \hat{w}$. When a phase ends we start a new phase for Algorithm 1 by selecting the same $\hat{w}$ as in Algorithm 2. Therefore it remains to consider step 4(c) with $w \neq \hat{w}$ and step 4(d).

Consider the case in which proposal t in Algorithm 2 performs step 4(c) with $w \neq \hat{w}$; that is, the algorithm found an IIC with $w = v_\ell$. We change the sequence of selections of $\hat{w}$ in Algorithm 1 so that v_ℓ is chosen to be $\hat{w}$ just before the current phase. That makes the previous phase of Algorithm 1 an improvement phase for $\hat{w} = v_\ell$ in which the cycle of the IIC is implemented. Continuing to run Algorithm 1 into the current phase will reveal the chain $(v_1, ..., v_{\ell-1})$ and reach the proposal to woman v_ℓ. Algorithm 2 will have the same chain and proposal following the implementation of the IIC, and since now $v_\ell \notin V \cup S$ the two algorithms will have an identical state at the end of the step.

Consider next the case in which proposal t in Algorithm 2 performs step 4(d); that is, the phase is declared to be a terminal phase because a woman $w \in S$ accepted the proposal. First, we show that if we continue to run Algorithm 1 it will uncover a rejection chain that ends in $\bar{\mathcal{W}}$ and declare the phase to be a terminal phase. If $w \in \bar{\mathcal{W}}$ this is immediate. Otherwise, we infer the remainder of the rejection chain as follows. Woman w was added to S in some previous terminal phase. Recursively build the rejection chain $C = \{w_1, w_2, ..., w_q\} \subset S$, where $w_1 = w$, $w_q \in \bar{\mathcal{W}}$, and each w_j was added to S in a terminal phase where w_j rejected her husband m_j, triggering a series of proposals by m_j that resulted in m_j making a proposal that was accepted by w_{j+1}. We next show that following proposal t Algorithm 1 will continue to uncover the rejection chain C.

From the construction of C, at the time of proposal t we have that $\mu(w_j) = m_j$ under Algorithm 2. To see that, recall that each w_j was added to S in a terminal phase which reverted her back to the husband m_j she rejected. Following that terminal phase the assignment of w_j did not change, as the algorithm never changes assignments of women who are already in S. Using the induction assumption, at the step of Algorithm 1 that corresponds to proposal t of Algorithm 2 we have that $\mu(w_j) = m_j$.

Now, consider the continuation of the phase under Algorithm 1. When $w = w_1$ rejects her husband m_1, he will make proposals in the same order as he did in the terminal phase in which w_1 was added to S under Algorithm 2. All women that m_1 prefers over w_2 rejected him back then, and since throughout the algorithm a woman's assignment can only be changed to a more preferred husband, these women will reject him again in the current phase. Therefore, m_1 will end up proposing to w_2. Since at that point $\mu(w_2) = m_2$ the proposal will be accepted by w_2, making m_2 the new proposer. By induction, m_j will make an accepted proposal to w_{j+1}, until a proposal is made to $w_q \in \bar{\mathcal{W}}$. At this step Algorithm 1 declares a terminal phase and rolls back to $\tilde{\mu}$, at which point the phase ends with the same μ and $\tilde{\mu}$ as in Algorithm 2.

The remaining difference between the algorithms is that under Algorithm 2 at step 4(d) all women in V are added to S. Continue to run Algorithm 1 by starting a phase with a selection of a new $\hat{w}$ from V, until we have that $V \subset S$ under Algorithm 1. We have shown above that all these phases will be terminal phases, and will therefore be rolled back. Following these selections, Algorithm 1 will have the same $\tilde{\mu}$ and S as in Algorithm 2 following proposal t.

Therefore, by the end of the run we find a sequence of selections of $\hat{w}$ for Algorithm 1 such that the two algorithms hold identical μ and $\tilde{\mu}$ at the end of Algorithm 2. Thus when Algorithm 2 terminates, Algorithm 1 terminates as well and outputs the same matching. By Proposition A.1 the output of Algorithm 1 is the WOSM. □

The following lemma shows how Algorithm 2 allows us to compare the WOSM and MOSM.

Lemma A.3. *The difference between the sum of mens' rank of wives under WOSM and the sum of mens' rank of wives under MOSM is equal to the number of proposals in improvement phases and IICs during Algorithm 2.*

Proof. Note that at the end of each terminal phase (Step 4(d)) we roll back all proposals made in that phase and return to $\tilde{\mu}$, the matching from the previous phase. Therefore, we can consider only improvement phases and IIC, in which each proposal increases the rank of the proposing man by one. □

A.1 Randomized algorithm

As we are interested in the behavior of Algorithm 2 on a random matching market, we transform the deterministic algorithm on random input into a randomized algorithm, which will be easier to analyze. The randomized, or coin flipping, version of the algorithm does not receive preferences as input, but draws them through the process of the algorithm.[5] This is often called the *principle of deferred decisions*.

The algorithm reads the next woman in the preference of a man in step 3 and whether a woman prefers a man over her current proposal in step 4. Since the algorithm ends a phase immediately when a woman $w \in S$ accepts a proposal, no man applies twice to the same woman during the algorithm, and therefore the algorithm never reads previously revealed preferences.[6] In step 3 the randomized algorithm selects the woman w uniformly at random from $\mathcal{W} \backslash R(m)$. In step 4 the probability that w prefers m over her current match can be given directly from $\nu(w)$ for $w \notin S$ or bounded for $w \in S$.[7] Table 1 describes the probabilities of the possible decisions of w. Note that the event in Step 4(a) is the complement of the union of the events in Table 1.

B Proof

We will prove the following quantitative version of the main theorem from the paper:

[5] The initialization step of the randomized version of Algorithm 2 uses the randomized version of Algorithm 1.

[6] There is an apparent exception to this in the case of an IIC, where we chose to describe the algorithm with a rollback of the proposal by man m' to v_ℓ that occurred just before the IIC, followed by remaking that proposal (which we know will be rejected by v_ℓ); cf. Step 4(c) of Algorithm 2. This was merely to simplify the exposition; in our stochastic analysis we proceed by directly considering the proposal by m' to his next most preferred woman after v_ℓ.

[7] The probability that a woman $w \in S$ accepts a man m can be calculated from the number of proposals she received during improvement phases or MPDA and the number of proposals she received during terminal phases. Since the bound on the acceptance probability we calculate from $\nu(w)$ is sufficient for our analysis we omit the additional counters from the algorithm.

Step	**Event**	**Probability**
4(b)	$w \notin S \cup \{\hat{w}\}$ prefers m over $\mu(w)$	$\frac{1}{\nu(w)+1}$
4(c)	$\hat{w}$ prefers m over $\tilde{\mu}(\hat{w})$	$\frac{1}{\nu(\hat{w})+1}$
4(d)	$w \in S \backslash \bar{\mathcal{W}}$ prefers m over $\mu(w)$	at least $\frac{1}{\nu(w)+1}$

Table 1: Probabilities in a run of Algorithm 2 on a random matching market.

Theorem 1. *Fix any $\epsilon > 0$. Consider a sequence of random matching markets, indexed by n, with n men and $n+k$ women, for arbitrary $1 \le k = k(n)$. There exists $n_0 < \infty$ such that for all $n > n_0$, with probability at least $1 - \exp\{-(\log n)^{0.4}\}$, we have*

(i) In every stable matching μ:

$$R_{\text{MEN}}(\mu) \le (1+\epsilon)\big((n+k)/n\big)\log\big((n+k)/k\big)$$

$$R_{\text{WOMEN}}(\mu) \ge n\big/\big[1 + (1+\epsilon)\big((n+k)/n\big)\log\big((n+k)/k\big)\big]\,.$$

(ii) Less than $n/(\log n)^{0.5}$ men, and less than $n/(\log n)^{0.5}$ women have multiple stable partners.

(iii) The men are almost as well off under the WOSM as under the MOSM:

$$\frac{R_{\text{MEN}}(\text{WOSM})}{R_{\text{MEN}}(\text{MOSM})} \le 1 + (\log n)^{-0.4}\,.$$

(iv) The women are almost as badly off under the WOSM as under the MOSM:

$$\frac{R_{\text{WOMEN}}(\text{WOSM})}{R_{\text{WOMEN}}(\text{MOSM})} \ge 1 - (\log n)^{-0.4}\,.$$

Remark 1. *In our proof of Theorem 1 (ii), we actually bound the number of different stable partners. We show that the sum over all men (or women) of the number of different stable partners of each man is no more than $n + n/\sqrt{\log n}$. Thus, in addition to the bound stated in Theorem 1 (ii), we rule out the possibility that there are a few agents who have a large number of different stable partners.*

Definition B.1. *Given a sequence of events* $\{\mathcal{E}_n\}$*, we say that this sequence occurs* with very high probability (wvhp) *if*

$$\lim_{n\to\infty} \frac{1-\mathbb{P}(\mathcal{E}_n)}{\exp\{-(\log n)^{0.4}\}} = 0\,.$$

Clearly, it suffices to show that (i)-(iv) in Theorem 1 hold wvhp.

To prove Theorem 1, we analyze the number of proposals in Algorithm 1 followed by Algorithm 2, which will provide us the average rank of wives in the women-optimal stable match. We partition the run of Algorithm 2 leading to the WOSM into three parts (Parts II through IV below).

1. **Part I is the run of DA (Algorithm 1)**, which by an analysis similar to that in (Pittel (1989)), takes no more than $3n\log(n/k)$ proposals wvhp.

2. **Part II are the proposals in Algorithm 2 that take place before the end of first terminal phase.** We show that wvhp,

   - Part II takes no more than $(n/k)(\log n)^{0.45} \le n(\log(n/k))^{0.45}$ proposals.
   - When part II ends the set S contains at least $n^{(1-\varepsilon)/2}$ elements.

3. **Part III are the proposals in Algorithm 2 after Part II that take place until $|S| \ge n^{0.7}$. Thus, this part ends at the end of a terminal phase when $|S|$ exceeds $n^{0.7}$ for the first time.** We show that, wvhp, part III requires $O(n^{0.47})$ phases, and $o(n)$ proposals.[8]

4. **Finally, Part IV includes the remaining proposals from the end of part III until Algorithm 2 terminates or $50n\log n$ total proposals have occurred (including proposals made in Parts I and II), whichever occurs earlier.** Because the set S is large, most phases are terminal phases containing no IICs, and most acceptances lead to eventual inclusion in S. We show that, wvhp, part IV ends with termination of the algorithm, and that the number of proposals in improvement phases and IICs is $o(n)$. But the increase in sum of men's rank of wives from the MOSM to the

[8]For any two functions $f:\mathbb{N}\to\mathbb{R}$ and $g:\mathbb{N}\to\mathbb{R}_{>0}$ we write $f=o(g)$ if $\lim_{n\to\infty}\frac{f(n)}{g(n)}=0$ and $f=O(g)$ if there exist a constant a such that $f(n)\le ag(n)$ for sufficiently large n. We write $f(n)=\Theta(g(n))$ if there exist constants $a\le b, n_0$ such that $ag(n)\le f(n)\le bg(n)$ for all $n>n_0$.

WOSM is exactly the number of proposals in improvement phases and IICs, yielding the result.

The definition of Part III based on when $|S|$ exceeds $n^{0.7}$ is for technical reasons, with the exponent of 0.7 being a choice for which our analysis goes through. Throughout this section, we consider the preferences on both sides of the market as being revealed sequentially as the algorithm proceeds, as discussed in Appendix A.1.

Lemma B.2. *Consider a man m, who is proposing at step 3 of Algorithm 2. Consider a subset of women $\mathcal{A} \subseteq \mathcal{W} \backslash R(m)$. Let $\nu(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \sum_{\tilde{w} \in \mathcal{A}} \nu(\tilde{w})$ be the average number of proposals received by women in $\mathcal{A}$. The man m proposes to some woman w in the current step. Conditional on $w \in \mathcal{A}$ and all preferences revealed so far, the probability that m is the most preferred man who proposed w so far, is at least $\frac{1}{\nu(\mathcal{A})+1}$.*

Proof. For any woman $\tilde{w} \notin R(m)$ the probability that m is the most preferred man who applied to w so far is $\frac{1}{\nu(\tilde{w})+1}$. Conditional on $w \in \mathcal{A}$, the probability that m is the most preferred man who applied to w so far is at least

$$\frac{1}{|\mathcal{A}|} \sum_{\tilde{w} \in \mathcal{A}} \frac{1}{\nu(\tilde{w})+1} \geq \frac{1}{\nu(\mathcal{A})+1}$$

by Jensen's inequality. □

The following lemma will be convenient and its proof is trivial:

Lemma B.3. *If all men have lists of length $|\mathcal{W}|$ and $|\mathcal{W}| > |\mathcal{M}|$, then no man ever reaches the end of his list in Algorithm 1 or Algorithm 2.*

B.1 Part I

For the analysis in this section we consider the following equivalent version of men proposing deferred acceptance:

Algorithm 2. *Index the men $\mathcal{M}$. Initialize $S_{\mathcal{M}} = \phi$, $\bar{\mathcal{W}} = \mathcal{W}$, $\nu(w) = 0$ $\forall w \in \mathcal{W}$ and $R(m) = \phi$ $\forall m \in \mathcal{M}$.*

1. *If $\mathcal{M} \backslash S_{\mathcal{M}}$ is empty then terminate. Else, let m be the man with the smallest index in $\mathcal{M} \backslash S_{\mathcal{M}}$. Add m to $S_{\mathcal{M}}$.*

2. *Man m proposes to his most preferred woman w whom he has not yet applied to (increment $\nu(w)$). If he is at the end of his list, go to Step 1.*

3. *Decision of w:*

   - *If $w \in \bar{\mathcal{W}}$, i.e, w is unmatched then she accepts m, remove w from $\bar{\mathcal{W}}$. Go to Step 1.*
   - *If w is currently matched, she accepts the better of her current match and m and rejects the other. Set m to be the rejected man, add w to $R(m)$ and continue at Step 2.*

Note that the output of Algorithm 2 is the same as the output of Algorithm 1, i.e., it is the man optimal stable match (we have just reordered the proposals). The output of Algorithm 2 is given as an input to Algorithm 2. Again, we think of preferences as being revealed as the algorithm proceeds, with the women only revealing preferences among the set of men who have proposed them so far.

The next lemma establishes upper bounds on the average and maximum men's rank of wives and a lower bound on the women's average rank of husbands. The upper bound for the worst possible men's rank of wifes is due to Pittel (1989) who obtained this bound in a balanced market (by adding more women to the market men are only becoming better off).

Lemma B.4. *Fix any $\epsilon > 0$. Let μ be the men-optimal stable matching. The following hold wvhp:*

(i) the men's average rank of wives in μ is at most $(1+\epsilon)\left(\frac{n+k}{n}\right)\log\left(\frac{n+k}{k}\right)$ and is at least $(1-\epsilon)\left(\frac{n+k}{n}\right)\log\left(\frac{n+k}{k}\right)$,

(ii) $\max_{m\in\mathcal{M}} \mathrm{Rank}_m(\mu(m)) \le 3(\log n)^2$,

(iii) the women's average rank of husbands in μ is at least $n\big/\left[1+(1+\epsilon)\left(\frac{n+k}{n}\right)\log\left(\frac{n+k}{k}\right)\right]$.

Proof. We first prove the upper bound in (i), then (ii), then the lower bound in (i), and finally (iii).

Tracking Algorithm 2 like in Pittel (1989), we claim that, wvhp, the sum of the men's rank of wives is at most $(1+\epsilon)(n+k)\log((n+k)/k)$ for small enough $\epsilon > 0$. This claim immediately implies the stated bound (i) on the men's average rank of wives. To prove the claim, we use the fact that the number of proposals is stochastically dominated by the number of draws in the coupon collector's problem, when n distinct coupons must be drawn from $n+k$ coupons. This latter quantity is a sum of Geometric$((n+k-i+1)/(n+k))$ random variables for $i = 1, 2, \ldots, n$. The mean is

$$\begin{aligned}\sum_{i=1}^{n} \frac{n+k}{n+k-i+1} &= (n+k)\Big(\frac{1}{k+1} + \frac{1}{k+2} + \ldots + \frac{1}{n+k}\Big) \\ &= (n+k)\Big(\log((n+k)/k) + O(1/k)\Big) \\ &= (n+k)\log((n+k)/k)\Big(1 + O(1/(k\log((n+k)/k)))\Big)\,.\end{aligned}$$

A short analytical exercise[9] shows that $1/(k\log((n+k)/k))$ is monotone decreasing in k, and is thus maximized at $k = 1$. It follows that $1/(k\log((n+k)/k)) \leq 1/\log(n+1) \leq 1/(\log n)$, which establishes that the error term $O(1/(k\log((n+k)/k))) = O(1/\log n)$ vanishes in the limit. Now routine arguments (e.g., Durrett (2010)) can be used to show that, in fact, this sum exceeds $(1+\epsilon)(n+k)\log((n+k)/k)$ with probability $\exp(-\Theta(n))$. This establishes the upper bound in (i).

Intuitively, the upper bound in (ii) should hold since it holds for a balanced market (see Pittel (1989)), and adding more women should presumably only make the bound tighter. We show that this is indeed the case. Our proof works as follows: we first show that for any man m, the probability that $\mathrm{Rank}_m(\mu(m)) > 3(\log n)^2$ is bounded above by $1/n^{1.2}$. We then use a union bound over the men to establish (ii).

Fix a man m and consider Algorithm 2, where one additional man is processed at a time. From McVitie and Wilson (1971), we know that the final outcome is the MOSM, and this does not depend on the order in which the men are processed. Therefore, we can assume that m is processed last.

Let $t = 1, 2, \ldots$ be the index of the proposals. From the upper bound in (i) proved above,

[9]Define $f(k) = k\log((n+k)/k)$, where we think of n as fixed, and $k \in (0,\infty)$ as varying. We obtain $f'(k) = \log(1+n/k) - 1 + 1/(1+n/k)$. It suffices to show that $f'(k) > 0$ for all $k > 0$. To show this, we define $g : (0,\infty) \to \mathbb{R}$ as $g(x) = \log(1+x) - 1 + 1/(1+x)$. Now $\lim_{x\to 0} g(x) = 0$ and $g'(x) = x/(1+x)^2 > 0$ for all $x > 0$, leading to $g(x) > 0$ for all $x > 0$. Hence, $f'(k) > 0$ for all $k > 0$, as required.

we know that with probability $1 - \exp(-\Theta(n))$, the MOSM is found before

$$\begin{aligned} t = T_* &= (1+\epsilon)(n+k)\log((n+k)/k) \\ &\leq (1+\epsilon)(n\log(1+n/k) + n\lim_{k\to\infty}(k/n)\log((n+k)/k)) \\ &\leq n(1+\epsilon)(\log(1+n/k)+1) \\ &\leq 1.1 n\log n\,, \end{aligned}$$

for small enough ϵ and large enough n. Let $\hat{\mathcal{E}}$ be the event that this bound holds. Then we know that

$$\mathbb{P}(\hat{\mathcal{E}}^c) \leq \exp(-\Theta(n))\,.$$

We track[10] Algorithm 2 until it terminates or index t exceeds T_*. If the index t exceeds T_*, i.e., we observe $\hat{\mathcal{E}}^c$, we declare failure and stop. Hence, we can use $t \leq T_*$ in what follows. Consider the i-th proposal by man m, and suppose $i \leq 3(\log n)^2$. Then there are at least $n + k - 3(\log n)^2$ women that m has not yet proposed to, and these women have together received no more than T_* proposals in total so far. Using Lemma B.2, the probability of the proposal being accepted is at least

$$1/(T_*/(n+k-3(\log n)^2)+1) \geq 1/(1.1n\log n/(n-3(\log n)^2)+1) \geq 1/(1.2\log n)$$

for large enough n. If the proposal is accepted by a woman $\hat{w}$, each subsequent proposal in the chain is, independently, at least as likely to go to an unmatched woman as it is to go to $\hat{w}$. Hence, the subsequent rejection chain has a probability at least 1/2 of terminating in an unmatched woman before there is another proposal to woman $\hat{w}$. As such, the probability that the i-th proposal will be the last proposal made by m is at least $(1/2)\cdot 1/(1.2\log n) = 1/(2.4\log n)$. It follows that the probability that m has to make more than $3(\log n)^2$ proposals before the MOSM is reached or failure occurs is no more than

$$\left(1 - \frac{1}{2.4\log n}\right)^{3(\log n)^2} \leq \left(\exp\left\{-\frac{1}{2.4\log n}\right\}\right)^{3(\log n)^2} \leq \exp(-1.25\log n) = 1/n^{1.25}\,.$$

Combined with the probability of failure, using a union bound, the overall probability that man m makes more than $3(\log n)^2$ proposals is bounded above by $1/n^{1.25} + \mathbb{P}(\hat{\mathcal{E}}^c) \leq 1/n^{1.25} +$

[10] As usual, we reveal information about each preference list as it is needed (cf. Appendix A.1): for men we reveal the next entry in the preference list just before a new proposal; for women, we reveal whether a new proposal is the best one so far, only when the proposal is made.

$\exp(-\Theta(n)) \le 1/n^{1.2}$, for large enough n. Since the same bound applies to any man, we can use a union bound to find that

$$\mathbb{P}(\text{Any man makes more than } 3(\log n)^2 \text{ proposals}) \le n \cdot 1/n^{1.2} = 1/n^{0.2}\,.$$

We conclude that wvhp, no man makes more than $3(\log n)^2$ proposals, establishing (ii).

We now establish the lower bound in (i), i.e., that the sum of men's rank of wives is at least $(1-\epsilon)(n+k)\log((n+k)/k)$. The proof is similar to that of the upper bound in (i). From (ii), we have that wvhp, no man makes more than $3(\log n)^2$ proposals. It follows that for each proposal that occurs during the search for the i-th unmatched woman, the probability that an unmatched woman is found is at most

$$p_i = \frac{n+k-i+1}{n+k-\min(3(\log n)^2, i-1)}\,.$$

It follows that the number of proposals needed to find the i-th woman stochastically dominates Geometric(p_i), conditional on what has happened. It follows that the mean total number of proposals is at least

$$\begin{aligned}\sum_{i=1}^{n} \frac{n+k-3(\log n)^2}{n+k-i+1} &= \big(n+k-3(\log n)^2\big)\Big(\frac{1}{k+1}+\frac{1}{k+2}+\ldots+\frac{1}{n+k}\Big)\\ &= (n+k)\log((n+k)/k)(1+O(1/(\log n))\,,\end{aligned}$$

where we bound the error term as above and using $3(\log n)^2/(n+k) = O((\log n)^2/n) = O(1/(\log n))$. Again, routine arguments (e.g., Durrett (2010)) can be used to show that, in fact, a sum of independent Geometric(p_i) random variables for $i = 1, 2, \ldots, n$ is less than $(1-\epsilon)(n+k)\log((n+k)/k)$ with probability $\exp(-\Theta(n))$. This establishes the lower bound on the men's average rank of wives.

Now consider the women's rank of husbands. For a woman w, who has received $\nu(w)$ proposals in Part I, the rank of her husband is a random variable that depends only on $\nu(w)$, and not anything else revealed so far. We have

$$\mathbb{E}[\mathrm{Rank}_w(\mu(w))] = \frac{n+1}{\nu(w)+1}\,.$$

Define

$$M \equiv \mathbb{E}\left[\sum_{w\in\mathcal{W}\backslash\bar{\mathcal{W}}} \mathrm{Rank}_w(\mu(w))\right] = (n+1)\sum_{w\in\mathcal{W}\backslash\bar{\mathcal{W}}} \frac{1}{\nu(w)+1}\,.$$

Using Azuma's inequality (see Durrett (2010)), we have

$$\mathbb{P}\Big((1/n) \sum_{w \in \mathcal{W} \setminus \bar{\mathcal{W}}} \mathrm{Rank}_w(\mu(w)) \leq \frac{M}{n} - \Delta\Big) \leq \exp\left\{-\frac{n\Delta^2}{2n^2}\right\},$$

since $\mathrm{Rank}_w(\mu(w)) \in [0, n]$. Plugging in $\Delta = n^{3/4}$ yields

$$\mathbb{P}\Big((1/n) \sum_{w \in \mathcal{W} \setminus \bar{\mathcal{W}}} \mathrm{Rank}_w(\mu(w)) \leq \frac{M}{n} - n^{3/4}\Big) \leq \exp\left\{-\frac{n^{1/2}}{2}\right\}. \tag{1}$$

Using Jensen's inequality in the definition of M, we have

$$\begin{aligned} M &\geq (n+1)n \cdot \frac{1}{1 + (1/n)\sum_{w \in \mathcal{W} \setminus \bar{\mathcal{W}}} \nu(w)} \\ &\geq \frac{n^2}{1 + (1+\epsilon/2)\left(\frac{n+k}{n}\right)\log\left(\frac{n+k}{k}\right)} \quad \text{wvhp}, \end{aligned}$$

where we used (i) with ϵ replaced by $\epsilon/2$, i.e., $(1/n)\sum_{w \in \mathcal{W} \setminus \bar{\mathcal{W}}} \nu(w) \leq (1+\epsilon/2)\left(\frac{n+k}{n}\right)\log\left(\frac{n+k}{k}\right)$ wvhp. Using $\left(\frac{n+k}{n}\right)\log\left(\frac{n+k}{k}\right) \leq 1.1\log n$ (see the bound on T_* in the proof of (ii)), i.e., the $n^{3/4}$ is negligible in comparison to M/n, we can deduce that

$$\frac{M}{n} - n^{3/4} \geq \frac{n}{1 + (1+\epsilon)\left(\frac{n+k}{n}\right)\log\left(\frac{n+k}{k}\right)} \quad \text{wvhp}, \tag{2}$$

for large enough n (notice that we used the $\epsilon/2$ slack).

Combining Eqs. (1) and (2), we obtain (iii). □

Lemma B.5. *Suppose $k \leq n^{0.1}$. Then, wvhp, there are fewer than $n^{0.99}$ women who each receive less than $(1/2)\log n$ proposals.*

Proof. Now consider a woman w'. For each proposal, it goes to w' with probability at least $1/(n+k)$, unless the proposing man has already proposed w'. Suppose w' receives fewer than $(\log n)/2 \leq (2/3)(1-2\epsilon)\log(n+k)/k$ proposals, where, for instance, we can define $\epsilon = 0.01$. Since, wvhp, each man makes at most $3(\log n)^2$ proposals (Lemma B.4(ii)), wvhp there are at most $(3/2)(\log n)^3$ proposals by men who have already proposed w'. Using Lemma B.4(i), wvhp, there are at least $(1-\epsilon)(n+k)\log((n+k)/k)$ proposals in total. It follows that, wvhp, there are at least $(1-\epsilon)(n+k)\log((n+k)/k) - (3/2)(\log n)^3 \geq (1-2\epsilon)(n+k)\log((n+k)/k)$ proposals by men who have not yet proposed w'. Using Fact E.1 (i), the probability that fewer than $(2/3)(1-2\epsilon)\log((n+k)/k)$ of these proposals go to w' is

$$\begin{aligned}&\mathbb{P}\big(\mathrm{Binomial}((1-2\epsilon)(n+k)\log((n+k)/k), 1/(n+k)) < (2/3)(1-2\epsilon)\log(n+k)/k)\big)\\ &\le 2\exp\{-(1-2\epsilon)\log((n+k)/k)/27\} \le ((n+k)/k)^{-1/28} \le n^{-0.02}\,,\end{aligned}$$

for $k \le n^{0.1}$. It follows that the expected number of women who receive fewer than $(1/2)\log n$ proposals is no more than $(n+k)n^{-0.02} \le 2n^{0.98}$. By Markov's inequality, the number of women who receive fewer than $(1/2)\log n$ proposals is no more than $n^{0.99}$, with probability at least $1 - 2n^{0.98}/n^{0.99} = 1 - o(\exp\{-(\log n)^{0.4}\})$. $\square$

B.2 Part II

Lemma B.7 below shows that by the end of Part II, the number of proposals by each man is "small", the number of proposals received by each woman is "small", and that the set S is "large". Since S will be large at the end of this part, Part III will terminate "quickly". The next lemma says that wvhp, there will not be too many proposals in Part II (this bound will be assumed in the proof of Lemma B.7).

Lemma B.6. *Part II completes in no more than $\frac{n+k}{k}(\log n)^{0.45} \le (n+1)(\log n)^{0.45}$ proposals wvhp.*

Proof. For each proposal (Step 3) in Part II, the probability of Step 4(d), which will end Part II, is the probability that the man m proposes to an unmatched woman $\frac{k}{|\mathcal{W}\backslash R(m)|} \ge \frac{k}{n+k}$. Therefore the probability that the number of proposals in part II exceeds $((n+k)/k)(\log n)^{0.45}$ is at most $\left(1 - \frac{k}{n+k}\right)^{((n+k)/k)(\log n)^{0.45}} \le \exp(-(\log n)^{0.45}) = o(\exp\{-(\log n)^{0.4}\})$, leading to the first bound. Noticing that $(n+k)/k = 1 + n/k \le 1 + n$, we obtain the bound of $(n+1)(\log n)^{0.45}$. $\square$

Lemma B.7. *Fix any $\varepsilon > 0$. At the end of Part II, the following hold wvhp:*

(i) No man has applied to a lot of women:

$$\max_{m\in\mathcal{M}} |R(m)| < n^{\varepsilon}. \tag{3}$$

(ii) The set S is large: $|S| \ge n^{(1-\varepsilon)/2}$.

(iii) No woman received many proposals

$$\max_{w \in \mathcal{W}} \nu(w) < n^{\varepsilon}. \tag{4}$$

Proof. Using Lemma B.4, we know that wvhp, there have been no more than $3n \log(n/k)$ proposals and no man has proposed more than $3(\log n)^2$ women in Part I. Assume that these two conditions hold for the rest of the proof.

We begin with (i). We say that a man m starts a **run of proposals** when m is rejected by a woman at step 4(b) or is divorced from $\hat{w}$ at step 2. We say that a *failure* occurs if a man starts more than $(\log n)^2$ runs or if the length of any run exceeds $(\log n)^3$ proposals. We associate a failure with a particular proposal t, when for the first time, a man starts his $(\log n)^2 + 1$-th run, or the proposal is the $(\log n)^3 + 1$-th proposal in the current run.

Consider the number of runs of a given man m. Man m starts at most one run at step 2. The other runs start when the proposing man $m' \neq m$ proposes to the women m is currently matched with and m' is accepted. At any proposal the probability that m' proposes to any particular woman is no more than the probability that he proposes to $\bar{\mathcal{W}}$. Now if the latter happens, Part II ends. Therefore, it follows that the number of runs man m has in part II is stochastically dominated[11] by $1 + \text{Geometric}(1/2)$. Hence, the probability that a man has more than $(\log n)^2$ runs is bounded by $\left(\frac{1}{2}\right)^{(\log n)^2 - 1} \leq 1/n^2$, showing that man m has fewer than $(\log n)^2$ runs in Part II with probability at least $1 - 1/n^2$. It follows from a union bound over all men $m \in \mathcal{M}$ that wvhp failure due to number of runs does not occur.

Assume failure did not occur before or at the beginning of a run of man m. The number of proposals man m accumulates until either the run ends or a failure occurs is bounded by

$$(\log n)^2 \cdot (\log n)^3 \leq n/2$$

for sufficiently large n. In each proposal in the run before failure, man m proposes to a uniformly random woman in $\mathcal{W} \backslash R(m)$. Since there were at most $4n \log n$ proposals so far, we have that

$$\nu\left(\mathcal{W} \backslash R(m)\right) \leq \frac{4n \log n}{n/2} = 8 \log n\,.$$

[11] A real values random variable with cumulative distribution F_1 is said to be stochastically dominated by another r.v. with cumulative distribution F_2 if $F_2(x) \leq F_1(x)$ for all $x \in \mathbb{R}$.

From Lemma B.2, we have that the probability of acceptance at each proposal is at least $\frac{1}{\nu(\mathcal{W}\backslash R(m))+1} \geq \frac{1}{8\log n+1}$. Therefore the probability of man m making $(\log n)^3$ proposals without being accepted is bounded by[12]

$$\left(\frac{1}{8\log n+1}\right)^{(\log n)^3} \leq \frac{1}{n^3}\,.$$

Thus, the run has length no more than $(\log n)^3$ with probability at least $1 - 1/n^3$. Now the number of runs is bounded by n^2, so we conclude that wvhp failure due to number of runs does not occur. Finally, assuming no failure,

$$|R(m)| \leq (\log n)^2 \cdot (\log n)^3 < n^{\varepsilon}$$

establishing (i).

We now prove (ii). If $k \geq n^{(1-\varepsilon)/2}$, the set S is already large enough at the beginning of Part II, and there is nothing to prove. Suppose $k < n^{(1-\varepsilon)/2}$. Consider the evolution of $|V|$ during Part II. We first provide some intuition. Part II contains about $n/k = \omega(n^{1/2})$ proposals before it ends. We start with $|V| = 0$, and $|V|$ initially builds up without any new stable matches found. We can estimate the size of $|V|$ when Step 4(c) (new stable match) occurs as follows[13]: Suppose we reach $|V| \sim N$. Consider the next accepted proposal. Ignoring factors of $\log n$, the probability that the woman who accepts is in V is $\sim |V|/n \sim N/n$. Thus, for one of the next N accepted proposals to include a woman in V, we need $N \cdot N/n \sim 1$, i.e., $N \sim \sqrt{n}$. Thus, when $|V|$ reaches a size of about $\sqrt{n}$, then an IIC forms over the next $\sim \sqrt{n}$ proposals, reducing the size of $|V|$. This occurs repeatedly, with $|V|$ converging to an 'equilibrium' distribution with mean of order $\sqrt{n}$, and this distribution has a light tail. Thus, when the phase ends, we expect $|V| \sim \sqrt{n}$.

We now formalize this intuition. Whenever $|V| < n^{(1-\varepsilon)/2}$, for the next proposal, the probability that:

- The proposal goes to a woman in V is less than $2/n^{(1+\varepsilon)/2}$. Such a proposal is necessary to creating an IIC.
- The proposal goes to a woman in $S = \bar{\mathcal{W}}$, is at most $2k/(n+k) \leq 2k/n \leq 2/(n^{(1+\varepsilon)/2})$. Such a proposal would terminate the phase.

[12] Again this inequality holds for large enough n. We omit the explicit mention of the condition "for large enough n" when such inequalities appear subsequently in this section.

[13] This analysis is analogous to that of the birthday paradox.

- The proposal goes to a woman in $\mathcal{W}\backslash(S \cup V)$, who accepts it, is at least $1/(5\log n)$, using the fact that there have been no more than $4n\log n$ proposals so far and Lemma B.2.

Suppose we start with any $|V| < n^{(1-\varepsilon)/2}$, for instance we have $|V| = 0$ at the start of the phase, we claim that with probability at least $1 - 3k/\sqrt{n}$, we reach $|V| = n^{(1-\varepsilon)/2}$ (call this an 'escape') before there is a proposal to S and before $\sqrt{n}$ proposals occur. We prove this claim as follows: There is a proposal to S among the next $\sqrt{n}$ proposals with probability no more than $2k/\sqrt{n}$. Suppose that $|V|$ stays less than $n^{(1-\varepsilon)/2}$. Then there are $n^{\varepsilon/4}$ or more proposals to women in $|V|$ among $\sqrt{n}$ total proposals with probability no more than $2^{-n^{\varepsilon/4}}$ using Fact E.1 (ii) on $\mathrm{Binomial}(\sqrt{\mathrm{n}}, 2/\mathrm{n}^{(1+\varepsilon)/2})$. Also, the probability that there are less than $n^{1/2-\varepsilon/8}$ proposals accepted by women in $\mathcal{W}\backslash(V \cup S)$ (these women are added to V) is at most $2^{-n^{1/2-\varepsilon/8}}$, using Fact E.1 (ii) on $\mathrm{Binomial}(\sqrt{\mathrm{n}}, 1/(5\log \mathrm{n}))$, since at each proposal, such a woman is added with probability at least $(1/(5\log n))$. But if there are

- less than $n^{\varepsilon/4}$ proposals to V, each such proposal reducing $|V|$ by at most $n^{1/2-\varepsilon/2}$,
- no proposal to S, and
- at least $n^{1/2-\varepsilon/8}$ women added to V,

then we must reach $|V| = n^{(1-\varepsilon)/2}$. Thus, the overall probability of *not* reaching $|V| = n^{(1-\varepsilon)/2}$ before there is a proposal to S and before $\sqrt{n}$ proposals occur is at most $2k/\sqrt{n} + 2^{-n^{\varepsilon/2}} + 2^{-n^{1/2-\varepsilon/4}} \leq 3k/\sqrt{n}$. In particular, the probability of a failed escape is at most $3k/\sqrt{n}$.

We now bound the number of times $|V|$ reduces from a value larger than $n^{(1-\varepsilon)/2}$ to a value smaller than $n^{(1-\varepsilon)/2}$. Suppose $|V| \geq n^{(1-\varepsilon)/2}$. The probability that a proposal goes to S is at least $k/(n+k) \geq k/(2n)$. The probability that a proposal goes to one of the first $n^{(1-\varepsilon)/2}$ women in $|V|$ is at most $2n^{(1-\varepsilon)/2}/n \leq 2/n^{(1+\varepsilon)/2}$. Thus, the number of times the latter occurs is stochastically dominated by $\mathrm{Geometric}(k/(4n^{(1-\varepsilon)/2})) - 1$. Thus, the total number of escapes needed to ensure $|V| \geq n^{(1-\varepsilon)/2}$, including the one at the start of the phase, is stochastically dominated by $\mathrm{Geometric}(k/(4n^{(1-\varepsilon)/2}))$, which exceeds $n^{1/2-\varepsilon/4}/k$ with probability at most

$$\left(1 - k/(4n^{(1-\varepsilon)/2})\right)^{n^{1/2-\varepsilon/4}/k} \leq \exp(-n^{\varepsilon/4}/4)\,.$$

Assuming no more than $n^{1/2-\varepsilon/4}/k$ escapes are needed, one of these escapes fails with probability at most $(n^{1/2-\varepsilon/4}/k)\cdot(3k/\sqrt{n}) = 3n^{-\varepsilon/4}$. Thus, the overall probability of $|V| < n^{(1-\varepsilon)/2}$ when the phase ends is bounded by $\exp(-n^{\varepsilon/4}/4)+n^{-\varepsilon/4} = o(\exp\{-(\log n)^{0.4}\})$. Thus, wvhp, $|V| \geq n^{(1-\varepsilon)/2}$ for all phases in Part II, including the terminal phase. This establishes (ii).

Finally, we establish (iii). Again we assume in our proof that Parts I and II end in no more than $4n\log n$ proposals in total, and that (i) holds (if not, we abandon our attempt to establish (iii), but this does not happen wvhp). Fix a woman w. For each proposal, the probability that *she* receives the proposal is no more than $2/n$, using (i). Thus, the total number of proposals she receives is no more than Binomial$(2\mathrm{n}\log\mathrm{n}, 2/\mathrm{n})$ which is less than n^{ε}, except with probability $2^{-n^{\varepsilon}/(4\log n)}$ by Chernoff bound (see Fact E.1 in Appendix E). Union bound over the women gives us that (iii) holds wvhp. □

Lemma B.8. *Wvhp, the number of accepted proposals in Part II is no more than $n/(2\sqrt{\log n})$ and the improvement in sum of women's rank of husbands during Part II is no more than $n^2/(2(\log n)^{3/2})$.*

Proof. If $k \geq n^{0.1}$, then we already know that wvhp the number of proposals is no more than $n^{0.95}$ using Lemma B.6.

If $k < n^{0.1}$, then we know from Lemma B.5, that, wvhp, fewer than $n^{0.99}$ women each received fewer than $\log n/2$ proposals in Part I. Further, from Lemma B.7, wvhp, no man has proposed to more than n^{ε} women in Parts I and II. It follows that for each proposal in Part II, it goes to a woman who has already received $\log n/2$ or more proposals with probability at least $1-n^{-0.01}/2$. Hence, the probability that the proposal is accepted is at most $2.5/\log n$. But the total number of proposals in Part II, wvhp, is less than $(n+1)(\log n)^{0.45}$ from Lemma B.6. It follows using Fact E.1 that, wvhp, fewer than $3(n+1)/(\log n)^{0.55} \leq n/(10\sqrt{\log n})$ proposals are accepted in Part II.

We now bound the improvement in the sum of women's rank of husbands. Using Markov's inequality, there are, wvhp, at most $n^{0.995}$ proposals to women who have received fewer than $\log n/2$ proposals so far. The maximum possible improvement in rank from these proposals is $(n+k)n^{0.995} \leq 2n^{1.995}$. The number of proposals accepted by women who have received at least $\log n/2$ proposals so far is, wvhp, at most $n/(10\sqrt{\log n})$, as we showed above. For such a proposal accepted by a woman w' who has received $\nu(w') \geq (\log n)/2$ previous proposals, the expected improvement in rank is

$$\frac{n-\nu(w')}{\nu(w')+1} - \frac{n-\nu(w')-1}{\nu(w')+2} \le n/((\log n)/2)^2 \le 4n/\log n\,,$$

since $\log n \ge 1$. Further the improvement in rank is in the interval $[1, n-1]$. Thus, the total improvement in rank is stochastically dominated by a sum of independent X_i, for $i = 1, 2, \ldots, n/(10\sqrt{\log n})$, with $\mathbb{E}[X_i] \le 4n/\log n$ and $1 \le X_i \le n-1$. It follows using Azuma's inequality (see, e.g., Durrett (2010)) that this sum exceeds $n/(10\sqrt{\log n}) \cdot 4n/\log n + n^{1.6} \le n^2/(2.1(\log n)^{3/2})$ with probability at most

$$2\exp\left\{-\frac{(n^{1.6})^2}{2 \cdot n/(10\sqrt{\log n}) \cdot n^2}\right\} = \exp\{-n^{0.1}\} = o(\exp\{-(\log n)^{0.4}\})\,.$$

Thus, the total improvement in the sum of women's rank of husbands is, wvhp, no more than $2n^{1.995} + n^2/(2.1(\log n)^{3/2}) \le n^2/(2(\log n)^{3/2})$. □

B.3 Part III

Let S_{II} be the set S at the end of part II.

The next lemma provides upper bounds (that are achieved wvhp) on the number of proposals each man makes and the number of proposals each woman receives throughout Parts III and IV.

Let $\mathcal{E}_t$ be the event that until proposal t, no man has applied to more than $n^{0.6}$ women in total or to more than $n^{3\varepsilon}$ women in S_{II}, and no woman has received $n^{2\varepsilon}$ or more proposals. Let $\mathcal{E}_\infty$ be the event that these same conditions hold when Part IV ends.

Lemma B.9. *The event $\mathcal{E}_\infty$ occurs wvhp.*

Proof. By Lemma B.7, we know that at the end of Part II, no man has made more than n^ε proposals, that $|S_{\mathrm{II}}| \ge n^{(1-\varepsilon)/2}$, and that no woman has received more than n^ε proposals, wvhp. We assume that all these conditions hold.

Fix a man m. We argue that if m makes a successful proposal to a woman in $S_{\mathrm{II}} \cup \bar{\mathcal{W}}$, then he makes no further proposals in Algorithm 2: If m makes a successful proposal to a woman in S, this ends the phase making the phase a terminal one, man m goes back to the woman to whom he was matched at the beginning of the phase, and this woman becomes a member of S. Thus, if a m makes a successful proposal to a woman in S, he makes no further proposals. In particular, if m makes a successful proposal to a woman in $S_{\mathrm{II}} \backslash \bar{\mathcal{W}}$, then he makes no further proposals.

Suppose man m is proposing in proposal t and that $\mathcal{E}_t$ holds. Then m has not yet applied to at least $3n^{(1-\varepsilon)/2}/4$ women in S_{II}. Hence the probability of applying to a woman in S_{II} is at least $n^{(1-\varepsilon)/2}/2$. Further, since no woman has received $n^{2\varepsilon}$ or more proposals, the probability of the proposal being accepted is at least $1/n^{2\varepsilon}$. Hence, the probability of the proposal going to a woman in S_{II} and being accepted is at least $n^{-3\varepsilon-1/2}$. Hence, the man makes fewer than $n^{0.6}/2$ proposals in Part IV, and proposes to fewer than $n^{3\varepsilon}/2$ additional women in S_{II}, except with probability $\exp(-n^{\varepsilon}/2)$. Using a union bound over the men, wvhp, no man has applied to more than $n^{0.6}$ women in total or to more than $n^{3\varepsilon}$ women in S_{II} until the end of Part IV.

Fix a woman w. Each time a proposal occurs, since no man has proposed to more than $n^{0.6}$ women (assuming $\mathcal{E}_t$ holds), the probability of the proposal going to w is less than $2/n$. Since there are at most $50n \log n$ proposals in total, the number of proposals received by w in Part III is more than n^{ε} with probability less than $\mathbb{P}(\text{Binomial}(50n \log n, 2/n) \geq n^{\varepsilon}) \leq 2^{-n^{\varepsilon}}$, using Chernoff bounds (see Fact E.1(ii) in Appendix E). Using a union bound over the women, wvhp, no woman has received more than n^{ε} proposals until the end of Part IV.

The result follows combining the analyses in the two paragraphs above. □

We now focus on Part III. We show (Lemma B.10) that for every phase in Part III, whp:

- the phase is a terminal phase, and
- that $|V|$ at the end of the phase is at least $n^{0.25}$.

For each such phase, $|S|$ increases by at least $n^{0.25}$. In addition, we show that phases are short, with the expected length of a phase being $O(n^{1/2+3\varepsilon})$. We infer that, wvhp, we reach $|S| \geq n^{0.7}$, i.e., the end of Part III, in $o(n^{0.47})$ phases, containing $o(n)$ proposals. Lemma B.11 below formalizes this.

Lemma B.10. *Assume $|S_{\mathrm{II}}| \geq n^{(1-\varepsilon)/2}$, cf. Lemma B.7. Consider a phase during Part III. Suppose $\mathbb{I}(\mathcal{E}_t) = 1$ at the start of the phase. Then, whp, either $\mathbb{I}(\mathcal{E}_{t'}) = 0$ at the end of the phase, or we have:*

- *The phase is a terminal phase.*

- *At the end of the phase is at least $|V| \geq n^{0.25}$.*

Proof. Assume $\mathbb{I}(\mathcal{E}_\tau) = 1$ throughout the phase (otherwise there is nothing to prove). Since we are considering a phase during Part III, we know that $|S| < n^{0.7}$. Also, $|S| \geq |S_{\text{II}}| \geq n^{(1-\varepsilon)/2}$ by assumption. For each proposal, there is a probability of at least $|S|n^{-2\varepsilon}/(2n) \geq n^{-1/2-3\varepsilon}$ and at most $2|S|/n \leq 2n^{-0.3}$, that the proposal is to a woman in S and is accepted. It follows that, whp, the phase is a terminal phase, and the number of proposals in the phase is in $[n^{0.28}, n^{0.52}]$. It is easy to see that with probability at least $(1-2n^{0.28}/n)^{n^{0.28}} = 1-o(1)$, all of the first $n^{0.28}$ proposals in the phase are to distinct women, meaning that there are no IICs. For each proposal, the probability of acceptance is at least $1/(1+n^{2\varepsilon})$, since no woman has received $n^{2\varepsilon}$ proposals, so whp, there are at least $n^{0.25}$ accepted proposals among the first $n^{0.28}$ proposals, using Fact E.1 (ii) on $\mathrm{Binomial}(\mathrm{n}^{0.26}, 1/(1+\mathrm{n}^{2\varepsilon}))$. Now, consider the first $n^{0.25}$ women in V. These women receive no further proposals during the phase with a probability at least $(1-2n^{0.25}/n)^{n^{0.52}} = 1-o(1)$. Hence, whp, these women are part of V at the end of the phase, establishing $|V| \geq n^{0.25}$ at the end of the phase as needed. $\square$

Lemma B.11. *Wvhp, Part III contains less than $n^{0.99}$ proposals.*

Proof. We first show that the next $n^{0.47}$ phases after the end of Part II complete in fewer than $n^{0.99}$ proposals. Since, wvhp, $|S_{\text{II}}| \geq n^{(1-\varepsilon)/2}$, if $\mathbb{I}(\mathcal{E}_t) = 1$ then for proposal t the probability of ending the phase (due to acceptance by a woman in S) is at least $n^{-1/2-3\varepsilon}$. It follows that either $\mathbb{I}(\mathcal{E}_\infty) = 0$ or wvhp, the next $n^{0.47}$ phases after the end of Part II complete in no more than $n^{0.47+1/2+4\varepsilon} \leq n^{0.99}$ proposals, using Fact E.1 (ii) on $\mathbb{P}(\mathrm{Binomial}(\mathrm{n}^{0.97+4\varepsilon}, \mathrm{n}^{-1/2-3\varepsilon}) \geq \mathrm{n}^{0.47})$.

Now we show that wvhp, Part III contains fewer than $n^{0.47}$ phases. Suppose this is not the case, then, by our definition of Part III, at most $n^{0.45}$ of these phases increase $|S|$ by $n^{0.25}$ or more. But using Lemma B.10, either $\mathbb{I}(\mathcal{E}_\infty) = 0$, or this occurs with probability at most

$$\mathbb{P}\big(\mathrm{Binomial}(\mathrm{n}^{0.47}, 1-\varepsilon) \leq \mathrm{n}^{0.45}\big) \leq \mathbb{P}\big(\mathrm{Binomial}(\mathrm{n}^{0.47}, 1/2) \leq \mathrm{n}^{0.47}/4\big) \leq 2\exp(-\mathrm{n}^{0.47}/24)\,,$$

using Fact E.1 (i). In other words, either $\mathbb{I}(\mathcal{E}_\infty) = 0$ or, wvhp, Part III contains fewer than $n^{0.47}$ phases.

But Lemma B.9 tells us that $\mathbb{I}(\mathcal{E}_\infty) = 1$ wvhp. Combining the above, we deduce that wvhp, Part III contains fewer than $n^{0.47}$ phases and fewer than $n^{0.99}$ proposals. $\square$

B.4 Part IV

Lemma B.12. *Suppose we are at Step 3 (time t) of Algorithm 2 during Part IV, we have $\mathbb{I}(\mathcal{E}_t) = 1$, and man m is proposing. Then, for large enough n, the probability that:*

(i) Man m proposes to S and is accepted is at least $n^{-0.31}$.

(ii) Man m proposes to $\mathcal{W}\backslash(R(m) \cup \hat{w})$ and is accepted is at least $0.9n/t$.

Proof. Note that $|S| \geq n^{0.7}$, whereas by definition of $\mathcal{E}_t$ the man has proposed to no more than $n^{0.6}$ women and no woman has received more than $n^{2\varepsilon}$ proposals. (i) follows from Lemma B.2.

Proof of (ii): Since m has not applied to more than $n^{0.6}$ women so far, we know that $|\mathcal{W}\backslash(R(m) \cup \hat{w})| \geq 0.95n$. Also, the total number of proposals so far is $t-1$. Using Lemma B.2, the probability of applying to $\mathcal{W}\backslash(R(m) \cup \hat{w})$ and being accepted is at least

$$\frac{1}{1 + (t-1)/(0.95n)} \geq \frac{0.9n}{t},$$

for large enough n, using $t > n$ since Part I itself requires at least n proposals. $\square$

Lemma B.13. *Wvhp, Part IV ends due to termination of the algorithm.*

Proof. Suppose Part IV does not end with termination (if not we are done) and that $\mathcal{E}_\infty$ occurs (Lemma B.9 guarantees this wvhp). Reveal each proposal sequentially.

For $t \leq 40n \log n$, call proposal t a 'seemingly-good' proposal when acceptance by $w' \in \mathcal{W}\backslash(R(m) \cup \hat{w})$ occurs. Denote the set of seemingly good proposals by $\mathcal{A}$. We use Lemma B.12. For each proposal t, there is a probability at least $0.9n/t$ of it being a seemingly-good proposal, conditioned on the history so far. Define independent $X_t \sim \text{Bernoulli}(0.9n/t)$ for $t = t_0, t_0 + 1 \ldots, 40n \log n$, where t_0 is the first proposal in Part III. Then we can set up a coupling so that proposal $t \in \mathcal{A}$ whenever $X_t = 1$. Now

$$\sum_{t=4n\log n}^{40n\log n} 1/t \geq (0.99)\ln(10) \geq 2.27\,,$$

$$\Rightarrow \qquad \sum_{t=4n\log n}^{40n\log n} \mathbb{E}[X_t] \qquad \geq 2n$$

Using Fact E.1, we deduce that

$$\sum_{t=4n\log n}^{40n\log n} X_t \geq 7n/4$$

wvhp, implying

$$|\mathcal{A}| \geq \sum_{t=t_0}^{40n\log n} X_t \geq 7n/4 \tag{5}$$

wvhp, since we know that $t_0 \leq 4n \log n$ wvhp using Lemmas B.4 and B.6.

We call a seemingly-good proposal $t \leq 40n \log n$ a 'good' proposal if the following conditions are satisfied:

- During the current phase, there is no proposal to a woman in V. In particular, there are no IICs.
- The phase is a terminal phase that ends during Part IV.

We denote the set of good proposals by $\mathcal{G} \subseteq \mathcal{A}$. We now argue that

$$|S_{\mathrm{II}}| \geq |\mathcal{G}|, \tag{6}$$

where S_{II} is the set S at the end of Part IV. If $w' \notin (S \cup \bar{w})$, then w' becomes part of S at the end of the phase if the proposal is a good proposal. For each terminal phase, there is exactly one good proposal to a woman in $S \cup \bar{w}$, which we think of as accounting for $\hat{w}$, which also becomes a part of S. Thus, for every good proposal, one woman joins S during Part IV, establishing Eq. (6).

Now consider any phase in Part IV that starts before the $40n \log n$-th proposal. Call such a phase an *early* phase. Using Lemma B.12, the phase contains more than $n^{0.32}$ proposals with probability at most $(1 - n^{-0.31})^{n^{0.32}} \leq \exp(-n^{0.01}) \leq 1/n^2$. But the total number of early phases is no more than $40n \log n$. It follows that using a union bound that, wvhp, there is no early phase that contains more $n^{0.32}$ proposals.

Now, the probability of a phase containing fewer than $n^{0.32}$ proposals, and containing a proposal to a woman in V is at most $n^{0.32} \cdot 2n^{0.32}/n \leq n^{-0.35}$, since $|V| \leq n^{0.32}$ throughout such a phase. Further, there are at most $40n \log n \leq n^{1.01}$ early phases. It follows, using Fact E.1 (ii) on $\mathrm{Binomial}(\mathrm{n}^{1.01}, \mathrm{n}^{-0.35})$, that the number of early phases containing a proposal to a woman in V is, wvhp, no more than $n^{0.67}$. It follows that, wvhp, no more than $n^{0.67} \cdot n^{0.32} =$

$n^{0.99}$ proposals occur in early phases containing a proposal to V. But all proposals in $\mathcal{A}\backslash\mathcal{G}$ must occur in such phases. We deduce that, wvhp,

$$|\mathcal{A}\backslash\mathcal{G}| \le n^{0.99}\,. \tag{7}$$

Combining Eqs. (5) and (7), we deduce that $|\mathcal{G}| \ge 3n/2 \ge |\mathcal{W}|$ wvhp. Plugging in Eq. (6), we obtain $|S_{\mathrm{II}}| \ge |\mathcal{W}|$ at the end of Part IV wvhp, which we interpret[14] as "With wvhp, our assumption that Part IV does not end with termination was incorrect. In other words, Part IV ends with termination wvhp". □

Lemma B.14. *The number of proposals in improvement phases and in IICs in Part IV is no more than $n^{0.99}$ wvhp.*

Proof. In the proof of Lemma B.13, we in fact showed that whp in Part IV, the number of proposals in phases that include a proposal to a woman in V is no more than $n^{0.99}$. (Actually, we showed this bound for 'early' phases, and also showed that wvhp, the algorithm terminates with an early phase, so that all phases in Part IV are early phases). But improvement phases and phases containing IICs must include a proposal to a woman in V. The result follows. □

We now establish two claims that follow from elementary calculus.

Claim B.15. *For large enough n and any $k \ge 1$ we have $k\log(1+n/k) \ge (\log n)/2$.*

Proof. We divide possible values of k into three ranges, and establish the bound for each range.

First, suppose $k \le 10\log n$. Then $\log(1+n/(10\log n)) \ge \log\sqrt{n} \ge \log n/2$ for large enough n. It follows that $k\log(1+n/k) \ge \log n/2$ since $k \ge 1$.

Next, suppose $k \in (10\log n, 10n]$. Now $\log(1+n/k) \ge \log(1+n/(10n)) = \log 1.1 \ge 0.09$. The bound follows by multiplying with $k \ge 10\log n$.

Finally, consider $k > 10n$. Now, $n/k \le 1/10$ leading to $\log(1+n/k) \ge n/k - (n/k)^2/2 \ge 0.95n/k$. It follows that $k\log(1+n/k) \ge 0.95n \ge \log n/2$ for large enough n. □

Claim B.16. *For any $n \ge 1$ and any $k \ge 1$, we have $(1+1/n)\log(1+n) \ge (1+k/n)\log(1+n/k) \ge 1$.*

[14] Recall our initial assumption that Part IV does not end with termination. Finding that $|S| \ge n$ under this assumption simply means that Part IV did, in fact, end with termination.

Proof. Consider $f : (0, \infty) \to \mathbb{R}$ defined by $f(x) = (1 + 1/x)\log(1 + x)$. Then

$$f'(x) = \frac{x - \log(1+x)}{x^2} > 0 \quad \forall x > 0\,,$$

using $\log(1 + x) < x$ for all $x > 0$. It follows that for any $x > 0$, we have

$$f(x) \geq \lim_{x\to 0} f(x) = 1,$$

using $\lim_{x\to 0} \log(1+x)/x = 1$. The lower bound in the claim follows by plugging in $x = n/k$. The upper bound follows by plugging in $k = 1$, since this maximizes n/k for fixed n. □

Finally, using these lemmas and claims we give the proof of Theorem 1.

Proof of Theorem 1. Using Lemma A.3, we can calculate the sum of men's rank of wives by summing up the rank under the MOSM and the number of proposals made during improvement phases and IICs during the run of Algorithm 2. By Lemma B.6, B.11 and Lemma B.14, the total number of proposals that occur in improvement phases and IICs (in Parts II-IV) of Algorithm 2 is, wvhp, no more than $(1 + n/k)(\log(n))^{0.45} + 2n^{0.99}$. Using Lemma A.3, we get that

$$\begin{aligned}&\text{Sum of men's rank of wives(WOSM)} - \text{Sum of men's rank of wives(MOSM)}\\ &\leq (1 + n/k)(\log(n))^{0.45} + 2n^{0.99}\end{aligned} \tag{8}$$

But

$$\begin{aligned}&\text{Sum of men's rank of wives(MOSM)}\\ &\geq 0.99(n + k)\log((n + k)/k)\\ &\geq \max(0.49(1 + n/k)\log n, 0.99n)\end{aligned} \tag{9}$$

wvhp, from Lemma B.4 (i), along with Claims B.15 and B.16. We deduce from Eqs. (8) and (9) that

$$\frac{R_{\text{MEN}}(\text{WOSM}) - R_{\text{MEN}}(\text{MOSM})}{R_{\text{MEN}}(\text{MOSM})} \leq (\log n)^{-0.55}/0.49 + 2n^{-0.01}/0.99 \leq (\log n)^{-0.4}$$

wvhp, immediately implying Theorem 1 (iii).

The only agents whose partner changes in going from the MOSM to the WOSM are the ones who make or receive accepted proposals during improvement phases and IICs. But this number on each side of the market is, wvhp, no more than $n/(2\sqrt{\log n})$ in Part II (Lemma B.8), and no more than $n^{0.99}$ each in Part III (Lemma B.11) and Part IV (Lemma B.14, leading to a bound of $n/(2\sqrt{\log n}) + 2n^{0.99} \leq n/\sqrt{\log n}$ on the total number of agents with multiple stable partners on each side of the market, establishing Theorem 1 (ii).

By Lemma B.8, wvhp, the improvement in the sum of women's rank of husbands in Part II is at most $n^2/(2(\log n)^{3/2})$. In Parts III and IV, wvhp there are at most $2n^{0.99}$ women who obtain better husbands (since each such woman must have received a proposal during an improvement phase or IIC; see above), and the rank improves by less than n for each of these women, so the improvement in sum of ranks is less than $2n^{1.99}$. It follows that the total improvement in sum of ranks is, wvhp, less than $n^2/(\log n)^{3/2}$. On the other hand, using Lemma B.4 (iii) and the upper bound in Claim B.16, we obtain that wvhp

$$\text{Sum of women's rank of husbands(MOSM)} \geq n^2/(2\log n)\,.$$

Theorem 1 (iv) follows.

Lemma B.4 (i) and (iii) with $\epsilon' = \epsilon/2$, combined with Theorem 1 (iii) and (iv) (established above) yields Theorem 1 (i).

□

C Many-to-one matching markets

This section discusses the extension of our results to many-to-one matching markets, in which colleges are matched with more than one student. We consider many-to-one markets in which colleges have a small capacity relative to the size of the market, each student has an independent, uniformly random complete preference list over colleges, and each college has responsive preferences (Roth, 1985) and an independent, uniformly random complete preference list over individual students. In Section 4.4, we presented computational experiments demonstrating that imbalance in such markets again leads to a small core and allows the short side to approximately "choose." We now follow to describe how our theoretical results can be extended.

Assume that each college has a constant number of seats q. Students are on the short side of the market if there are fewer students than seats, and, symmetrically, colleges are on the short side of the market if there are more students than seats. We denote the extreme stable matchings by SOSM (the student-optimal stable matching) and COSM (the college-optimal stable matching). The students' average rank of their colleges is defined as before. We define the colleges' rank of students to be the average rank of students assigned to the college.

We argue that the bounds on average rank stated in Theorem 2 will deteriorate by a factor that depends on q,[15] whereas Theorem 1 will hold as stated. Thus, with high probability, the core will be small and the short side will "choose."

Our proof can be extended as follows. First, using results from Roth and Sotomayor (1989), one can decompose the many-to-one market into a one-to-one market as follows. For each seat in a college, create an "agent" that ranks students according to that college's preferences. Students rank all seats according to their preferences for the corresponding schools, and all students rank seats within a college in the same order (i.e., there is a "top" seat and a "bottom" seat in each college). A matching is stable in the original market if and only if there is a corresponding stable matching in the decomposed market. With this decomposition, all our results from Appendix A extend to the many-to-one case, allowing us to use our algorithms to calculate the extreme stable matchings via a sequence of proposals by agents on the short side.

Next, the stochastic analysis can be extended to the many-to-one case as follows. Suppose there are n colleges, each with q seats. First consider the case in which there are fewer than qn students. The first part of the analysis is student-proposing DA, and we can bound the number of proposals in this stage by considering q repetitions of the coupon collector's problem. The next steps in our proof follow with slight modifications, using the fact that rejection chains have the same structure as in the one-to-one case. Whenever a seat rejects a student, the student matched with the bottom seat in the college gets rejected, and that bottom student in turn applies to a randomly drawn college that ranks the student uniformly at random. A college will accept the student if the applying student is more preferred than the q-th best student currently at the college and will reject that q-th best student if it accepts the applying student. A phase (chain), initiated by a college c rejecting a student, can terminate either with an application to a school that has not filled its seats, or with a successful application to college c. Therefore, a phase consists of a series of proposals to random colleges that in case of acceptance, always reject their lowest-ranked student. The

[15]That is, there is a similar upper bound on the average rank of partners for the short side of the market, and there is a similar lower bound on the average rank for the long side of the market.

main difference is that in order to calculate the acceptance probability, we need to calculate the probability that a (randomly ranked) proposing student is better than the q-th best proposal the college received (instead of being better than the best proposal the woman received in the one-to-one case). Bounds on this probability will be affected by at most a constant factor.

Similar arguments allow us to extend our proof to the case where there are more students than seats. The first part of the analysis is college-proposing DA, and we can again bound the number of proposals in this stage by considering the coupon collector's problem. The rest of our analysis, controlling rejection chains, is almost unchanged. Whenever a student in a college rejects a seat, the seat accepts the student matched with the next lower seat at the college, and so on until the bottom seat in the college is rejected. This seat proposes to a randomly drawn student who ranks the college uniformly at random and accepts if she prefers the college over her current match. A phase (chain) initiated by a student s rejecting a seat can either terminate with an offer to an unmatched student, or with a successful offer to student s. Overall, the analysis in this case will be almost identical to our original proof.

We note that these results do not imply that colleges cannot gain from manipulation in unbalanced matching markets, as a college can potentially manipulate even if there is a unique stable matching. Kojima and Pathak (2009) show that a college can manipulate only if a rejection of one of the students assigned to that college triggers a rejection chain that cycles back to the college. We therefore conjecture that when the imbalance is larger than a college's capacity, even colleges have a limited scope for manipulation, but this conjecture does not directly follow from our analysis, which only establishes that the core is small.

D Average rank estimates

Table 1 in Section 4 includes values of the following function for each unbalanced market.

$$\mathrm{EST} = \mathrm{EST}(|\mathcal{M}|, |\mathcal{W}|) = \begin{cases} \frac{|\mathcal{W}|}{|\mathcal{M}|} \log\left(\frac{|\mathcal{W}|}{|\mathcal{W}|-|\mathcal{M}|}\right) & \text{for } |\mathcal{W}| > |\mathcal{M}| \\ |\mathcal{W}| \Big/ \left(1 + \frac{|\mathcal{M}|}{|\mathcal{W}|} \log\left(\frac{|\mathcal{M}|}{|\mathcal{M}|-|\mathcal{W}|}\right)\right) & \text{for } |\mathcal{W}| < |\mathcal{M}| \end{cases} \tag{10}$$

The definition of EST is based on Theorem 2, as justified by the following facts.

Remark 2. *Fix any $\epsilon > 0$.*

- *For the case $|\mathcal{W}| > |\mathcal{M}|$, Theorem 2 and its proof imply that, with high probability (asymptotically in $|\mathcal{M}|$), $R_{\text{MEN}}/\text{EST} \in (1-\epsilon, 1+\epsilon)$ under all stable matches (including the MOSM and WOSM). Note that the upper bound on R_{MEN} is part of the statement of Theorem 2, whereas the lower bound follows from the proof (though it is not part of the statement). Hence, we think of* EST *as a heuristic estimate for R_{MEN} in finite markets with $|\mathcal{W}| > |\mathcal{M}|$.*

- *For the case $|\mathcal{W}| < |\mathcal{M}|$, Theorem 2 implies that, with high probability (asymptotically in $|\mathcal{W}|$), $R_{\text{MEN}}/\text{EST} \geq 1-\epsilon$ under all stable matches (including the MOSM and WOSM). Hence, we think of* EST *as a heuristic lower bound for R_{MEN} in finite markets with $|\mathcal{W}| < |\mathcal{M}|$.*

E Chernoff bounds

Fact E.1 Chernoff bounds (see Durrett (2010))**.** *Let $X_i \in \{0,1\}$ be independent with $\mathbb{P}[X_i = 1] = \theta_i$ for $1 \leq i \leq n$. Let $X = \sum_{i=1}^n X_i$ and $\lambda = \sum_{i=1}^n \theta_i$.*

(i) Fix any $\delta \in (0,1)$. Then

$$\mathbb{P}(|X-\lambda| \geq \lambda\delta) \leq 2\exp\{-\delta^2\lambda/3\}. \tag{11}$$

(ii) For any $R \geq 6\lambda$, we have

$$\mathbb{P}(X \geq R) \leq 2^{-R}. \tag{12}$$

References

Durrett, R., *Probability: theory and examples*, Cambridge University Press, 2010.

Immorlica, N. and M. Mahdian, "Marriage, honesty, and stability," in "Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete algorithms" 2005, pp. 53–62.

Kojima, F. and P. A. Pathak, "Incentives and Stability in Large Two-Sided Matching Markets," *American Economic Review*, 2009, *99*, 608–627.

McVitie, D.G. and L.B. Wilson, "The stable marriage problem," *Communications of the ACM*, 1971, *14(7)*, 486–490.

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

Roth, A.E., "The college admissions problem is not equivalent to the marriage problem," *Journal of Economic Theory*, 1985, *36* (2), 277–288.

_ , "On the allocation of residents to rural hospitals: a general property of two-sided matching markets," *Econometrica*, 1986, pp. 425–427.

_ **and M. Sotomayor**, "The college admissions problem revisited," *Econometrica*, 1989, *57* (3), 559–570.