Online Appendix
Two-Sided Matching with (almost) One-Sided Preferences

Guillaume Haeringer and Vincent Iehlé

Contents

A Blocks: additional results

A.1 Basic Properties

In the next proposition, we show that if conditions 1 and 2 of a block hold, then condition 2 also holds when we consider all possible sets $\mathbf{K} \subseteq \mathbf{J} \backslash \mathbf{J}_{i,d}$. For a given block at (i, d), the proposition also establishes that if there are less than q_d candidates ranked higher than i **in the block** then i cannot be an impossible match for d. So, a necessary condition for i to be an impossible match is that there are q_d or more such candidates in the block. Finally, we slightly reinforce the conclusion of Theorem 2. We show that, in the presence of a block at (i, d), whenever candidate i is matched to d it must be that one of the candidate j from the block must remain unmatched, and is ranked higher than a candidate matched at one of the acceptable departments for j.

Proposition A.1 *Every block $\mathbf{J}$ at (i, d) satisfies the following properties:*

1. *For each nonempty $\mathbf{K} \subseteq \mathbf{J} \setminus \mathbf{J}_{i,d}$, Hall's marriage condition (3) does not hold in $\overline{P}^d(\mathbf{J}, \mathbf{K})$.*
2. *There are at least q_d candidates in $\mathbf{J}$ ranked higher than i at department d, i.e., $|\mathbf{J}_{i,d}| \geq q_d$.*
3. *If μ is a feasible matching such that $\mu(i) = d$ then there exists $k \in \mathbf{J}$ and $j \in \mathbf{J} \cup \{i\}$ such that $\mu(j) \neq j$, $\mu(k) = k$ and $kP_{\mu(j)}j$.*

To prove Propositions A.1 we need the following lemma.

Lemma A.2 *Let P be a proto-matching problem and $(i, d) \in A$. Let $\mathbf{J} \subseteq I \setminus \{i\}$, with $d \in A_{\mathbf{J}}$, be a nonempty subset of candidates such that for some $\mathbf{J}' \subseteq \mathbf{J}$ with $|\mathbf{J}'| = \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}}$, Hall's condition (3) is satisfied in $P(\mathbf{J}', \emptyset)$. The following three conditions are equivalent,*

(a) For each nonempty $\mathbf{K} \subseteq \mathbf{J} \setminus \mathbf{J}_{i,d}$, Hall's marriage condition (3) does not hold in $\overline{P}^d(\mathbf{J}, \mathbf{K})$.

(b) For each $\mathbf{K} \subseteq \mathbf{J} \setminus \mathbf{J}_{i,d}$, with $|\mathbf{K}| = |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1$, Hall's marriage condition does not hold in $\overline{P}^d(\mathbf{J}, \mathbf{K})$.

(c) *For each* $\mathbf{K}' \subseteq \mathbf{J}\backslash\mathbf{J}_{i,d}$ *such that* $|\mathbf{K}'| = |\mathbf{J}| - \sum_{\hat{d}\in A_{\mathbf{J}}} q_{\hat{d}} + 1$, *there is no perfect match between* $\mathbf{J}\backslash\mathbf{K}'$ *and* $A_{\mathbf{J}}$ *in the problem* $\overline{P}^d(\mathbf{J},\mathbf{K}')$.

Proof Obviously, *(a)* $\Rightarrow$ *(b)* $\Rightarrow$*(c)*. Thus it remains to show $\neg$*(a)* $\Rightarrow$ $\neg$*(c)*. To this end, suppose there exists a set $\mathbf{K} \subseteq \mathbf{J}\backslash\mathbf{J}_{i,d}$ such that Hall's condition (3) is satisfied in $\overline{P}^d(\mathbf{J},\mathbf{K})$. So, there exists a matching μ such that each candidate admissible in $\overline{P}^d(\mathbf{J},\mathbf{K})$ is matched to a department. Define the matching μ' that leaves the match of admissible candidates unchanged in $\overline{P}^d(\mathbf{J},\mathbf{K})$ under μ and such that $\mu'(j) = j$ for every other candidate j. Then $\mu\prime$ is also comprehensive in $\overline{P}^d(\mathbf{J},\emptyset)$. Note that every candidate in $\mathbf{J}_{i,d}$ is matched to a department under μ' since these candidates are all admissible in $\overline{P}^d(\mathbf{J},\mathbf{K})$.

Applying successively Lemma 1 we obtain a comprehensive and maximum matching $\tilde{\mu}$ in $\overline{P}^d(\mathbf{J},\emptyset)$, where every candidate $j \in \mathbf{J}_{i,d}$ is matched to a department. By assumption, there exists a perfect match between a set $\mathbf{J}' \subseteq \mathbf{J}$ such that $|\mathbf{J}'| = \sum_{\hat{d}\in A_{\mathbf{J}}} q_{\hat{d}}$ and $A_{\mathbf{J}}$. It follows that each department in $A_{\mathbf{J}}$ shall fill its capacity at a maximum matching in the problem $\overline{P}^d(\mathbf{J},\emptyset)$. Thus $\tilde{\mu}$ is comprehensive for $\overline{P}^d(\mathbf{J},\emptyset)$ and every candidate in $\mathbf{J}_{i,d}$ is matched to a department and every department in $A_{\mathbf{J}}$ fills its capacity under $\tilde{\mu}$. Consider now $\mathbf{K}' = \{j \in \mathbf{J} \ : \ \tilde{\mu}(j) = j\}$. Clearly, $|\mathbf{K}'| = |\mathbf{J}| - \sum_{\hat{d}\in A_{\mathbf{J}}} q_{\hat{d}} + 1$. In addition, since $\tilde{\mu}$ is comprehensive for $\overline{P}^d(\mathbf{J},\emptyset)$, every candidate $j \in \mathbf{J} \setminus \mathbf{K}'$ is admissible in $\overline{P}^d(\mathbf{J},\mathbf{K}')$. We conclude that $\tilde{\mu}$ is a perfect match between $\mathbf{J}\backslash\mathbf{K}'$ and $A_{\mathbf{J}}$ in the problem $\overline{P}^d(\mathbf{J},\mathbf{K}')$, as was to be proved. ■

Proof of Proposition A.1

1. See Lemma A.2.

2. One can use the alternative to condition 2 given in (a) of Lemma A.2. The set $\mathbf{J}_{i,d}$ cannot be empty, for otherwise Hall's marriage condition trivially holds in $\overline{P}^d(\mathbf{J},\mathbf{K})$ with $\mathbf{K} = \mathbf{J}$, which is a contradiction. If $\mathbf{J}_{i,d} = \mathbf{J}$ then condition 1 of a block implies $|\mathbf{J}_{i,d}| \geq q_d$. Otherwise, consider the nonempty set $\mathbf{K} = \mathbf{J} \setminus \mathbf{J}_{i,d}$. The set of admissible candidates in $\overline{P}^d(\mathbf{J},\mathbf{K})$ is $\mathbf{J}_{i,d}$. Since Hall's marriage condition (3) does not holds in $\overline{P}^d(\mathbf{J},\mathbf{K})$, one cannot assign all candidates in $\mathbf{J}_{i,d}$ to department d at any feasible matching in $\overline{P}^d(\mathbf{J},\mathbf{K})$. That is, $|\mathbf{J}_{i,d}| > q_d - 1$.

3. Let $\mathbf{J}$ be a block at (i,d), and let μ be a feasible matching such that $\mu(i) = d$. Let $\mathbf{K} = \{j \in \mathbf{J} \ : \ \mu(j) = j\}$. Since $\mu(i) = d$, we have that $\mathbf{K} \neq \emptyset$ from condition 1 of the definition of a block. Suppose now by way of contradiction that the conclusion of *3*. is not satisfied. Since $\mu(i) = d$ we have then necessarily $\mathbf{K} \subseteq \mathbf{J} \backslash \mathbf{J}_{i,d}$. Let μ' be such that $\mu'(i) = i$, $\mu'(j) = \mu(j)$ if $j \in \mathbf{J}$, and $\mu'(j) = j$ if $j \notin \mathbf{J}$. We observe that μ' is a feasible matching in $\overline{P}^d(\mathbf{J}, \mathbf{K})$, where, by construction, every admissible candidate is matched to a department under μ'. This contradicts the statement (a) of Lemma A.2. ■

A.2 A simple sufficient condition: exact block

We consider a particular case of a block where the number of candidates in $\mathbf{J}$ is **equal to** the total number of positions in the acceptable departments. We will see that the conditions 1 and 2 specialize respectively into a perfect match between $\mathbf{J}$ and $A_{\mathbf{J}}$, and Hall's marriage condition (3) with singleton sets $\mathbf{K}$.

Definition A.3 Given a proto-matching problem $P = (I, D, (P_d, q_d)_{d \in D}, A)$, an **exact block at** $(i, d) \in A$ is a nonempty set $\mathbf{J} \subseteq I \backslash \{i\}$, with $d \in A_{\mathbf{J}}$, such that:

1. $|\mathbf{J}| = \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}}$ and Hall's marriage condition (3) holds in $P(\mathbf{J}, \emptyset)$;
2. For each $k \in \mathbf{J} \setminus \mathbf{J}_{i,d}$ Hall's marriage condition does not hold in $\overline{P}^d(\mathbf{J}, \{k\})$.

The above tractable conditions are sufficient for the existence of impossible matches. In the next section, we will show that those conditions are also necessary under a mild restriction on the proto-matching problem.

The following result is deduced from Theorem 2 in the paper.

Corollary A.4 *If P admits an exact block at (i, d) then i is an impossible match for d.*

Proof Let $\mathbf{J}$ be an exact block at (i, d). It is easy to check that $\mathbf{J}$ is also a block at (i, d). Condition 1 of the definition of a block holds with $\mathbf{J}' = \mathbf{J}$. Since $|\mathbf{J}| = \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}}$, the second condition of an exact block is equivalent to the condition 2. of a block with $|\mathbf{K}| = |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1 = 1$. ■

The next two examples illustrate the interplay between conditions 1 and 2 (for blocks or exact blocks) in order to get impossible matches.

P_{d_0}	P_{d_1}
i_1	i_2
i_0	i_1

Table 1: Impossible match and exact block

Example A.5 Let d_0 and d_1 be two departments and i_0, i_1 and i_2 three candidates. Consider the problem P depicted in Table 1, where $A_{i_1} = \{d_0, d_1\}$, $A_{i_2} = \{d_1\}$, $q_{d_0} = q_{d_1} = 1$.

Clearly, i_0 is an impossible match for d_0, and it is easy to see that $\mathbf{J} = \{i_1, i_2\}$ is an exact block at (i_0, d_0). Condition 1 holds trivially. As for condition 2, $\mathbf{K} = \{i_2\}$ is the only set $\mathbf{K}$ for which we have to consider condition 2. So we only have to check whether we can match i_1 in $\overline{P}^{d_0}(\mathbf{J}, \{i_2\})$. This is impossible since d_2 is not admissible and d_1 has no position. So, $\mathbf{J}$ is an exact block. $\square$

An exact block $\mathbf{J}$ does not always exist. A counter-example, where an impossible match does not admit any exact block, is the following.

Example A.6 Consider four departments, d_0, d_1, d_2 and d_3, with one position at each department, and six candidates i_h, $h = 0, \ldots, 5$. The problem P is depicted in Table 2. It is easy to see that candidate i_0 is an impossible match for d_0.

P_{d_0}	P_{d_1}	P_{d_2}	P_{d_3}
i_1	i_2	i_2	i_3
i_0	i_3	i_4	i_4
	i_5	i_1	i_5
	i_1		

Table 2: Impossible match and no exact block

To begin, notice that any block must contain at least 4 candidates, otherwise condition 1 cannot be satisfied.[1] The set $\mathbf{J^1} = \{i_1, i_2, i_3, i_4\}$ satisfies condition 1, but not condition 2.

[1]Candidate i_1 must be part of the block, and thus condition 1 implies that we must include at least two other candidates to fill departments d_1 and d_2. So, any block must contain at least candidate i_3, or i_4 or i_5. If any of these three candidates is added to the block, we span all four departments, and thus any block must contain at least four candidates.

To see this, consider the matching μ (feasible for $\overline{P}^{d_0}$) defined by $\mu(i_1) = d_1$, $\mu(i_2) = d_2$ and $\mu(i_3) = d_3$. Candidate i_4 is ranked lower than the matched candidates for all his acceptable departments. So, condition 2 is not satisfied for the block $\mathbf{J}^1$, since Hall's condition (3) is satisfied in $\overline{P}^{d_0}(\mathbf{J}^1, \{i_4\})$.

A similar reasoning applies for the sets $\mathbf{J}^2 = \{i_1, i_2, i_3, i_5\}$, $\mathbf{J}^3 = \{i_1, i_2, i_4, i_5\}$, and $\mathbf{J}^4 = \{i_1, i_3, i_4, i_5\}$. That is, there is no block $\mathbf{J}$ at (i_0, d_0) such that $|\mathbf{J}| = 4$,, hence no exact block, while the reader can check that $\mathbf{J} = \{i_1, i_2, i_3, i_4, i_5\}$ is the (unique) block at (i_0, d_0). $\square$

A.3 Finding blocks and impossible matches

According to Proposition 2 in the paper, we would have to consider all possible maximal and comprehensive matchings to check whether a candidate is an impossible match for a department. Theorem 2 in the paper greatly simplifies this task, for we only have to consider all the sets $\mathbf{K} \subseteq \mathbf{J}$ of a particular size. The set $\mathbf{J}$ turns out to be easy to construct; the proof of the necessity part of Theorem 2 suggests a procedure to identify impossible matches, which we now detail.

Let P be a proto-matching problem and let $i \in A_d$.

Algorithm: Impossible Matches Algorithm

1. Delete candidate i from the preference of all departments but department d, without changing the relative ranking of the other candidates.

2. Construct a maximum and comprehensive matching μ.[2]

3. If $\mu(i) = d$ then i is not an impossible match for d. Otherwise go to step 4.

4. Constructing the set $\mathbf{J}$:

    4.1 Let $\mathbf{J}^0$ be the set of all candidates matched to a department, and D^0 all the departments that do not fill their capacity at μ. If D^0 is empty then go to Step 5. Otherwise iterate through h until obtaining an empty set D^h:

[2] We describe in the proof of Lemma 1 in the Appendix a way to construct a maximum and comprehensive matching for a proto-matching problem. This construction consists of an iterative search of augmenting paths starting from a comprehensive matching (which could be empty) and can be performed in polynomial time —See the Section A in Appendix for a definition of augmenting paths.

4.2 Define $\mathbf{J}^h$ as the set $\mathbf{J}^{h-1}$ minus the set of candidates made by every candidate who finds acceptable at least one department in D^{h-1} ,

$$\mathbf{J}^h = \{j \in \mathbf{J}^{h-1} \ : \ j \notin A_{D^{h-1}}\} \, ,$$

and let D^h be the set of departments that do not fill their capacities with candidates in $\mathbf{J}^h$,

$$D^h = \{\hat{d} \ : \ |\mu(\hat{d}) \cap \mathbf{J}^h| < q_{\hat{d}}\} \, .$$

Set $h := h+1$ and repeat step 4.2 until $D^\ell = \emptyset$ for some $h = \ell$. Let $\bar{\mathbf{J}} = \mathbf{J}^\ell$.

5. If $d \notin A_{\bar{\mathbf{J}}}$ then i is not an impossible match for d. Otherwise check condition 2 of Definition 4 at $\mathbf{J}$:

5.1 Let $\bar{\mathbf{T}}$ be the set of unmatched candidates at μ (except i) such that every such candidate finds acceptable some department in $A_{\bar{\mathbf{J}}}$. Let $\mathbf{J} = \bar{\mathbf{J}} \cup \bar{\mathbf{T}}$.

5.2 For each $\mathbf{K} \subseteq \mathbf{J} \backslash \mathbf{J}_{i,d}$ such that $|\mathbf{K}| = |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1$, find a maximum matching in the problem $\overline{P}^d(\mathbf{J}, \mathbf{K})$. If there is no set $\mathbf{K}$ such that all admissible candidates of $\overline{P}^d(\mathbf{J}, \mathbf{K})$ are matched to a department then i is an impossible match for d. Otherwise i is not an impossible match.

We show in the proof of Theorem 2 that the set $\mathbf{J}$ constructed in Step 5.1 satisfies the first condition of a block, and Step 5.2 consists of checking condition 2. If there is a set $\mathbf{K}$ satisfying the conditions set in Step 5.2 and a matching μ such that all admissible candidates in $\overline{P}^d(\mathbf{J}, \mathbf{K})$ can be matched to a department then i is not an impossible match for d. This is so because μ is comprehensive for $P(\mathbf{J}, \emptyset)$. By construction no candidate better ranked than i at d is in $\mathbf{K}$, so these candidates are all matched. Also, in the problem $\overline{P}^d(\mathbf{J}, \mathbf{K})$ the capacity of d is $q_d - 1$, so it then suffices to match i to d to have a comprehensive matching for P, the original proto-matching problem. On the other hand, if no such set $\mathbf{K}$ can be found then Hall's condition (3) is not satisfied for any set $\mathbf{K}$, which simply establishes that the second condition of a block holds.

Note that the set $\bar{\mathbf{J}}$ constructed in Step 4 is clearly easy to obtain, so the complexity of our algorithm is determined by Step 5.2. This is where the added value of Theorem 2 comes in, for we only have to truncate by taking sets $\mathbf{K}$ that have the size specified in condition 2, a clear improvement compared to Proposition 2.

In practice, it is possible to further reduce the sets $\mathbf{K}$ one has to consider by not including the candidates who are necessarily matched at a maximal and comprehensive matching.

Definition A.7 Given a proto-matching problem P, a candidate i is **prevalent** if for each job market problem $\succ \in \Theta(P)$ and each $\mu \in \Sigma(\succ)$, it holds that $\mu(i) \in D$.

Finding whether a candidate i is prevalent is easy. It suffices to consider the profile P restricted to the departments A_i truncated at i, and then check whether there is at least one department in A_i that does not fill its capacity at a maximum matching. Given a proto-matching problem P, let $\mathbf{J}^*$ denote the set of prevalent candidates. We can then re-write the second condition of a block in the following way,

2'. For each $\mathbf{K} \subseteq \mathbf{J} \setminus (\mathbf{J}_{i,d} \cup \mathbf{J}^*)$, with $|\mathbf{K}| = |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1$, Hall's marriage condition (3) does not hold in $\overline{P}^d(\mathbf{J}, \mathbf{K})$.

Proposition A.8 *Let P be a proto-matching problem and $(i, d) \in A$. Let $\mathbf{J} \subseteq I \setminus \{i\}$, with $d \in A_{\mathbf{J}}$, be a nonempty subset of candidates satisfying condition 1. of a block. Then the condition 2. of a block and the condition 2' are equivalent.*

Proof Clearly 2 implies $2'$. To show the converse implication suppose that there exists $\mathbf{K} \subseteq \mathbf{J} \setminus \mathbf{J}_{i,d}$ with $|\mathbf{K}| = |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1$, such that Hall's marriage condition (3) holds in $\overline{P}^d(\mathbf{J}, \mathbf{K})$. From Lemma A.2 ($c$), there exists a perfect match between $\mathbf{J} \backslash \mathbf{K}$ and $A_{\mathbf{J}}$ in the problem $\overline{P}^d(\mathbf{J}, \mathbf{K})$, *i.e.*, where each department fills its capacity. Let j be a prevalent candidate. If $j \in \mathbf{K}$, then it is possible to fill all the departments in A_j with candidates that are ranked higher than j at these departments. This would contradict the fact that j is prevalent. So $j \notin \mathbf{K}$. It follows that $\mathbf{K} \subseteq \mathbf{J} \setminus (\mathbf{J}_{i,d} \cup \mathbf{J}^*)$. ■

Finally, we are also able to show that our algorithm yields a computationally easy task by imposing a slight restriction on the proto-matching problem. We assume that there exists a matching for P such that all candidates are matched to a department.

Assumption A.9 Hall's Marriage condition is satisfied in P.

Checking whether Hall's Marriage condition is satisfied for a given problem P simply amounts to computing a maximum matching and checking whether all candidates are matched. There are various algorithms doing this in polynomial time. A well-known algorithm is the Edmonds–Karp algorithm (Edmonds and Karp (1972)) which is similar to the procedure outlined in the proof of Lemma 1 (see also Remark 5). Another (and faster) algorithm is the Hopcroft–Karp algorithm (Hopcroft and Karp (1973)).

Proposition A.10 *Under Assumption A.9, candidate i is an impossible match for department d if, and only if, P admits an exact block at (i, d). Furthermore, finding whether a candidate is an impossible match can be done in polynomial time.*

Proof From Corollary A.4, the existence of an exact block $\mathbf{J}$ at (i, d) implies that i is an impossible match for department d.[3]

Suppose now that i is an impossible match for d in P. We show that there is an exact block $\mathbf{J}$. To begin with, note that we cannot have $\mathbf{J} = I \backslash \{i\}$. To see this, note that Assumption A.9 implies that $|I| \leq \sum_{\hat{q}} q_{\hat{d}}$. Since we necessarily have $i \notin \mathbf{J}$ we have $|I \backslash \{i\}| < \sum_{\hat{q}} q_{\hat{d}}$.

Let $\widetilde{P}$ be the problem identical to P except that $A_i^{\widetilde{P}} = \{d\}$. By Assumption A.9, there exists a matching $\widetilde{\mu}$ for P such that $\widetilde{\mu} \neq j$ for each $j \in I$. Let μ be the matching such that $\mu(j) = \widetilde{\mu}(j)$ for each $j \neq i$ and $\mu(i) = i$. Note that $\widetilde{\mu}$ is necessarily comprehensive in P. Since i is an impossible match for d in P, $\widetilde{\mu}(i) \neq d$. So, μ is a feasible matching for $\widetilde{P}$ and $|\{j \,:\, \mu(j) \neq j\}| = |I| - 1$. So any maximum and comprehensive matching for $\widetilde{P}$ necessarily matches $|I| - 1$ candidates. So we can assume without loss of generality that μ is also a comprehensive matching.

We can now construct a set $\mathbf{J}^\ell$ following the procedure detailed in the proof of Theorem 2.[4] Now consider the set $\mathbf{T}$. It contains all the unmatched candidates different from i that are acceptable for a department in $A_{\mathbf{J}^\ell}^{\widetilde{P}}$. But since $\widetilde{\mu}(j) \neq j$ for each $j \neq i$, we have $\mathbf{T} = \emptyset$.

From there it suffices to follow the arguments in the necessity part of Theorem 2 (after the claim that $\mu(d_0) \subseteq \mathbf{J}^\ell$). Since $\mathbf{T} = \emptyset$, $|\mathbf{J}| = \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}}$, the two conditions for a block become the two conditions for an exact block.

We show now that under Assumption A.9 we can check whether a candidate is an impossible match for a department in polynomial time. To this end, consider the Impossible Matches Algorithm. The matching constructed in Step 2 is done in polynomial time (see Remark 5 in the paper). At step 4, the matching is such that all candidates but candidate i are matched to a department. Hence, at step 5.1, $\mathbf{T} = \emptyset$ and $|\mathbf{J}| = \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}}$. At Step 5.2, the maximal number of cases one has to consider is thus equal to $|I| - 1$. That is, we only have to check that, for each candidate $j \in \mathbf{J} \setminus \mathbf{J}_{i,d}$, Hall's marriage condition (3) is not

[3] Note that Assumption A.9 is not needed here.

[4] Observe that the set D^0 is nonempty. This is so because $\widetilde{\mu}(i) \neq i$ and thus $|\mu(\widetilde{\mu}(i))| \leq q_{\widetilde{\mu}(i)} - 1$.

satisfied in the problem $\overline{P}^d(\mathbf{J}, \{j\})$. Again, it amounts to finding a maximum matching for the problem $\overline{P}^d(\mathbf{J}, \{j\})$ which can also be done in polynomial time.[5] ■

Remark A.11 The identification of impossible matches can be sped up when scrutinizing the blocks. If we identify a block (resp. exact block) $\mathbf{J}$ at some pair candidate–department (i, d), then for each candidate ranked lower than i at d who is not an element of the block (resp. exact block) $\mathbf{J}$, the set $\mathbf{J}$ is also a block (resp. exact block) for (i, d). That is, the candidate is also an impossible match for d.

Remark A.12 The identification can also be sped up by considering the connected components of the graph defined by the mutually acceptable pairs, in the proof of Theorem 2. If a block at (i_0, d_0) involves two disconnected components of the acceptability graph, then there exists one block in the connected component containing d_0. This obvious property can be helpful in practice. It means one can work separately on each connected component of the market when looking at impossible matches one by one.

B The French academic job market

Here we provide more details about the French academic job market.

B.1 Organization

Disciplines in French academics are divided into *sections*, that correspond to fields. For instance the section corresponding to economics is section ♯ 5, section ♯ 10 corresponds to comparative literature. There are 77 different sections, but medical positions sections are often split into sub-sections corresponding to sub-specialities, bringing the total to about 120.

For candidates, the job market begins in the Fall. Candidates first request an evaluation by a national committee, the *Conseil National des Universités*. There is a committee for each section. A positive evaluation by a committee is valid for four years, after which the

[5] We can use for instance the Hopcroft–Karp algorithm. Contrary to Step 2 in Step 5.2 we only need to look for a maximum matching (i.e., not necessarily comprehensive).

candidate must ask for another evaluation, should he or she want to participate again in the job market.

Positions are advertised by the Ministry of Higher Education in winter. Shortly after the publication of job openings the candidates send their applications to the departments where they want to apply. By law a candidate need only to have a positive evaluation by one section to apply to any job. That is, a candidate having a positive evaluation from the committee for section, say, 17 (philosophy), can apply to a position advertised in section, say, 34 (astronomy and astrophysics). In practice, recruiting committees reject candidates who do not have a positive evaluation for the section corresponding to the job opening.

Each open position is assigned a recruiting committee of about 15–20 people. The composition of the committee is constrained by a double parity on the members' origin (internal or external) and the type of position (junior or senior). Upon receiving the candidates' applications, the recruiting committees decide first which candidate to interview. The interviews usually take place in May, and usually last between 20 and 30 minutes. In most cases all the interviews are made the same day. Contrary to, for instance, the (international) job market for economists, interviews in the French job market are held at the universities. That is, candidates must travel to attend their interviews. So, a candidate may have an interview scheduled in Paris and Marseilles (900 km away) the same day, at the same time. Committees try to be flexible to permit candidates to tune their schedules in case of conflicting interviews.

After the interviews, recruiting committees rank the candidates (usually the same day) and the rankings are submitted to the Ministry of Higher Education after validation by the University board.[6] Once the interview season is over, candidates are asked to submit (in June) their preferences over the positions for which they are ranked. When logging in, candidates see *all* the positions for which they are ranked (and their rank), independently of the section (i.e., field) in which the positions were advertised, and have to rank them. For instance, if a candidate was ranked for positions that were advertised for sections, say, 26 and 34, the candidate will have to submit a unique ranking, not one for each field. In other words, the Ministry of Higher Education is considering one job market encompassing

[6]The university board ("*Conseil d'administration*" is composed of (mostly) faculty, staff and students' representative. The board has the power to alter the ranking decided by the recruiting committee. In mathematics this does happen, albeit rarely. We consider in the paper the rankings decided by the recruiting committees.

all sections.

The Ministry then computes a matching, which is announced about a week after.[7] Most positions start on September 1st of the same year.

Remark B.1 Parallel to the university job market there is another job market involving research institutions (CNRS, INRA, INSERM, INRIA, etc.). These positions (also tenured) are research-only positions and are usually preferred by the candidates. These institutions usually announce which candidate is offered a job in April, i.e., before the interviews. However, candidates have little control over the institution where they will be assigned. The assignment is usually announced in June, when candidates must submit their choice lists for the assistant professor positions. That is, a candidate hired by the CNRS (in April) does not know before June *where* (e.g., Paris, Marseilles or Toulouse) he will be assigned.

Remark B.2 There are each year between 80 and 130 positions open in sections 25 and 26. This high number of positions is rather unusual. For instance, in economics there are on average about 40–50 positions a year. This high number is partly explained by the fact that mathematics departments often provide the mathematics courses taught in other disciplines. Each year, there are on average between 400 and 500 candidates having passed the evaluation of the CNU of sections 25 or 26, so this gives an average of 5 candidates per opening.[8] On average about 30–40% of the candidates do not pass the CNU evaluation. The main reasons of reject are misfit (the field of specialization of the candidate does not match the field of the section), incomplete application and weakness of the scientific production (about 5% of the applicants).

B.2 Changes during the time framework

Our observations cover 15 years (from 1999 to 2013) of the market for assistant professors in Mathematics. We observe four major institutional changes along those years. In chronological order:

[7]For 2016 for instance, the deadline for submitting the preferences over the positions is June 16 and the match is announced on June 20.

[8]There are slightly more candidates qualified for section 26 than for section 25. The former often includes candidates from other disciplines (e.g., economics, physics, biostatistics, chemistry, computer science)

No more linked positions - since 2008 It can happen that a department offers two or three identical positions at the same faculty (i.e., positions with identical job descriptions). Until 2007, those positions were considered "linked," which implied that the recruiting committee had to establish a unique ranking for all the linked positions.

a new decentralized market Since September 2008, departments must choose between **two** separate methods of recruiting before entering the market. The first one is a "synchronized session," where the timeline for recruiting is imposed by the Ministry of Higher Education. Departments recruiting in this market are competing with each other *at the same time*. This is the market we consider in the paper. The other market is not coordinated. That is, departments can choose the dates of the interview and make job offers to candidates immediately thereafter. This market is called "*au fil de l'eau*" (on the fly). A cause for concern is merely the existence of overlaps between the decentralized and centralized markets for the candidates who matter in our data. But the few positions of the uncoordinated market are usually published well before with the respect to the synchronized market, so that the candidates accepting such early positions do not appear in the latter market.[9]

No more second session – since 2009 The Ministry used to organize a second job market in September which would sometimes include positions that were not filled in the first job market session (although in general the second session positions were new positions). This session had a very small number of positions and will not be considered in this paper (e.g., for mathematics, there were 6 positions in 2006, 2 in 2007, 1 in 2008).

No more constraints for the ranking length – since 2009 Until 2008 departments could rank at most 5 candidates per position. In case of "linked positions" (until 2007), the length of the rank list was $5 \times \sharp$ linked positions. For instance, if a department had 3 linked positions it could rank up to 15 candidates. The Ministry of Higher Education dropped the constraint on the number of candidates that could be ranked. Our data

[9]The co-existence of a second market may have important strategic implications for both departments and candidates. Some of them are already well-documented in the matching literature, including strategic unraveling and the problem of early resolution of uncertainty (e.g., Halaburda (2010), Echenique and Pereyra (2016)). We do not discuss those issues here.

shows that it took some time for the departments to fully integrate this new possibility in their strategies. By 2013, most of the departments ranked more than 5 candidates per position, suggesting that the constraints on the length of the rankings was binding. However, the length of the rankings have not increased substantially. For the last year in our data (2013), the distribution of the length of the rankings are given in Table 3. When looking at the rank of hired candidates (Table 4), we can see that there is no substantial impact of the abandon of the constraint on the length of the rankings.[10]

Length	3	4	5	6	7	8	9	10
♯ positions	1	9	17	27	12	4	2	2
% of positions	1.4%	12.2%	23%	36.5%	16.2%	5.4%	2.7%	2.7%

Table 3: Distribution of ranking lengths (year 2013)

Rank	1	2	3	4	5	6	7	8	9 and more
♯ cand. (year 2013)	26	15	11	11	2	3	1	0	2
% cand. (year 2013)	36.6%	21.1%	15.5%	15.5%	2.8%	4.2%	1.4%	0%	2.8%
avg. years $\leq$ 2008	32.8%	20.9%	16.7%	10.4%	7.2%	3.6%	3.7%	2.2%	2.5%
avg. years $>$ 2008	29.3%	20.2%	13.8%	12.7%	6.2%	6.9%	5.0%	3.1%	2.9%

Table 4: Rank of hired candidates (for the position that hired them)

The effects of those institutional changes are globally ambiguous. On the one hand, we observe that the joint abandon of quotas and linked positions leads to a thicker centralized market, all things being equal. But on the other hand, the effect is mitigated by the creation of a second decentralized market that deprives the centralized market of candidates and positions.

[10]The average distribution for the years prior to 2009 in Table 4 is calculated not taking into account the linked positions. If two positions are linked then the candidate ranked 2nd could be considered as ranked first (he is sure to get either position if he desires so).

Year	Number Violations	Number Violators
1999	4	4
2000	3	3
2001	16	15
2002	9	6
2003	6	4
2004	6	5
2005	7	6
2006	11	11
2007	6	6
2008	12	8
2009	19	15
2010	5	4
2011	16	13
2012	12	9
2013	9	6

Table 5: Number of violations of the acceptability assumption

C More on acceptability

C.1 Violation of the acceptability assumption on the French market

Table 5 provides the distribution of the *obvious* violations of the assumption by candidates. See the discussion in sub-section IV.C in the Empirical Analysis.

C.2 Acceptable or achievable?

As we explained in the Introduction, our goal is to analyze the *consequences* of short preference lists and not why preference lists are short. We can nevertheless use the concept of impossible matches to shed some light on the structure of those lists. A natural hypothesis to explain the length of the departments' preference lists is that departments are naturally constrained in their choices. Any candidate that appears in a department's submitted preference list must be interviewed by that department. Obviously, in a sufficiently large job market like the one we are analyzing it is not materially possible for departments to interview all candidates who applied to the positions offered by the departments. But departments are likely to be strategic, too. That is, when establishing the rankings of candidates departments may focus on a subset of candidates, namely candidates that they may have a chance to hire. In other words, a candidate not appearing in a department's (submitted) preference list is not necessarily unacceptable for the department. A reasonable assumption could be then that departments consider in their ranking only candidates that are *achievable*, that is, candidates that departments can be matched with at some stable matching (in the job market problem). The next result shows that if this were the case then we should not identify impossible matches in the data. Therefore the presence of impossible matches in our data shows that departments do not rank only achievable candidates.[11]

Before going further, some definitions are in order. Let $\succ$ be a job market problem. A candidate i is **achievable** for a department d (or d is achievable for i) if there exists a matching μ stable for $\succ$ such that $\mu(i) = d$. Given a job market $(I, D, (\succ_d, q_d)_{d\in D}, (\succ_i)_{i\in I})$, we define the **achievable proto-matching problem associated to** $\succ$ the proto-matching problem $(I, D, (\widetilde{P_d}, q_d)_{d\in D}, \widetilde{A})$ such that

- for each $i \in I$, $d \in D$, $(i, d) \in \widetilde{A}$ if, and only if, i is achievable for d in $\succ$
- for each $d \in D$, and $i, i' \in I$, $i\widetilde{P}_d i'$ if, and only if, $iP_d i'$.

Proposition C.1 *Let $\succ$ be a job market problem and let $\widetilde{P}$ be the achievable proto-matching problem associated to $\succ$. If a candidate i is achievable for a department d in $\succ$ then i is not an impossible match for d in $\widetilde{P}$.*

Proof Let $\succ$ be a job market problem and let i_0 be achievable for d_0 in $\succ$, i.e., there exists a matching $\mu \in \Sigma(\succ)$ such that $\mu(i_0) = d_0$. Let $P = (I, D, (P_d, q_d)_{d\in D}, A)$ be the

[11] Intuitively, this result is not too surprising since matching markets tend to have a unique stable matchings as the market gets larger. In this case, most departments would rank only one candidate per position, the one who is matched to that position at the unique stable matchings. One issue with that reasoning is that large markets do *not necessarily* have a unique stable matchings. Even though we may suspect that in the French job market there is only one stable matching, we cannot show for sure that it is the case.

proto-matching problem such that $\succ \in \Theta(P)$. Since μ is stable, μ is maximal and comprehensive for P. Now consider $\widetilde{P}$, the achievable proto-matching problem associated to $\succ$. By construction, for each $j \in I$, $\mu(j) \neq j$ implies $j \in A_d^{\widetilde{P}}$. Note that for each department d, $A_d^{\widetilde{P}} \subseteq A_d$ and for each $j, j' \in A_d^{\widetilde{P}}$, $j \widetilde{P}_d j'$ if, and only if, $j P_d j'$. It follows that μ is also maximal and comprehensive for $\widetilde{P}$, and thus $\mu(i_0) = d_0$ implies that i_0 is not an impossible match for d_0 in $\widetilde{P}$. ■

D Data analysis

D.1 A remark on reshuffled profiles

In this case, the ranking of a department simply consists of a random reshuffling of the original ranking. This ensures that the distribution of number of times candidates are ranked is identical to the original distribution.

Until 2007 some positions were linked, i.e., recruiting committees had to give a unique ranking for all the linked positions. In that case we opted for reshuffling in the same way for all linked positions. For instance, if two positions are linked and rank (in this order) candidates a, b, c, d and e, the reshuffled ranking of the two positions would be the same (e.g., c, e, a, d and b).

D.2 Market stars

In the main paper the definition of a star given a proto-matching problem P is a candidate that is ranked more than once and always first (at some stage of the Predicting Algorithm). In the program we use an additional definition of stars, that we call *pseudo stars*. Such a candidate is a candidate who is always ranked weakly higher than the hired candidate (but not necessarily always first). Table 6 illustrates this, where the candidates in boxes are the matched (hired) candidates.

In this example candidate i_1 is a pseudo star. He is ranked by d_1 and d_2, and his rank is always weakly higher than the rank of the hired candidate. Candidate i_2, on the contrary, is not a pseudo star, because for d_4 he is ranked below the hired candidate.

Table 6: Pseudo stars.

P_{d_1}	P_{d_2}	P_{d_3}	P_{d_4}
i_2	i_2	i_5	$\boxed{i_4}$
i_1	$\boxed{i_1}$	$\boxed{i_2}$	i_2
$\boxed{i_5}$	i_5	i_6	i_7
i_4	i_3	i_4	i_3

Since we know that the final matching is stable, a pseudo market star is similar to a market star in the following sense: the department that hires that candidate is necessarily the top choice among all the other positions that rank that candidate. Indeed, if it is not the case the matching would not be stable. So, in our case we know that i_1 necessarily prefers d_2 to d_1. Note, however, that this deduction is only valid for a given proto-matching problem at a given step of the algorithm. In our predicting algorithm it could be that at a previous step i_1 was ranked by, say, d_4, but *below* the hired candidate, i_4. In this case we obviously cannot infer the relative ranking of d_4 and d_1 in i_1's preferences, and thus at that previous step i_1 was not a pseudo star.

The structure of our code works as follows when considering stars (and pseudo stars). The algorithm is similar to the algorithm presented in the paper, except that we also search for pseudo stars. This search is only triggered when there are no market stars.

Predicting Algorithm

Input: A pair (P, μ), where μ is the output of a regular matching algorithm with input $\succ \in \Theta(P)$.

Step 1 Identify and remove all impossible match pairs in P and construct $P^1 := R(P)$.

Step h for $h \geq 2$

h.1 Identify the market stars in P^h.

If there are no market stars, identify the pseudo stars in P^h.

If there are no stars and no pseudo stars, then STOP.

h.2 If there are stars (or pseudo stars), eliminate them from the rankings of the positions that do not hire them. Let $\widetilde{P}^h$ be the proto-matching problem obtained.

h.3 Construct $P^{h+1} := R(\widetilde{P}^h)$ and go to Step $h+1$.

It is routine to check that with this version of the Predicting Algorithm the results of Section III continue to hold. The proof of Proposition 7, which is the unique result dealing with market stars, simply needs some minor modifications in the phrasing (paragraphs 2 and 3 in the proof).

Table 7 reports the number of market stars and pseudo stars used in the Predicting Algorithm for every year. Note that the total number of stars and pseudo stars outnumber the number of positions extracted. Indeed, very few pseudo stars (at most three per year) identified at step $h.2$ of the above algorithm are not mapped to a position by the algorithm. Those candidates are identified by the algorithm and possibly used to clear the market since they are eliminated from the rankings of the positions that do not hire them. However, the positions hiring them cannot be cleared at any step of the algorithm, that is *extracted.* In the example given in Table 6, candidate i_1 is a pseudo star but the algorithm cannot clear any of the four positions (even after iterations). In that example, there is no position extracted and one pseudo star.

D.3 Simulations

Table 8 reports the detailed results of the simulations that we discuss in sub-section IV.D of the empirical analysis.

References

[1] **Echenique, Federico and Juan S. Pereyra** 2016. “Strategic complementarities and unraveling in matching markets.” *Theoretical Economics*, vol. 11, 1–39.

[2] **Edmonds, Jack and Richard M. Karp** 1972. “Theoretical improvements in algorithmic efficiency for network flow problems.” *Journal of the ACM. Association for Computing Machinery*, vol. 19, 248–264.

[3] **Halaburda, Hanna** 2010. “Unravelling in two-sided matching markets and similarity of preferences.” *Games and Economic Behavior*, vol. 69, 365–393.

year	Positions extracted	Number Stars	Number Pseudo Stars
1999	28	21	11
2000	14	9	6
2001	22	20	2
2002	14	13	2
2003	17	15	3
2004	18	18	0
2005	16	14	5
2006	29	22	9
2007	25	22	4
2008	16	13	6
2009	18	16	4
2010	14	9	6
2011	28	24	7
2012	19	15	5
2013	13	10	4
Average	19.4	16	4.7

Table 7: Extractions

[4] **Hopcroft, John E. and Richard M. Karp** 1973. "An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs." *SIAM Journal on computing*, vol. 2, 225–231.

year	Original	Random	Reshuffle
1999	32.2%	31%	33.1%
2000	49.4%	40.8%	43.8%
2001	44.2%	41.1%	47.1%
2002	52.9%	42.6%	48.3%
2003	44.3%	40.7%	38%
2004	50.7%	44.1%	47.8%
2005	44.9%	38%	46.3%
2006	39.1%	27.8%	40.3%
2007	47%	36.7%	45.3%
2008	36%	28.1%	34.4%
2009	33.9%	29.3%	30.2%
2010	34.2%	27.3%	36%
2011	38.8%	28.5%	35.3%
2012	60.2%	53.3%	51%
2013	58.9%	48.3%	52.2%
average	44.4%	37.2%	41.9%

Table 8: % of predicted positions