SIAM J. DISCRETE MATH.
Vol. 31, No. 4, pp. 2348–2377

POPULAR MATCHINGS WITH TWO-SIDED PREFERENCES AND ONE-SIDED TIES*

ÁGNES CSEH†, CHIEN-CHUNG HUANG‡, AND TELIKEPALLI KAVITHA§

Abstract. We are given a bipartite graph $G = (A \cup B, E)$ where each vertex has a preference list ranking its neighbors: In particular, every $a \in A$ ranks its neighbors in a strict order of preference, whereas the preference list of any $b \in B$ may contain ties. A matching M is *popular* if there is no matching M' such that the number of vertices that prefer M' to M exceeds the number of vertices that prefer M to M'. We show that the problem of deciding whether G admits a popular matching or not is NP-hard. This is the case even when every $b \in B$ either has a strict preference list or puts all its neighbors into a single tie. In contrast, we show that the problem becomes polynomially solvable in the case when each $b \in B$ puts all its neighbors into a single tie. That is, all neighbors of b are tied in b's list and b desires to be matched to any of them. Our main result is an $O(n^2)$ algorithm (where $n = |A \cup B|$) for the popular matching problem in this model. Note that this model is quite different from the model where vertices in B have no preferences and do *not* care whether they are matched or not.

Key words. popular matching, NP-complete, polynomial algorithm, ties

AMS subject classifications. 05C85, 68R10, 68Q17

DOI. 10.1137/16M1076162

1. Introduction.

We are given a bipartite graph $G = (A \cup B, E)$ where the vertices in A are called applicants and the vertices in B are called posts, and each vertex has a preference list ranking its neighbors in an order of preference. Here we assume that vertices in A have strict preferences while vertices in B are allowed to have ties in their preference lists. Thus each applicant ranks all posts that she finds interesting in a strict order of preference, while each post need not come up with a total order on all interested applicants—here applicants may get grouped together in terms of their suitability, thus equally competent applicants are tied together at the same rank.

Our goal is to compute a *popular* matching in G. The definition of popularity uses the notion of each vertex casting a "vote" for one matching versus another. A vertex v *prefers* matching M to matching M' if either v is unmatched in M' and matched in M or v is matched in both matchings and $M(v)$ (v's partner in M) is ranked better than $M'(v)$ in v's preference list. In an election between matchings M and M', each vertex v votes for the matching that it prefers or it abstains from voting if M and M' are equally preferable to v. Let $\phi(M, M')$ be the number of vertices that vote for M in an election between M and M'.

DEFINITION 1. *A matching M is* popular *if $\phi(M, M') \geq \phi(M', M)$ for every matching M'.*

If $\phi(M', M) > \phi(M, M')$, then we say M' is *more popular* than M and denote it

*Received by the editors May 19, 2016; accepted for publication (in revised form) August 7, 2017; published electronically October 18, 2017. A preliminary version of this paper appeared in ICALP 2015 [4].
http://www.siam.org/journals/sidma/31-4/M107616.html

†Hungarian Academy of Sciences and Corvinus University of Budapest, Budapest 1112, Hungary (cseh.agnes@krtk.mta.hu).
‡École Normale Supérieure, 75005 Paris, France (cchuang@di.end.fr).
§School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India (kavitha@tcs.tifr.res.in).

by $M' \succ M$; else $M \succeq M'$. Observe that popular matchings need not always exist. Consider an instance where $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2, b_3\}$, and for $i = 1, 2, 3$, each a_i has the same preference list which is b_1 followed by b_2 followed by b_3 while each b_i ranks a_1, a_2, a_3 the same, i.e., a_1, a_2, a_3 are tied together in b_i's preference list (see bottom left instance in Figure 2). It is easy to see that for any matching M here, there is another matching M' such that $M' \succ M$, thus this instance admits no popular matching.

The popular matching problem is to determine if a given instance $G = (A \cup B, E)$ admits a popular matching or not, and if so, to compute one. This problem has been studied in the following two models:

– 1-*sided model:* Here it is only vertices in A (also called agents) that have preferences and cast votes; vertices in B are objects with no preferences or votes.
– 2-*sided model:* Vertices on both sides have preferences and cast votes.

Popular matchings need not always exist in the 1-sided model and the problem of deciding whether a given instance admits one or not can be solved efficiently using the characterization and algorithm from [1]. In the 2-sided model when all preference lists are strict, it is known that any stable matching is popular [7]; thus a popular matching can be found in linear time using the Gale–Shapley algorithm. However, when ties are allowed in preference lists on both sides, Biró, Irving, and Manlove [3] showed that the popular matching problem is NP-complete. One of our models discusses this case further, strengthening the above result. Our other model deals with the following variant of the two-sided model with ties:

∗ Vertices on *both* sides cast votes. However, it is only vertices of A that *rank* their neighbors in a strict order of preference, in other words, the preference list of each vertex of A is strict while the preference list of each vertex of B contains a single "large tie."

That is, in the above model, vertices in B have no ranking over their neighbors—however, each $b \in B$ desires to be matched to *any* of its neighbors. Thus in an election between two matchings, b abstains from voting if it is matched in both or unmatched in both, else it votes for the matching where it is matched.

The above model is a natural variant of the 1-sided model (recall that A is a set of agents and B is a set of objects here) where each object has an owner who gains a fixed profit by allocating the object to an agent. Such fixed price markets occur, for example, in housing markets where the house owner earns rent when his house gets allotted to a tenant. Thus agents have preferences over objects and each object-owner wants his object to get matched to some agent so as to earn the cost of the object. That is, each object has a vote and does not care who is matched to it as long as it is matched to someone.

We will see in section 2 that the above problem is significantly different from the popular matching problem in the 1-sided model where vertices in B do not cast votes. We show the following results here, complementing our polynomial time algorithm in Theorem 2 with an NP-hardness result in Theorem 3. Note that Theorem 3 deals with the case when vertices of B are also allowed to have strict preference lists.

THEOREM 2. *Let $G = (A \cup B, E)$ be a bipartite graph where each $a \in A$ has a strict preference list while each $b \in B$ puts all its neighbors into a single tie. The popular matching problem in G can be solved in $O(n^2)$ time, where $|A \cup B| = n$.*

THEOREM 3. *Let $G = (A \cup B, E)$ be a bipartite graph where each $a \in A$ has a strict preference list while each $b \in B$ either has a strict preference list or puts all its* *neighbors into a single tie. The popular matching problem in G is* NP*-complete.*

Thus Theorem 3 tells us that the popular matching problem with two-sided preferences and one-sided ties is NP-hard even in the restricted case where the preference list of each $b \in B$ is either *strict* or a *single tie*. We know that the two extreme cases admit polynomial time algorithms, i.e., (i) when the preference list of every vertex of B is strict (popular matchings always exist in this case [7]) or (ii) when the preference list of every vertex of B is a single tie (by Theorem 2).

When proving Theorem 2 we show that a graph G, where each $a \in A$ has a strict preference list and each $b \in B$ puts all its neighbors into a single tie, admits a popular matching if and only if a new graph H that we construct here (H is essentially a subgraph of G) admits an A-complete matching, i.e., one that matches all vertices in A. The graph H is based on a partition $\langle X, Y, Z \rangle$ of B, where the first set X is a subset of top posts and, roughly speaking, the second set Y consists of *mid-level* posts, while the third set Z consists of *unwanted* posts. We show that corresponding to any popular matching in G, there is a partition $\langle L_1, L_2, L_3 \rangle$ of B into *top posts*, *mid-level posts*, and *unwanted posts* such that $X \supseteq L_1$ and $Z \subseteq L_3$, where $\langle X, Y, Z \rangle$ is the partition computed by our algorithm to construct H. This allows us to show that if H does not admit an A-complete matching, then G has no popular matching. In fact, not every popular matching in G becomes an A-complete matching in H (section 3 has such an example). However, it will be the case that if G admits popular matchings, then at least one of them becomes an A-complete matching in H.

Theorem 3 follows from a simple reduction from the (2,2)-E3-SAT problem. The (2,2)-E3-SAT problem takes as its input a Boolean formula $\mathcal{I}$ in CNF, where each clause contains three literals and every variable appears exactly twice in unnegated form and exactly twice in negated form in the clauses. The problem is to determine if $\mathcal{I}$ is satisfiable or not. This problem is NP-complete [2] and our reduction shows that the following version of the 2-sided popular matching problem in $G = (A \cup B, E)$ with 1-sided ties is NP-complete:

– every vertex in A has a strict preference list of length 2 or 4;
– every vertex in B has either a strict preference list of length 2 or a single tie of length 2 or 3 as a preference list.

Note that our NP-hardness reduction needs B to have $\Omega(|B|)$ vertices with strict preference lists and $\Omega(|B|)$ vertices with single ties as their preference lists.

Background. Popular matchings have been well-studied in the 1-sided model [1, 15, 16, 17, 19, 20] where only vertices of A have preferences and cast votes. Abraham et al. [1] gave efficient algorithms to determine if a given instance admits a popular matching or not—their algorithm also works when preference lists of vertices in A admit ties. The capacitated, many-to-one matching extension of the problem was studied by Sng and Manlove [17], while many-to-many markets were considered by Paluch [21]. The notions of *least unpopular* matchings [18] and popular *mixed* matchings [14] were also proposed to deal with instances that had no popular matchings. For markets with edge weights, McDermid and Irving [19] gave a structural characterization of popular matchings. Mestre [20] showed that in the presence of vertex weights, a maximum weight maximum cardinality popular matching or a proof of its nonexistence can be found in polynomial time even in the presence of ties.

Gärdenfors [7], who introduced the notion of popularity, considered the popular matchings problem in the domain of 2-sided preference lists and showed that in any instance with *strict* preference lists, a stable matching is popular. Later, Biró, Irving, and Manlove [3] gave polynomial-time algorithms to test a given matching for popularity. Efficient algorithms for computing a maximum size popular matching in an instance $G = (A \cup B, E)$ with 2-sided strict preference lists were given in [10, 12]. Various structural properties of popular matchings in such instances have also been investigated, such as identifying which vertices can get matched in some popular matching [9], determining "popular edges" (those that belong to some popular matching) [5], solving the optimal popular half-integral matching problem [13], and studying the polytope of popular fractional matchings [11].

Organization of this paper. Section 2 has preliminaries. Section 3 contains our algorithm and its proof of correctness. Section 4 shows our NP-hardness result. We conclude with some open problems.

2. Preliminaries. For any $a \in A$, let $f(a)$ denote a's most desired, first choice post. Let $F = \{f(a) : a \in A\}$ be the set of these top posts. We will refer to posts in F as f-posts and to the ones in $B \setminus F$ as non-f-posts. For any $a \in A$, let r_a be the rank of a's most preferred non-f-post in a's preference list; when all of a's neighbors are in F, we set $r_a = \infty$. The following theorem characterizes popular matchings in the 1-sided voting model.

THEOREM 4 (from [1]). *Let $G = (A \cup B, E)$ be an instance of the 1-sided popular matching problem, where each $a \in A$ has a strict preference list. Let M be any matching in G. M is popular if and only if the following two properties are satisfied:*
(i) *M matches every $b \in F$ to some applicant a such that $b = f(a)$;*
(ii) *M matches each applicant a to either $f(a)$ or its neighbor of rank r_a.*

Thus the only applicants that may be left unmatched in a popular matching here are those $a \in A$ that satisfy $r_a = \infty$.

If b_1 is ranked better than b_2 in a's preference list (where $a \in A$), then we write $b_1 > b_2$ in a's list. Let us consider the following example where $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2, b_3\}$: both a_1 and a_2 have the same preference list which is $b_1 > b_2$ while a_3's preference list is $b_1 > b_2 > b_3$ (see the top left figure in Figure 2). Assume first that only applicants cast votes. The only posts that any of a_1, a_2, a_3 can be matched to in a popular matching here are b_1 and b_2. As there are three applicants and only two possible partners in a popular matching, there is no popular matching here. However, in our 2-sided voting model, where posts also care about being matched and all neighbors of a post are in a single tie in its preference list, we have a popular matching $\{(a_1, b_1), (a_2, b_2), (a_3, b_3)\}$. Note that b_3 is ranked third in a_3's preference list, which is worse than $r_{a_3} = 2$. However, such edges are permitted in popular matchings in our 2-sided model.

Consider the following example (see the middle figure in Figure 2): $A = \{a_0, a_1, a_2, a_3\}$ and $B = \{b_0, b_1, b_2, b_3\}$; both a_1 and a_2 have the same preference list which is $b_1 > b_2$ while a_3's preference list is $b_1 > b_0 > b_2$ and a_0's preference list is $b_0 > b_3$. There is again no popular matching here in the 1-sided model. However, in our 2-sided voting model, we have a popular matching $\{(a_0, b_3), (a_1, b_1), (a_2, b_2), (a_3, b_0)\}$. Note that $b_0 \in F$ and here it is matched to a_3 and $f(a_3) \neq b_0$; also a_3 is matched to its second ranked post: This is neither its top post nor its r_{a_3}-th ranked post ($r_{a_3} = 3$ here).

Thus popular matchings in our 2-sided voting model are quite different from the characterization given in Theorem 4 for popular matchings in the 1-sided model. Our algorithm (presented in section 3) uses the following decomposition.

Dulmage–Mendelsohn decomposition [6]. Let M be a maximum matching in a bipartite graph $G = (A \cup B, E)$. Using M, we can partition $A \cup B$ into three disjoint sets: A vertex v is *even* (similarly, *odd*) if there is an even (resp., odd) length alternating path with respect to M from an unmatched vertex to v. Similarly, a vertex v is *unreachable* if there is no alternating path from an unmatched vertex to v. Denote by $\mathcal{E}$, $\mathcal{O}$, and $\mathcal{U}$ the sets of even, odd, and unreachable vertices, respectively. The following properties (proved in [8]) will be used in our algorithm and analysis.

- $\mathcal{E}$, $\mathcal{O}$, and $\mathcal{U}$ are pairwise disjoint. Let M' be any maximum matching in G and let $\mathcal{E}'$, $\mathcal{O}'$, and $\mathcal{U}'$ be the sets of even, odd, and unreachable vertices with respect to M', respectively. Then $\mathcal{E} = \mathcal{E}'$, $\mathcal{O} = \mathcal{O}'$, and $\mathcal{U} = \mathcal{U}'$.
- Every maximum matching M matches all vertices in $\mathcal{O} \cup \mathcal{U}$ and has size $|\mathcal{O}| + |\mathcal{U}|/2$. In M, every vertex in $\mathcal{O}$ is matched with some vertex in $\mathcal{E}$, and every vertex in $\mathcal{U}$ is matched with another vertex in $\mathcal{U}$.
- The graph G has no edge in $\mathcal{E} \times (\mathcal{E} \cup \mathcal{U})$.

3. Finding popular matchings in a 2-sided voting model. The input is $G = (A \cup B, E)$ where each applicant $a \in A$ has a strict preference list while each post $b \in B$ has a single tie as its preference list. If G has a popular matching, then we will construct such a matching in this section; else we return the message "G has no popular matching." Recall that F is the set of top posts and r_a is the rank of a's most preferred non-f-post, for every $a \in A$.

Our goal is to construct a graph H such that G admits a popular matching if and only if H admits an A-complete matching. Note that the algorithm for popular matchings in the 1-sided popular matching problem is also based on the same idea: The algorithm in [1] constructs a graph based on the partition $\langle F, B \setminus F\rangle$ of B. While it is the case that in the 1-sided popular matching problem every applicant has to be matched to either its most preferred post in F or its most preferred post in $B \setminus F$, we saw in the examples given in section 2 that in our 2-sided popular matching problem an applicant a can be matched to a neighbor of rank worse than r_a; also posts in F can be matched to applicants who do not regard them as top posts.

Let M be any matching in G and let us label each edge (a, b) in $G \setminus M$ by the vote of a for b versus $M(a)$, i.e., if a prefers b to $M(a)$, then $\mathsf{label}(a, b) = +1$, otherwise $\mathsf{label}(a, b) = -1$. In case a is not matched in M, then $\mathsf{label}(a, b) = +1$ for any neighbor b of a. If M is popular in G, then the following two necessary conditions must hold on these edge labels:

(i) There is no alternating path ρ such that the edge labels in $\rho \setminus M$ are $\langle +1, +1, +1, \cdots\rangle$, i.e., no three consecutive nonmatching edges are labeled $+1$.

(ii) There is no alternating path ρ where the edge labels in $\rho \setminus M$ are $\langle +1, +1, -1, +1, +1, \cdots\rangle$, i.e., no five consecutive nonmatching edge labels add up to 3.

Otherwise $M \oplus \rho \succ M$. Inspired by the above two conditions that are necessary for a matching M to be popular, our algorithm will construct a 3-level partition $\langle X, Y, Z\rangle$ of B such that the following properties hold:

- $X \subseteq F$ and $Z \subseteq B \setminus F$,
- $Y \subseteq F \cup \{b \in B \setminus F : b \text{ has rank } r_a \text{ in some } a\text{'s preference list}\}$.

The graph H that we will construct here will be based on this partition $\langle X, Y, Z\rangle$. Using the partition $\langle X, Y, Z\rangle$ of B, we will build a graph H where each applicant keeps at most two edges: either (1) to its most preferred post in X and also in Y or (2) to its most preferred post in Z and also in Y. Later, we will define dummy posts that may be included in Y towards the end of our algorithm.

Our algorithm performs the partition of B into X, Y, and Z over several iterations. Initially $X = F$, $Y = B \setminus F$, and $Z = \emptyset$. In each iteration, certain non-top posts get *demoted* from Y to Z and similarly, certain top posts get demoted from X

to Y.

We will decide which f-posts belong to X and which belong to Y by trying to maintain the following property which will enable us to show that M obeys necessary condition (ii) stated above. Let $M(U)$ be the set of applicants who are matched to posts in U for any $U \subseteq B$.

$(*)$ There will be no edge in G between an applicant in $M(X)$ and a post in Z.

In order to maintain property $(*)$, we will partition A into two subsets: $A \setminus \mathsf{nbr}(Z)$ and $\mathsf{nbr}(Z)$, where $\mathsf{nbr}(U)$ (similarly, $\mathsf{nbr}_H(U)$) is the set of neighbors in G (resp., in H) of the vertices in U, for any subset U of vertices. Our algorithm will maintain $\mathsf{nbr}_H(X) \subseteq A \setminus \mathsf{nbr}(Z)$ and thus it is only the applicants in $A \setminus \mathsf{nbr}(Z)$ who will get matched to vertices in X in M and so property $(*)$ will be maintained. In each iteration, we have new posts entering Z from Y and this causes some applicants to move from $A \setminus \mathsf{nbr}(Z)$ to $\mathsf{nbr}(Z)$.

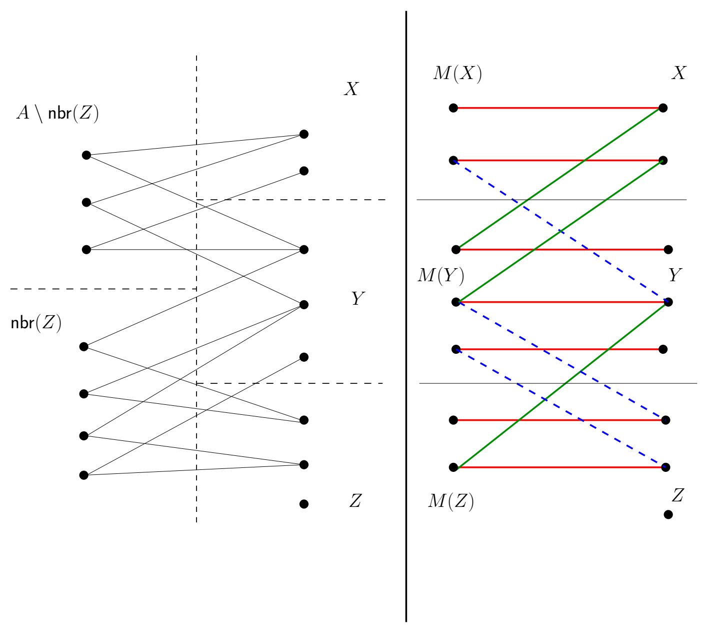

FIG. 1. *The set B gets partitioned into X, Y, and Z. We have $\mathsf{nbr}_H(X) \cap \mathsf{nbr}(Z) = \emptyset$. In the figure on the right, the horizontal edges belong to M. These belong to the graph H and the dashed edges are other possible edges in the graph H. It will be the case that only the edges of $(M(Y) \times X) \cup (M(Z) \times (X \cup Y))$ in G can be labeled $+1$.*

If M is the matching that is returned by our algorithm, then it will be the case that any edge that is labeled $+1$ by our edge labeling in $G \setminus M$ has to be either in $M(Y) \times X$ or in $M(Z) \times (X \cup Y)$ (see Figure 1). There will be *no* $+1$ edge incident on any applicant who is matched to a post in X. This will enable us to show that M obeys necessary condition (i) stated earlier.

We are now ready to formally describe our algorithm. Initialize $X = F$, $Y = B \setminus F$, and $Z = \emptyset$.

(I) While true do
 0. H is the empty graph on $A \cup B$.
 1. For each $a \in A \setminus \mathsf{nbr}(Z)$ do:
  – if $f(a) \in X$, then add the edge $(a, f(a))$ to H.

2. For every $b \in X$ that is isolated in H do:
   – delete b from X and add b to Y.
3. For each $a \in A$ do:
   – let b be a's most preferred post in the set Y; if the rank of b in a's preference list is $\leq r_a$ (i.e., r_a or better), then add (a, b) to H.
4. Consider the graph H constructed in steps 1–3. Compute a maximum matching in H. *(This is to identify "even" posts in H.)*
   – If there exist even posts in Y, then delete all even posts from Y and add them to Z.
   – Else quit the While-loop.

(II) Every $a \in \mathsf{nbr}(Z)$ adds the edge (a, b) to H where b is a's most preferred post in the set Z.

(III) Add all posts in $D = \{\ell(a) : a \in A \text{ and } r_a = \infty\}$ to Y, where $\ell(a)$ is the *dummy* last resort post of applicant a. For every applicant a such that $\mathsf{nbr}(\{a\}) \subseteq X$, add the edge $(a, \ell(a))$ to H.

Note that introducing dummy posts does not interfere with the voting for popular matchings because dummy posts do not vote—they are only present in the "helper" graph H constructed above and not in the given instance G. For any applicant a, being matched to $\ell(a)$ in H is equivalent to a being left unmatched in G. Thus any matching M in H can be projected to a matching in G, by deleting all $(a, \ell(a))$ edges from M. For convenience, we will refer to the resulting matching also as M.

The condition for exiting the While-loop ensures that all posts in Y are odd/unreachable in the subgraph of H with the set of posts restricted to *real* posts in $X \cup Y$ (i.e., the nondummy ones). This implies that all posts in X are also odd/unreachable in this subgraph—this is because if a post $b \in X$ is *even* in this subgraph, then b's neighbor a in this subgraph is odd (by Dulmage–Mendelsohn decomposition). So the applicant a has degree more than 1 and hence it has a neighbor b' in the set Y. Note that then b' has to be *even* in this subgraph, otherwise there would be no odd length alternating path from an unmatched vertex to a in any maximum matching here (recall that every applicant has degree at most 2 in H).

Thus all posts in $X \cup Y$ are odd/unreachable in the subgraph of H with the set of posts restricted to the nondummy ones. So starting with a maximum matching in this subgraph and augmenting it after adding the edges on posts in Z in phase (II) and the edges on dummy posts in phase (III), we get a maximum matching in H that matches all real posts in $X \cup Y$. After the construction of H, our algorithm for the popular matching problem in G is given below.

– If H admits an A-complete matching, then return one that matches all real posts in $X \cup Y$; else output "G has no popular matching."

In the rest of this section, we prove the following theorem.

THEOREM 5. *G admits a popular matching if and only if H admits an A-complete matching, i.e., one that matches all vertices in A.*

3.1. Some examples. We present some examples in Figure 2 and describe how our algorithm builds the graph H on these examples. Let X_i, Y_i, Z_i denote the sets X, Y, Z at the end of the ith iteration of our algorithm and let H_i denote the graph H in step 4 of the ith iteration of our algorithm.

In the first example (top left of Figure 2), we have $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2, b_3\}$ and the preferences of applicants are denoted on the edges. By our initialization, we have $X_0 = \{b_1\}$, $Y_0 = \{b_2, b_3\}$, and $Z_0 = \emptyset$. In step 4 of our

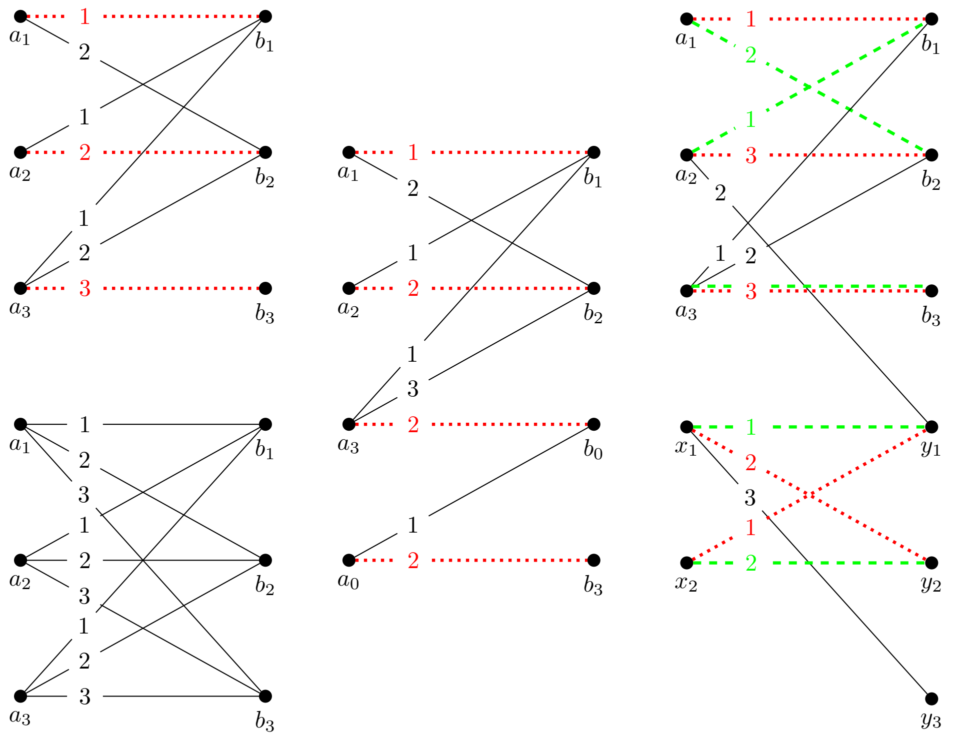

FIG. 2. *We have four examples here: except for the graph in bottom left, all the other graphs admit popular matchings and these are highlighted. In the graph on the extreme right, both the red dotted and green dashed matchings are popular. However, the matching* $\{(a_1, b_1), (a_2, b_2), (a_3, b_3), (x_1, y_1), (x_2, y_2)\}$ *in their union is* not *popular. (Figure appears in color online.)*

first iteration, we identify b_3 as an even post in H_1. So $Y_1 = \{b_2\}$ and $Z_1 = \{b_3\}$. In the second iteration, $a_3 \in \mathsf{nbr}(Z_1)$ and so it has no edge to b_1 in H_2. This is the last iteration of our algorithm. Our final graph H has the edge set $\{(a_1, b_1), (a_2, b_1), (a_1, b_2), (a_2, b_2), (a_3, b_2), (a_3, b_3)\}$.

While the above example admits a popular matching, consider the graph in the bottom left of Figure 2. The first iteration of our algorithm is exactly the same on this graph as it was with the earlier graph. We have $X_1 = \{b_1\}$, $Y_1 = \{b_2\}$, and $Z_1 = \{b_3\}$. However, in the second iteration all the applicants a_1, a_2, a_3 become elements of $\mathsf{nbr}(Z_1)$ and b_1 becomes an isolated vertex in step 2, so b_1 becomes an element of Y_2. In step 4 of the second iteration, b_2 is identified as an even post in H_2 as it is isolated in H_2. So $Y_2 = \{b_1\}$ and $Z_2 = \{b_2, b_3\}$. No demotions happen in the third iteration, which is the last iteration of our algorithm. Our final graph H has the edge set $\{(a_1, b_1), (a_2, b_1), (a_3, b_1), (a_1, b_2), (a_2, b_2), (a_3, b_2)\}$. Observe that H has no A-complete matching.

In the third example (middle of Figure 2), we have $A = \{a_0, a_1, a_2, a_3\}$ and $B = \{b_0, b_1, b_2, b_3\}$ and the preferences of applicants are again denoted on the edges. In step 4 of the first iteration of this algorithm, the post b_3 is identified as an even vertex in Y_0 and it becomes an element of Z_1. So $a_0 \in \mathsf{nbr}(Z_1)$ and b_0 becomes isolated in step 2 of the second iteration. So b_0 becomes an element of Y_2 and this is the last iteration of our algorithm. Our final graph H has the edge set

$\{(a_1, b_1), (a_2, b_1), (a_3, b_1), (a_1, b_2), (a_2, b_2), (a_3, b_0), (a_0, b_0), (a_0, b_3)\}$. This graph admits an A-complete matching $\{(a_1, b_1), (a_2, b_2), (a_3, b_0), (a_0, b_3)\}$.

The fourth example here (the rightmost graph in Figure 2) is that of a graph G with several popular matchings. It is not the case that H contains all these matchings. At the end of our entire algorithm, we have $X = \{b_1, y_1\}$, $Y = \{b_2, y_2\}$, and $Z = \{b_3, y_3\}$. The graph H does not contain the edges (a_3, b_1) and (x_1, y_1) since a_3 and x_1 belong to $\mathsf{nbr}(Z)$. The subgraph H admits an A-complete matching $M = \{(a_1, b_1), (a_2, b_2), (a_3, b_3), (x_2, y_1), (x_1, y_2)\}$ and this is a popular matching in G. However, H does not contain $M' = \{(a_1, b_2), (a_2, b_1), (a_3, b_3), (x_1, y_1), (x_2, y_2)\}$, which is another popular matching in G. In fact, any subgraph that contains both M and M' would also contain the following A-complete matching $N = \{(a_1, b_1), (a_2, b_2), (a_3, b_3), (x_1, y_1), (x_2, y_2)\}$, which is *not* popular. This is because the matching $N' = \{(a_1, b_1), (a_2, y_1), (a_3, b_2), (x_1, y_3), (x_2, y_2)\}$ is more popular than N: Observe that the vertices a_2, a_3, and y_3 prefer N' to N and the vertices x_1 and b_3 prefer N to N' while the remaining vertices are indifferent between the two matchings.

3.2. Proof of Theorem 5: The sufficient part. We first show that if H admits an A-complete matching, then G admits a popular matching. We have already observed that if H admits an A-complete matching, then H has an A-complete matching M that matches all real posts in $X \cup Y$.

A useful observation is that $Z \subseteq B \setminus F$. This is because in step 4 of the While-loop in our algorithm, all f-posts in Y are odd/unreachable in H as they are the only neighbors in H of applicants who regard them as f-posts, i.e., their neighbors have degree 1 in H in step 4.

We now assign edge labels in $\{\pm 1\}$ to all edges in $G \setminus M$ as described at the beginning of section 3, i.e., each edge (a, b) in $G \setminus M$ is labeled $\mathsf{label}(a, b)$ which is a's vote for b versus $M(a)$ and $\mathsf{label}(a, b) = +1$ if a is unmatched in M. Figure 1 is helpful here. For any $U \in \{X, Y, Z\}$, let $M(U) \subseteq A$ be the set of applicants matched in M to posts in U. The following lemma is important.

LEMMA 6. *Every edge of G in $M(X) \times Y$ is labeled -1; similarly, every edge in $M(Y) \times Z$ is labeled -1. Any edge labeled $+1$ has to be either in $M(Y) \times X$ or in $M(Z) \times (X \cup Y)$.*

Proof. Every edge of $\mathsf{nbr}(X) \times X$ that is present in H is a top ranked edge. Since M belongs to H, the edges of M from $\mathsf{nbr}(X) \times X$ are top ranked edges. Thus it is clear that every edge of G in $M(X) \times Y$ is labeled -1. Regarding $M(Y) \times Z$, every edge of $\mathsf{nbr}(Y) \times Y$ that is present in the graph H is an edge (a, b) where the rank of b in a's preference list is $\le r_a$ (i.e. r_a or better); on the other hand, every edge of $\mathsf{nbr}(Z) \times Z$ that is present in the graph G is an edge (a, b') where the rank of b' in a's preference list is $\ge r_a$ (because $b' \in B \setminus F$). Since M belongs to H, every edge of G in $M(Y) \times Z$ is labeled -1.

We now show that any edge labeled $+1$ has to be in either $M(Y) \times X$ or $M(Z) \times (X \cup Y)$ (see Figure 1). Consider any edge $(a, b) \notin M$ such that $b \in U$ and $a \in M(U)$, where $U \in \{X, Y, Z\}$. It follows from the construction of the graph H that a vertex in $\mathsf{nbr}(U)$ can be adjacent in H to only its most preferred post in U. Thus any edge $(a, b) \notin M$ where $b \in U$ and $a \in M(U)$ is labeled -1. We have already seen that all edges in $M(X) \times Y$ and in $M(Y) \times Z$ are labeled -1. There are no edges in $M(X) \times Z$ since $M(X) \subseteq A \setminus \mathsf{nbr}(Z)$. Thus any edge labeled $+1$ has to be in either $M(Y) \times X$ or $M(Z) \times (X \cup Y)$. ∎

Let M' be any matching in G. The symmetric difference of M' and M is denoted by $M' \oplus M$: this consists of alternating paths and alternating cycles—note that edges here alternate between M and M'. Recall that last resort posts are not used in M' (which is a matching in G) whereas last resort posts may be present in M (which is a matching in H).

LEMMA 7. *Consider $M' \oplus M$. The following three properties hold:*

(i) *In any alternating cycle in $M' \oplus M$, the number of -1 edges is at least the number of $+1$ edges.*

(ii) *In any alternating path in $M' \oplus M$, the number of $+1$ edges is at most* two plus *the number of -1 edges; in case one of the endpoints of this path is a last resort post, then the number of $+1$ edges is at most* one plus *the number of -1 edges.*

(iii) *In any even length alternating path in $M' \oplus M$, the number of -1 edges is at least the number of $+1$ edges; in case one of the endpoints of this path is a last resort post, then the number of -1 edges is at least* one plus *the number of $+1$ edges.*

Proof. Property (i). Let $C \in M \oplus M'$ be an alternating cycle. Let C be b_0-a_0-b_1-a_1-b_2-$\cdots$-a_{k-1}-b_0, where $(a_i, b_i) \in M$ for $0 \le i \le k-1$. If C contains no vertex of Z, then there cannot be two consecutive nonmatching edges labeled $+1$ in C. That is, if (a_i, b_{i+1}) is labeled $+1$, then $b_{i+1} \in X$ and there is no $+1$ edge incident on $M(b_{i+1}) = a_{i+1}$, thus the nonmatching edge incident on a_{i+1} in C has to be labeled -1. Hence the number of -1 edges is at least the number of $+1$ edges.

Suppose C contains a vertex of Z: let b_i be such a vertex. There can be two consecutive nonmatching edges labeled $+1$ now: Let b_i-a_i-b_{i+1}-a_{i+1}-b_{i+2} be such an alternating path within C, where both (a_i, b_{i+1}) and (a_{i+1}, b_{i+2}) are labeled $+1$. Then $b_i \in Z$, $b_{i+1} \in Y$, and $b_{i+2} \in X$. In the first place, there is no $+1$ edge incident on a_{i+2} and the crucial part is that there is *no* edge in G between a vertex in $\mathsf{nbr}_H(X)$ and a vertex in Z. Thus once we reach a vertex $a_{i+2} \in M(X)$, we have to see an edge (a_{i+2}, b_{i+3}) labeled -1 where $b_{i+3} \in X \cup Y$ (since a_{i+2} has no neighbor in Z). In order to reach a vertex in Z, we need to see at least *two* consecutive nonmatching edges labeled -1. Thus it again follows that the number of -1 edges is at least the number of $+1$ edges.

Property (ii). Let $\rho \in M \oplus M'$ be an alternating path. Let ρ be b_0-a_0-b_1-a_1-b_2-$\cdots$-a_{k-1}-b_k-a_k, where $(a_i, b_i) \in M$ for $0 \le i \le k$. The same argument that was used in the proof of property (i) shows us that there can be at most two consecutive nonmatching edges labeled $+1$ in ρ and once we traverse such an alternating path b_i-a_i-b_{i+1}-a_{i+1}-b_{i+2} in ρ (where b_i has to be in Z), we are at a vertex $b_{i+2} \in X$. Thereafter, we have to see at least two more nonmatching edges labeled -1 than those labeled $+1$ to again reach a vertex in Z. Thus it follows that the difference between the number of $+1$ edges and the number of -1 edges is at most two.

In fact, for the difference between the number of $+1$ edges and the number of -1 edges to be exactly two, it has to be the case that b_0 is in Z. In case b_0 is in Y, then it is easy to see that the difference between the number of $+1$ edges and the number -1 edges is at most one. Note that all last resort posts belong to Y. Thus when b_0 is a last resort post, then the number of $+1$ edges in ρ is at most *one plus* the number of -1 edges.

Property (iii). Let $\rho = b_0$-a_0-b_1-a_1-b_2-$\cdots$-a_{k-1}-b_k be an even length alternating path where $(a_i, b_i) \in M$ for $0 \le i \le k-1$. The post b_0 is unmatched in M' and b_k is unmatched in M. Recall that M is A-complete, thus any even length alternating path with respect to M has to have vertices in B as its endpoints (since one of them is left unmatched in M). Since b_k is a post that is matched in M' but not in M, it follows that $b_k \in Z$ (as all nondummy posts in $X \cup Y$ are matched in M).

Now the argument is similar to the proof of property (ii). In order to maximize the difference between the number of edges labeled $+1$ and those labeled -1, we assumed that the starting vertex $b_0 \in Z$. For the final vertex b_k to be in Z, it follows that the number of -1 edges is at least the number of $+1$ edges. In particular, when b_0 is a last resort post, then the starting vertex is in Y and so the number of -1 edges is at least *one plus* the number of $+1$ edges. ◻

Lemma 8 uses the above lemma to show the popularity of M. This completes the proof that if H admits an A-complete matching, then G admits a popular matching.

LEMMA 8. *For any matching M' in G, we have $\phi(M, M') \geq \phi(M', M)$.*

Proof. Recall that M is A-complete (where some of the posts used in M can be last resort posts). Consider $M \oplus M'$. We will now investigate every component of $M \oplus M'$—each of which is an alternating cycle, or an odd length alternating path, or an even length alternating path—and show $\phi(M, M') \geq \phi(M', M)$ for each of them.

- For any alternating cycle $C \in M \oplus M'$ among the vertices of C, the difference between those who prefer M' and those who prefer M is equal to $\sum_{a \in C} \mathsf{label}(a, M'(a))$. It follows from part (i) of Lemma 7 that this sum is at most 0.
- Consider any odd length alternating path $\rho \in M \oplus M'$: Its endpoints are an applicant a' and a post b' that are unmatched in M'. Assume b' is a nondummy post. Then among the vertices of ρ that are matched in M', the difference between those who prefer M' and those who prefer M is equal to $\sum_{a \in \rho} \mathsf{label}(a, M'(a))$. It follows from part (ii) of Lemma 7 that this sum is at most 2. The two vertices a' and b' prefer M to M' as they are matched in M and unmatched in M'. Thus summed over all vertices of ρ, the difference between those who prefer M' and those who prefer M is again at most 0.
  Now suppose b' is a dummy post. Then it follows from part (ii) of Lemma 7 that among the vertices of ρ that are matched in M', the difference between those who prefer M' and those who prefer M is at most 1. The vertex a' prefers M to M'. Thus summed over all real vertices of ρ, the difference between those who prefer M' and those who prefer M is again at most 0.
- Consider any even length alternating path $\rho \in M \oplus M'$: Its endpoints are a post b_0 that is unmatched in M' and a post b_k that is unmatched in M. Assume b_0 is a nondummy post. Then summed over all vertices of ρ (this includes b_0 who prefers M and b_k who prefers M'), the difference between those who prefer M and those who prefer M' is at least 0 (by part (iii) of Lemma 7).
  Now suppose b_0 is a dummy post. Then summed over all real vertices of ρ that are matched in M, the difference between those who prefer M and those who prefer M' is at least 1 (by part (iii) of Lemma 7). Thus summed over all real vertices of ρ (this includes b_k who prefers M'), the difference between those who prefer M and those who prefer M' is at least 0.

All vertices whose partners in M and in M' are different belong to some alternating path or cycle in $M \oplus M'$. Hence the difference between the number of vertices that prefer M and those that prefer M' is nonnegative. In other words, $\phi(M, M') \geq \phi(M', M)$. ◻

Bounding the size of M. We know that M is an A-complete matching in H and it matches all real posts in $X \cup Y$. We would now like to bound from below the number of "real edges" in M, i.e., we would like to bound from below the size of the matching obtained after deleting those edges from M that are incident on dummy posts.

We claim that any popular matching in G has size at least $2/3 \cdot |M_{\max}|$, where $M_{\max}$ is a maximum size matching in G. This is because there can be no length-3 augmenting path ρ with respect to any popular matching N in $N \oplus M_{\max}$. Suppose $\rho = b_0$-a_1-b_1-a_0 is such an augmenting path where a_0 and b_0 are unmatched in N. Then by matching a_0 to b_1 and a_1 to b_0, we get a matching $N \oplus \rho$ that is preferred by both a_0 and b_0 while one vertex a_1 prefers N to $N \oplus \rho$. Thus $N \oplus \rho$ is more popular than N, a contradiction to the popularity of N. Hence every augmenting path in $N \oplus M_{\max}$ has length 5 or more. Thus we have $|N| \geq 2/3 \cdot |M_{\max}|$.

The above bound is tight as shown by the example in Figure 3. Here $A = \{a_0, a_1, a_2\}$, $B = \{b_0, b_1, b_2\}$, and let b_1 be the top post of all the three applicants, let b_2 be the second ranked post of both a_1 and a_2, and let b_0 be the third ranked post of a_2. Consider the matching $N = \{(a_1, b_1), (a_2, b_2)\}$. This is a popular matching. In fact, the matching $M = N \cup \{(a_0, \ell(a_0))\} = \{(a_1, b_1), (a_2, b_2), (a_0, \ell(a_0))\}$ is an A-complete matching in H that matches all real posts in $X \cup Y$ (i.e., the posts b_1 and b_2).

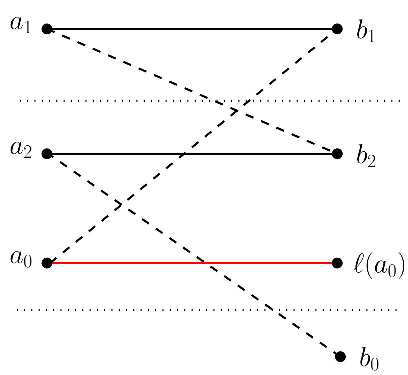

FIG. 3. *The graph H corresponding to the above instance G. Let $N = \{(a_1, b_1), (a_2, b_2)\}$. In the graph G, the matching N has a length-5 augmenting path a_0-b_1-a_1-b_2-a_2-b_0 with respect to it.*

The maximum size matching in G is $M_{\max} = \{(a_0, b_1), (a_1, b_2), (a_2, b_0)\}$. Note that $M_{\max}$ is also a popular matching, thus popular matchings in G can have different sizes. Observe that our algorithm could return either the matching $M_{\max}$ or the matching M since all we ask of our algorithm is to return an A-complete matching in H that matches all real posts in $X \cup Y$. We could easily modify our algorithm so that it always returns an A-complete matching in H that matches all real posts in $X \cup Y$ and the *fewest* number of dummy posts. We will describe this modification in section 3.4 and show the resulting matching to be a max-size popular matching whenever G admits popular matchings.

3.3. Proof of Theorem 5: The necessary part. We now show the other side of Theorem 5. That is, if G admits a popular matching, then H admits an A-complete matching. Let M^* be a popular matching in G. We label the edges of $G \setminus M^*$ by $+1$ or -1 as done at the beginning of section 3: For any edge (a, b) in $G \setminus M^*$, we have $\mathsf{label}(a, b) =$ vote of a for b versus $M^*(a)$. In case a is not matched in M^*, then $\mathsf{label}(a, b) = +1$ for any neighbor b of a.

A crucial property is that there is no alternating path ρ such that the edge labels in $\rho \setminus M^*$ are $\langle +1, +1, +1, \cdots \rangle$, i.e., *no three consecutive nonmatching edges in any alternating path are labeled* $+1$: Note that this is the same as necessary condition (i) stated in the early part of section 3. We will use this property to prove Lemma 9.

LEMMA 9. *If $(a, b) \in M^*$ and $b \in F$, then b has rank better than r_a in a's preference list.*

Proof. Suppose $(a, b) \in M^*$, where $b \in F$, and b has rank worse than r_a in a's preference list. Note that the rank of b cannot be exactly r_a since there is another post $b' \notin F$ that has rank r_a in a's preference list. We know that $a = M^*(b)$ prefers post b' to b. If post b' is unmatched, then consider $M^* \oplus p$ where $p = M^*(a_0)$-a_0-b-a-b', where a_0 is an applicant such that $f(a_0) = b$ (there exists such an applicant since $b \in F$). The matching $M^* \oplus p$ is more popular than M^*.

So suppose the post b' is matched and let $a_1 = M^*(b')$. If $a_0 = a_1$, then consider the alternating cycle $C = a_0$-b-a-b'-a_0; the matching $M^* \oplus C$ makes a_0 and a swap their partners and both applicants prefer $M^* \oplus C$ to M^* while nobody prefers M^* to $M^* \oplus C$. Thus $M^* \oplus C$ is more popular than M^*.

If $a_0 \neq a_1$, then consider the alternating path $\rho = a_0$-b-a-b'-a_1-$f(a_1)$, where $a_1 = M^*(b')$. The path ρ has three consecutive nonmatching edges (a_0, b), (a, b'), $(a_1, f(a_1))$ that are labeled $+1$, hence $M^* \oplus \rho$ is more popular than M^*. Thus we have contradicted the popularity of M^* in all the cases. ∎

Let $p = a_1$-b_1-a_2-b_2-a_3-b_3 and $p' = a'_1$-b'_1-a'_2-b'_2-a'_3-b'_3 be two length-5 alternating paths in G with all their nonmatching edges labeled $+1$. Note that if $b_i = b'_j$ for some $i, j \in \{1, 2, 3\}$, then $i = j$; otherwise, p and p' can be appropriately combined to create an alternating path with three consecutive nonmatching edges labeled $+1$ and we know there is no such alternating path in G.

Based on the matching M^* and the edge labels on $G \setminus M^*$, we partition B into $L_1 \cup L_2 \cup L_3$.

- Roughly speaking, L_3 consists of unwanted posts, so all posts that are unmatched in M^* belong to L_3. Similarly, posts like b_3 with a length-5 alternating path a_1-b_1-a_2-b_2-a_3-b_3 incident on them, with both the nonmatching edges labeled $+1$ (see Figure 4) are in L_3; mid-level posts like b_2 are in L_2 and top posts like b_1 are in L_1.

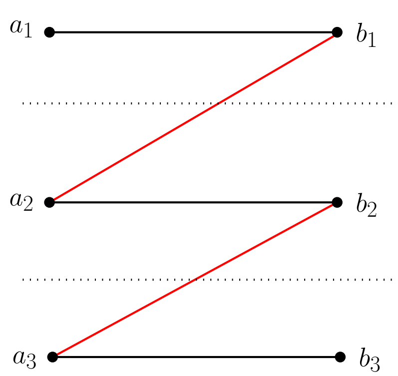

FIG. 4. *A length-5 alternating path a_1-b_1-a_2-b_2-a_3-b_3, where $M^*(b_i) = a_i$, for $i = 1, 2, 3$, and both (a_3, b_2), (a_2, b_1) are labeled $+1$.*

- There cannot be an edge in G between a_1 and b'_3 where a_1-b_1-a_2-b_2-a_3-b_3 and a'_1-b'_1-a'_2-b'_2-a'_3-b'_3 are two length-5 alternating paths with both the nonmatching edges labeled +1—note that this is the same as necessary condition (ii) stated in the early part of section 3. We will maintain this invariant that $M^*(L_1) \cap \mathsf{nbr}(L_3) = \emptyset$ while adding further posts to L_1, L_2, L_3.

More formally, we define the partition $B = L_1 \cup L_2 \cup L_3$ below.

0. Initialize $L_1 = L_2 = \emptyset$ and $L_3 = \{b \in B : b \text{ is unmatched in } M^*\}$. Let all posts that are matched in M^* be *unmarked*. In steps 1–4 we mark these posts and add them to the sets L_1, L_2, L_3 as described below.
1. For each length-5 alternating path $\rho = a_1\text{-}b_1\text{-}a_2\text{-}b_2\text{-}a_3\text{-}b_3$ where $(a_1, b_1), (a_2, b_2), (a_3, b_3) \in M^*$, and both (a_2, b_1) and (a_3, b_2) are labeled +1:
   – add b_i to L_i (if not already there) and mark b_i, for $i = 1, 2, 3$.
2. Repeat the following two steps until there is no unmarked post b to be added to $L_2 \cup L_3$ by these rules:
   – If $M^*(b)$ has no +1 edge incident on it and $M^*(b) \in \mathsf{nbr}(L_3)$, then add b to L_2 and mark b.
   – If $M^*(b)$ has a +1 edge to a vertex in L_2, then add b to L_3 and mark b.
3. For each unmarked b:
   (3.1) if $M^*(b)$ has no +1 edge incident on it then add b to L_1 and mark b;
   (3.2) else add b to L_2 and mark b.

Remark. If a post b is unmarked at the end of step 2, then there are two subcases:
– either $M^*(b)$ has no +1 edge incident on it and $M^*(b) \notin \mathsf{nbr}(L_3)$,
– or $M^*(b) = a$ has a +1 edge (a, b') incident on it and $b' \notin L_2$.

In the first subcase above, b will get added to L_1 in (3.1) and in the second subcase above, b will get added to L_2 in (3.2). Since every post $b \in B$ gets added to exactly one of L_1, L_2, L_3, steps 0-3 obtain a partition $\langle L_1, L_2, L_3 \rangle$ of B.

LEMMA 10. *We have $L_1 \subseteq F \subseteq L_1 \cup L_2$, where F is the set of top posts.*

Proof. We will first show that every post in L_1 is an f-post. Posts are added to L_1 in steps 1 and 3.1. Regarding posts added to L_1 in step 3.1, it follows from the description of step 3.1 that there is no +1 edge incident on the partner of such a post. Let b be any post added to L_1 in step 1. Note that there is no +1 edge incident to $M^*(b)$. Otherwise, we would have an alternating path ρ with respect to M^* with three consecutive nonmatching edges that are labeled +1 (which is forbidden since M^* is popular). Thus $L_1 \subseteq F$.

We will now show that every f-post belongs to either L_1 or L_2. Suppose $b_1 = f(a_0)$ belongs to L_3. The post b_1 has to be matched in M^*. Let $a_1 = M^*(b_1)$ and we also know from the construction of the set L_3 that there is an edge (a_1, b_2) with $b_2 \in L_2$ that is labeled +1. If the vertex a_0 is unmatched in M^*, then by promoting a_0 to b_1 and a_1 to b_2, and leaving $M^*(b_2)$ unmatched, we obtain a matching that is more popular than M^*.

Hence let us assume that a_0 is matched in M^*. Consider the alternating path $M^*(a_0)\text{-}a_0\text{-}b_1\text{-}a_1\text{-}b_2$ with respect to M^*: This has two consecutive nonmatching edges that are labeled +1. Thus it follows from our construction of L_1, L_2, L_3 that $M^*(a_0) \in L_3$, $b_1 \in L_2$, and $b_2 \in L_1$. This contradicts our assumption that $b_1 \in L_3$. ∎

LEMMA 11. $M^*(L_1) \cap \mathsf{nbr}(L_3) = \emptyset$.

Before we prove the above lemma, we will show the following claim which will be useful in proving Lemma 11.

CLAIM 1. *If $a \in \mathsf{nbr}(L_3)$ and $M^*(a) = f(a)$, then there is an alternating path ρ_a with respect to M^* with a as an endpoint such that either $|\rho_a|$ is even and the edge labels on $\rho_a \setminus M^*$ are $\langle -1, +1, -1, \ldots, +1, +1 \rangle$ or $|\rho_a|$ is odd and the edge labels on*

$\rho_a \setminus M^*$ *are* $\langle -1, +1, -1, \ldots, +1, -1 \rangle$, *where the last edge is incident on an unmatched post.*

Proof. Posts are added to L_3 in steps 0, 1, and 2. We now study each of these cases. The set L_3 was initialized to the set of posts left unmatched in M^*. So at the end of step 0, it is the case that every $a \in \mathsf{nbr}(L_3)$ has an odd length alternating path, which is in fact an edge (a, b) labeled -1, whose one endpoint is a and the other endpoint is an unmatched post b.

Let b_3 be a post that got added to L_3 in step 1. Then there is an alternating path b_3-a_3-b_2-a_2-b_1-a_1 such that $(a_i, b_i) \in M^*$ for $i = 1, 2, 3$, and both (a_3, b_2) and (a_2, b_1) are marked $+1$. Thus every neighbor $a \in \mathsf{nbr}(\{b_3\})$ with $M^*(a) = f(a)$ has an even length alternating path $\rho_a = a$-b_3-a_3-b_2-a_2-b_1-a_1 where the edge labels on $\rho_a \setminus M^*$ are $\langle -1, +1, +1 \rangle$. Note that $a \neq a_1$. Otherwise, ρ_a is an alternating cycle and $M^* \oplus \rho_a$ is more popular than M^*.

Thus the claim that every $a \in \mathsf{nbr}(L_3)$ with $M^*(a) = f(a)$ has a desired alternating path ρ_a is true at the end of step 1. Let b_3 be a post that got added to L_3 in step 2 and let us assume that till the point b_3 gets added to L_3, the claim holds. Since b_3 was added to L_3 in step 2, this was due to a $+1$ edge between $a_3 = M^*(b_3)$ and a post $b_2 \in L_2$ whose partner $a_2 = M^*(b_2)$ regards b_2 as a top post. The post $b_2 \in L_2$ because its partner $a_2 \in \mathsf{nbr}(L_3)$. This means there is a desired alternating path ρ_{a_2} incident on a_2. Neither b_3 nor b_2 lies on ρ_{a_2} since all the posts in ρ_{a_2} that belong to $L_2 \cup L_3$ were added to $L_2 \cup L_3$ prior to b_2 joining L_2 and b_3 joining L_3. Consider any neighbor a of b_3 that is in $\mathsf{nbr}(L_3)$ because $b_3 \in L_3$ and $M^*(a) = f(a)$. The desired alternating path ρ_a is a-b_3-a_3-b_2-a_2 followed by ρ_{a_2}. ◻

Proof of Lemma 11. We will use Claim 1 to show that $M^*(L_1) \cap \mathsf{nbr}(L_3) = \emptyset$. Posts get added to L_1 in steps 1 and 3.1 of the partition scheme. Let b_1 be a post that got added to L_1 in step 1, then there is an alternating path $p = a_3$-b_2-a_2-b_1-a_1 where both (a_3, b_2) and (a_2, b_1) are labeled $+1$. It has to be the case that $b_1 = f(a_1)$ (otherwise there would be an alternating path with respect to M^* with three consecutive nonmatching edges labeled $+1$); if $a_1 \in \mathsf{nbr}(L_3)$, then there is an alternating path ρ_{a_1} as described in Claim 1. If the applicants a_2 and a_3 do not appear in ρ_{a_1}, then consider the alternating path p' which consists of p followed by ρ_{a_1}. It is easy to see that $M^* \oplus p'$ is more popular than M^*: A contradiction to the popularity of M^*.

In case a_2 appears in ρ_{a_1}, then we have an alternating cycle C, which is ρ_{a_1} truncated until the vertex a_2 followed by a_2-b_1-a_1. This cycle has a stretch of alternating -1 and $+1$ labeled nonmatching edges along with two consecutive nonmatching edges labeled $+1$: these are the edge (a_2, b_1) and the edge incident on b_2 in ρ_{a_1} from a vertex in $M^*(L_3)$. Thus $M^* \oplus C$ is more popular than M^*: A contradiction again. If a_3 appears in ρ_{a_1}, then we can again construct an alternating cycle C' (using the $a_1 \rightsquigarrow a_3$ subpath of ρ_{a_1} followed by the alternating path p). The matching $M^* \oplus C'$ is more popular than M^* since C' has more $+1$ labeled nonmatching edges than -1 labeled nonmatching edges. This again contradicts the popularity of M^*.

Regarding any post b added to L_1 in step 3.1, as observed in the remark below our partition scheme, we have $M^*(b) \notin \mathsf{nbr}(L_3)$. This completes the proof that $M^*(L_1) \cap \mathsf{nbr}(L_3) = \emptyset$. ◻

We will use the partition $\langle L_1, L_2, L_3 \rangle$ of B to build the following subgraph $G' = (A \cup B, E')$ of G. For each $a \in A$, include the following edges in E':

(i) If $a \notin \mathsf{nbr}(L_3)$, then add the edge $(a, f(a))$ to E'.

(ii) If a has a neighbor of rank $\leq r_a$ in L_2, then add the edge (a, b) to E', where

b is a's most preferred neighbor in L_2.

(iii) If $a \in \mathsf{nbr}(L_3)$, then add the edge (a, b) to E', where b is a's most preferred neighbor in L_3.

LEMMA 12. *Every edge of the matching M^* belongs to the graph G'.*

Proof. The set B has been partitioned into $L_1 \cup L_2 \cup L_3$. We will now show that for each post b_0 that is matched in M^*, the edge $(M^*(b_0), b_0)$ belongs to G'. We distinguish three cases: $b_0 \in L_1$, $b_0 \in L_2$, and $b_0 \in L_3$.

Case 1. The post $b_0 \in L_1$. Hence there is no $+1$ edge incident on $a_0 = M^*(b_0)$, in other words, $b_0 = f(a_0)$. Lemma 11 tells us that $M^*(L_1) \cap \mathsf{nbr}(L_3) = \emptyset$; hence a_0 has no neighbor in L_3, and by rule (i) above, the edge $(a_0, f(a_0)) = (a_0, b_0)$ belongs to the edge set of G'.

Case 2. Next we consider the case when $b_0 \in L_2$. It is easy to see that b_0 has to be a_0's most preferred post in L_2, where $a_0 = M^*(b_0)$. Otherwise, there would have been an edge (a_0, b_1) labeled $+1$ with $b_1 \in L_2$, where b_1 is a_0's most preferred post in L_2. Then either $b_1 \in L_1$ or $b_0 \in L_3$ (from how we construct the sets L_1, L_2, L_3), a contradiction. We now have to show that the rank of b_0 in a_0's preference list is $\leq r_a$. Otherwise, the edge (a_0, b_0) does not belong to G'.

Suppose $b_0 \in F$. Since the edge $(a_0, b_0) \in M^*$, which is a popular matching, it follows from Lemma 9 that b_0 is ranked better than r_{a_0} in a_0's preference list; thus the edge (a_0, b_0) belongs to G'. So the case left is when $b_0 \notin F$. If b_0 is not a_0's most preferred post outside F, then there is the length-5 alternating path $\rho = b_0$-a_0-b_1-a_1-$f(a_1)$-$M^*(f(a_1))$, where b_1 is the most preferred post of a_0 outside F and $a_1 = M^*(b_1)$. The alternating path ρ has two consecutive nonmatching edges (a_0, b_1) and $(a_1, f(a_1))$ that are labeled $+1$. This contradicts the presence of b_0 in L_2 as such a post would have to be in L_3. Thus if $b_0 \notin F$, then b_0 has to be a_0's most preferred post outside F, i.e., b_0 has rank r_{a_0} in a_0's preference list.

Case 3. We finally consider the case when the post $b_0 \in L_3$. We need to show that b_0 is the most preferred post of $a_0 = M^*(b_0)$ in L_3. Suppose not. Let b_1 be a_0's most preferred post in L_3. The post b_1 has to be matched in M^*, otherwise $M^* \oplus \rho$ is more popular than M^*, where $\rho = b_0$-a_0-b_1. Since $b_1 \in L_3$ while $F \cap L_3 = \emptyset$ (by Lemma 10), we know that there is an edge labeled $+1$ incident on $a_1 = M^*(b_1)$. Let this edge be (a_1, b_2) and let a_2 be $M^*(b_2)$. So there is a length-5 alternating path $p = b_0$-a_0-b_1-a_1-b_2-a_2 where both the nonmatching edges (a_0, b_1) and (a_1, b_2) are labeled $+1$. This contradicts the presence of b_1 in L_3 as such a post would have to be in L_2. Thus b_0 is a_0's most preferred post in L_3. ∎

The following lemma shows the relationship between the partition $\langle L_1, L_2, L_3 \rangle$ and the partition $\langle X, Y, Z \rangle$ constructed by our algorithm that builds the graph H.

LEMMA 13. *The set $X \supseteq L_1$ and the set $Z \subseteq L_3$.*

Proof. In our algorithm that constructs the graph H and also the partition $\langle X, Y, Z \rangle$, the set X is initialized to F and the set Y is initialized to $B \setminus F$. As our algorithm progresses, in each iteration of the While-loop, some f-posts get demoted from X to Y and similarly, some non-f-posts get demoted from Y to Z until there is an iteration (say, iteration $h + 1$) where all posts in Y are odd/unreachable in H—this is the last iteration of the While-loop.

For any $1 \leq k \leq h+1$, let T_k (similarly, F_k) be the set of posts that got demoted from Y to Z (resp., X to Y) in the k-th iteration of the While-loop in our algorithm. We have $T_{h+1} = \emptyset$.

Note that $F_1 = \emptyset$ since Z is initialized to $\emptyset$, so in the first iteration of our algorithm, *every* f-post b has a neighbor $a \in A \setminus \mathsf{nbr}(Z)$ such that $f(a) = b$. Thus no post is demoted from X to Y in the first iteration.

The graph H_1 is the subgraph of G where each $a \in A$ has at most two neighbors: its top post and when $r_a < \infty$, its neighbor of rank r_a. The set T_1 is the set of even non-f-posts in H_1. We will use the following claims and finish the proof of this lemma (the proofs of Claims 2–4 are given after the proof of Lemma 13).

CLAIM 2. *The set $T_1 \subseteq L_3$.*

CLAIM 3. *For any $1 \le k \le h$, if $\bigcup_{i=1}^{k} T_i \subseteq L_3$, then $F_{k+1} \subseteq L_2$.*

CLAIM 4. *For any $2 \le k \le h$, if $\bigcup_{i=2}^{k} F_i \subseteq L_2$, then $T_k \subseteq L_3$.*

Claim 2 tells us that $T_1 \subseteq L_3$. We now use Claims 3 and 4 alternately to conclude that for every $1 \le k \le h$, we have $\cup_{i=2}^{k+1} F_i \subseteq L_2$ and $\cup_{i=1}^{k} T_i \subseteq L_3$.

Thus the set $Z = \cup_{i=1}^{h} T_i$ is a subset of L_3 and the set $F \setminus X = \cup_{i=2}^{h+1} F_i$ is a subset of L_2. Since $F \setminus X \subseteq L_2$, it follows that $X \supseteq F \setminus L_2$. We know that $F \setminus L_2 = L_1$ (by Lemma 10), thus $X \supseteq L_1$. ◻

Proof of Claim 2. Any post in T_1 that is left unmatched in M^* has to belong to L_3. Similarly, any $b_0 \in T_1$ that is matched in M^* to an applicant a_0 that ranks b_0 worse than r_{a_0} has to belong to L_3: This is because there is a length-5 alternating path $p = b_0\text{-}a_0\text{-}b_1\text{-}a_1\text{-}b_2\text{-}a_2$ where b_1 is a post of rank r_{a_0} in a_0's preference list, $a_1 = M^*(b_1)$, and $b_2 = f(a_1)$. The path p has two consecutive nonmatching edges that are labeled $+1$, so $b_0 \in L_3$.

Now consider any $b_0 \in T_1$ that is matched in M^* to an applicant a_0 such that the rank of (a_0, b_0) is r_{a_0}. So a_0 is a neighbor of b_0 in H_1. Since b_0 is even in H_1, all the neighbors of b_0 in H_1 are odd and thus they have to be of degree *exactly* 2 in H_1 (recall that all applicants have degree at most 2 in H_1). Thus the neighbors of these applicants are again even. Let C be the connected component containing b_0 in H_1. It is easy to see that in C, all posts are even, all applicants are odd and have degree exactly 2, and the number of posts is more than the number of applicants. (In fact, C is a tree with b_0 as the root and the number of posts in C is *one* plus the number of applicants in C.)

If $b_0 \in L_2$, then a_0's other neighbor in C, which is $f(a_0)$, has to be in L_1 since there is a $+1$ edge from a_0 to $f(a_0)$. This means $f(a_0)$ is matched to an applicant a_0' that ranks it as a top post, so the applicant a_0' is a neighbor of $f(a_0)$ in C. There has to be another neighbor of a_0' in C; call this vertex b_1. The important observation is that b_1 cannot be in L_3 as that would violate Lemma 11 since $a_0' \in M^*(L_1)$. So $b_1 \in L_2$ and this means b_1 is matched to an applicant a_1 that ranks it r_{a_1}, in other words, a_1 is a neighbor of b_1 in C. So $f(a_1)$ has to be in L_1 and we continue in this manner marking all f-posts in C as elements of L_1 and all non-f-posts in C as elements of L_2.

This means all posts in C are matched to their neighbors in C. However, this is not possible as there are more posts than applicants in C. This contradicts our assumption that $b_0 \in L_2$, in other words, b_0 has to be in L_3. Thus $T_1 \subseteq L_3$. ◻

Proof of Claim 3. The set F_{k+1} is the set of posts that got demoted from X to Y in the $(k+1)$th iteration of the While-loop: This means each post b in F_{k+1} had no applicant outside $\mathsf{nbr}(\cup_{i=1}^{k} T_i)$ that regarded b as an f-post. In other words, every applicant a such that $f(a) = b$ belongs to $\mathsf{nbr}(\cup_{i=1}^{k} T_i)$. Since $\cup_{i=1}^{k} T_i \subseteq L_3$, each such applicant a is present in $\mathsf{nbr}(L_3)$.

Let $F_{k+1} = \{b_1, \dots, b_h\}$. For $1 \le i \le h$, let $(a_i, b_i) \in M^*$: If $f(a_i) = b_i$, then $b_i \in L_2$ (because $a_i \in \mathsf{nbr}(L_3)$); else there is an edge $(a_i, f(a_i))$ that is labeled $+1$ incident on a_i and hence b_i cannot be in L_1. Thus $F_{k+1} \cap L_1 = \emptyset$, i.e., $F_{k+1} \subseteq L_2$ (by Lemma 10). ∎

Proof of Claim 4. Let us assume that we have proved Claim 4 for all smaller values of k. That is, for $j \le k-1$, we have shown that if $\cup_{i=2}^{j} F_i \subseteq L_2$, then the set $T_j \subseteq L_3$. This is indeed the case for $k = 2$ since we know $T_1 \subseteq L_3$ (by Claim 2). Using Claims 3 and 4 (for $j \le k-1$) alternately now, it follows that $T_j \subseteq L_3$ for $j \le k-1$. Thus $\cup_{i=1}^{k-1} T_i \subseteq L_3$. We will now show that T_k is also a subset of L_3.

Let H_k denote the graph H in step 4 in the kth iteration of the While-loop in our algorithm. This is the graph where we determine the even posts that will get demoted from Y to Z. In step 4 of the kth iteration of the While-loop, the set $X = F \setminus \cup_{i=2}^{k} F_i$ (call this set X_k), $Z = \cup_{i=1}^{k-1} T_i$ (call this set Z_k), and let Y_k be the set of posts outside $X_k \cup Z_k$. The edge set of H_k is as follows:

– for each $a \in A$: If the rank of a's most preferred post b in Y_k is $\le r_a$, then the edge (a, b) belongs to H_k.
– for $a \in A \setminus \mathsf{nbr}(Z_k)$: The edge $(a, f(a))$ is also present in H_k.

Let S be the set of non-f-posts that are odd/unreachable in the graph H_1. Let us refer to posts in S as *s-posts*. We will now show that all s-posts in L_2 are odd/unreachable in H_k; so every s-post that is even in H_k has to be in L_3, in other words, $T_k \subseteq L_3$. Let G'_0 be the subgraph of G' with the set of posts restricted to $L_1 \cup L_2$ (see Figure 5). Consider the subgraph G'_k of G'_0 obtained by deleting edges missing in H_k from G'_0.

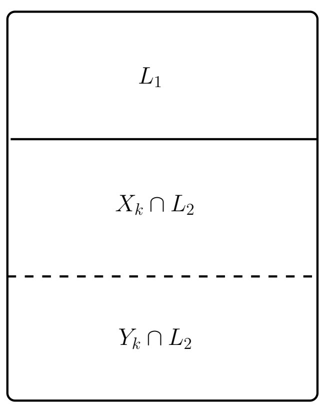

FIG. 5. *The set of posts in G'_0 can be viewed as $L_1 \cup (X_k \cap L_2) \cup (Y_k \cap L_2)$. All s-posts in L_2 are in $Y_k \cap L_2$.*

We now show that G'_k contains all edges in G'_0 incident on s-posts in L_2, by showing that any edge (a, b) incident on an s-post $b \in L_2$ in G'_0 is present in H_k also. Since the edge (a, b) belongs to G'_0, the post b has to be ranked r_a in a's preference list and moreover, there is no f-post in L_2 of rank better than r_a in a's list. If the edge (a, b) does not exist in H_k, then it means there is some f-post in Y_k that a prefers to b. All f-posts in Y_k are in $\cup_{i=2}^{k} F_i$ and we are given that $\cup_{i=2}^{k} F_i \subseteq L_2$. Since we know there is no f-post in L_2 that a prefers to b, it follows that b has to be a's most preferred post in Y_k and so the edge (a, b) belongs to H_k. Thus G'_k, whose edge set is the intersection of the edge sets of G'_0 and H_k, contains all edges in G'_0 incident on s-posts in L_2.

Every post in $L_1 \cup L_2$ is odd/unreachable in G'_0 since the matching M^* restricted to the edge set of G'_0 is $(L_1 \cup L_2)$-complete. We have shown that G'_k contains all edges in G'_0 incident on s-posts in L_2: Thus all s-posts in L_2 are odd/unreachable in G'_k. It is easy to see that all *top ranked* edges in G'_0 incident on f-posts in $Y_k \cap L_2$ are also present in G'_k: each such post has a degree 1 neighbor in G'_k, thus all f-posts in $Y_k \cap L_2$ are also odd/unreachable in G'_k.

We now claim that all posts in L_1 are also odd/unreachable in G'_k. We first show that all edges incident on L_1 in G'_0 are present in H_k. This is because each edge (a, b) in G'_0 such that $b \in L_1$ is incident on an applicant $a \in A \setminus \mathsf{nbr}(L_3)$ such that $b = f(a)$ and we know the graph H_k has $(a, f(a))$ edges for all $a \in A \setminus \mathsf{nbr}(Z_k) \supseteq A \setminus \mathsf{nbr}(L_3)$ since $Z_k = \cup_{i=1}^{k-1} T_i \subseteq L_3$.

In G'_k, each vertex $b \in L_1$ either has a degree 1 neighbor (in which case our claim is true) or it has a degree 2 neighbor a whose other neighbor is in $Y_k \cap L_2$, i.e., it is not in $X_k \cap L_2$. This is because a cannot have two neighbors in X_k in the graph H_k and we know $L_1 \subseteq X_k$ since all f-posts missing in X_k (these are posts in $\cup_{i=2}^{k-1} F_i$) are absent from L_1 also. Since all posts in $Y_k \cap L_2$ are odd/unreachable in G'_k, it follows that all posts in L_1 are also odd/unreachable in G'_k.

Let us now compare the graph H_k with the graph G'_k. The graph H_k has additional vertices: These are the ones in $Y_k \cap L_3$ and the new edges in H_k (new relative to G'_k) belong to the following two classes: (i) $\mathsf{nbr}(L_3) \times (Y_k \cap L_3)$ and (ii) $A \times (L_1 \cup (Y_k \cap L_2))$. This is because every edge incident on $X_k \cap L_2$ in H_k (these are all top-ranked edges) is present in G'_0 as well.

Consider any new edge (a, b) in H_k of type (i), i.e., $(a, b) \in \mathsf{nbr}(L_3) \times (Y_k \cap L_3)$. Since (a, b) belongs to H_k, it must be the case that a's most preferred neighbor in Y_k is b. So the post b is ranked r_a in a's list and a has no neighbor of rank better than r_a in Y_k. Recall that G'_0 has no edge in $\mathsf{nbr}(L_3) \times L_1$. So the only edge that can be incident on a in the graph G'_k is an edge to $f(a)$ in $X_k \cap L_2$.

Consider any connected component C in G'_k that contains an s-post in L_2: Every post here belongs to either L_1 or $Y_k \cap L_2$, in other words, there is no post in $X_k \cap L_2$ here. This is because there is no applicant a in G'_k with neighbors in $Y_k \cap L_2$ and $X_k \cap L_2$ as this means a has two neighbors in L_2, which is forbidden in G'_0. Similarly, there is no applicant a' in G'_k with neighbors in L_1 and $X_k \cap L_2$ as this means a has two neighbors in X_k, which is forbidden in H_k. Thus C has no post from $X_k \cap L_2$.

So the new edges in H_k of type (i) do not touch components in G'_k that contain s-posts in L_2. All the new edges incident upon these components have their endpoints in $L_1 \cup (Y_k \cap L_2)$. These posts are already odd/unreachable in G'_k. So these posts remain odd/unreachable in H_k. Hence every s-post in L_2 is odd/unreachable in H_k. This completes the proof of Claim 4. ◻

The augmented graph G'. The matching M^* need not be A-complete. However, it would help us to assume that M^* is A-complete, so we augment M^* by adding $(a, \ell(a))$ edges for every $a \in A$ that is unmatched in M^*. Recall that $\ell(a)$ is the dummy last resort post of a. However, the augmented matching M^* need not belong to the graph G' any longer—hence we augment G' also by adding some dummy vertices and some edges as described below.

The augmentation of G' is analogous to phase (III) of our algorithm—we augment G' as follows: Let $L_2 = L_2 \cup D$, where $D = \{\ell(a) : a \in A \text{ and } r_a = \infty\}$; if $\mathsf{nbr}(\{a\}) \subseteq L_1$, then add $(a, \ell(a))$ to G'. Thus when compared to G', the augmented G' has some new vertices (all these are dummy last resort posts) and some new edges—each new edge is of the form $(a, \ell(a))$ where $\ell(a)$ is a's only neighbor in $L_2 \cup L_3$. These new edges are enough to show the following lemma.

LEMMA 14. *The augmented matching M^* belongs to the augmented graph G'.*

Proof. Before the augmentations of G' and M^*, the matching M^* belonged to the graph G' (by Lemma 12). We now need to show that if a is left unmatched in M^* (before augmentation), then $r_a = \infty$ and all of a's neighbors belong to L_1.

Suppose a is left unmatched in M^* and $r_a < \infty$. Since $r_a < \infty$, there is a post $b \notin F$ such that the post b has rank r_a in a's preference list. Consider the alternating path $p = a\text{-}b\text{-}a'\text{-}f(a')\text{-}a''$, where $a' = M^*(b)$ and $a'' = M^*(f(a'))$. The matching $M^* \oplus p$ matches a to b and promotes a' to its top post $f(a')$ and leaves a'' unmatched. Thus $M^* \oplus p$ is more popular than M^*, a contradiction.

So let us assume $r_a = \infty$ and a was left unmatched in M^*. Suppose a has some neighbor b_0 outside L_1. The post b_0 has to be in F because $r_a = \infty$, i.e., a has no neighbors outside F. Since $F \subseteq L_1 \cup L_2$ (by Lemma 10), it follows that $b_0 \in L_2$. Let $a_0 = M^*(b_0)$; if $b_0 \neq f(a_0)$, then we again have an alternating path $p = a\text{-}b_0\text{-}a_0\text{-}f(a_0)\text{-}a_1$, where $a_1 = M^*(f(a_0))$ such that $M^* \oplus p$ is more popular than M^*. This contradicts the popularity of M^*.

So suppose $b_0 = f(a_0)$ and $b_0 \in L_2$ because $a_0 \in \mathsf{nbr}(L_3)$. We know from Claim 1 that there is a desired alternating path ρ_{a_0}, where either the last two nonmatching edges are labeled $+1$ or the last post in ρ_{a_0} is unmatched. Consider the alternating path ρ which is the path $a\text{-}b_0\text{-}a_0$ followed by the path ρ_{a_0}. It is easy to see that $M^* \oplus \rho$ is more popular than M^*, a contradiction to the popularity of M^*. ∎

Since the augmented M^* is an A-complete matching, it follows from Lemma 14 that the augmented graph G' admits an A-complete matching. Theorem 15 uses Lemma 13 to show that if the augmented graph G' admits an A-complete matching, then so does the graph H constructed by our algorithm.

THEOREM 15. *If H does not admit an A-complete matching, then the augmented graph G' cannot admit an A-complete matching.*

Proof. We will use G' to refer to the *augmented* graph G' in this proof. The rules for adding edges in H and in G' are exactly the same—the only difference is in the partition $\langle X, Y, Z\rangle$ on which H is based versus the partition $\langle L_1, L_2, L_3\rangle$ on which G' is based. If $\langle X, Y, Z\rangle = \langle L_1, L_2, L_3\rangle$, then the graphs H and G' are exactly the same.

Figure 6 denotes how the partition $\langle X, Y, Z\rangle$ can be modified to the partition $\langle L_1, L_2, L_3\rangle$. We know from Lemma 13 that $X \supseteq L_1$ and $Z \subseteq L_3$. Consider the subgraph G'' of G' induced on the vertex set $A' = (A \setminus \mathsf{nbr}(Z)) \cup (\mathsf{nbr}(Z) \cap \mathsf{nbr}_H(Y \cap L_3))$ and $B' = X \cup Y$. This is the part bounded by the box in Figure 6. In our analysis, we can essentially separate G' into G'' and the part outside G'' due to the following claim that says G' has no edges between A' and Z.

CLAIM 5. *G' has no edge (a, b) where $a \in A'$ and $b \in Z$.*

Proof. Any applicant $a \in A'$ has to belong to either $A \setminus \mathsf{nbr}(Z)$ or to $\mathsf{nbr}(Z) \cap \mathsf{nbr}_H(Y \cap L_3)$ (see Figure 6). There is obviously no edge in G between a vertex in $A \setminus \mathsf{nbr}(Z)$ and any vertex in Z. So suppose $a \in \mathsf{nbr}(Z) \cap \mathsf{nbr}_H(Y \cap L_3)$. For $b \in L_3$, if the edge (a, b) is in G', then b has to be a's most preferred post in L_3. We will now show that $b \in Y \cap L_3$, equivalently $b \notin Z$. Thus G' has no edge (a, b) where $a \in A'$ and $b \in Z$.

Since $a \in \mathsf{nbr}_H(Y \cap L_3)$, the graph H contains an edge between a and some $b' \in Y \cap L_3$. Recall that an element of $Y \cap L_3$ is a real post in Y. By the rules of including edges in H, it follows that the rank of b' in a's preference list is $\leq r_a$. The

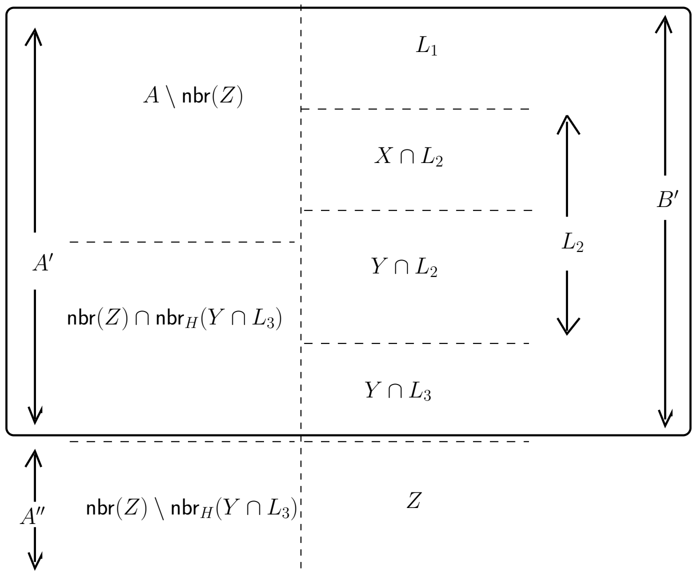

FIG. 6. *The part of G' inside the box will be called G''. We show that the graph G' has no edge between any applicant in A' and any post in Z.*

entire set L_3 cannot contain any post of rank better than r_a for any $a \in A$ since any post of rank better than r_a in a's list belongs to F while $L_3 \cap F = \emptyset$ (by Lemma 10). So b' has rank r_a in a's list. Thus a's most preferred neighbor in L_3 belongs to $Y \cap L_3$. ∎

Let G_0 be the subgraph of G'' obtained by deleting from G'' the edges that are absent in H. Thus G_0 is a subgraph of both G' and H. The following claim (whose proof is given after the proof of Theorem 15) will be useful to us.

CLAIM 6. *All posts in $(X \cap L_2) \cup (Y \cap L_3)$ are odd/unreachable in G_0. Moreover, every edge (a, b) in G' that is missing in H satisfies $b \in (X \cap L_2) \cup (Y \cap L_3)$.*

Consider the graph G_1 whose edge set is the intersection of the edge sets of G' and H. Equivalently, G_1 can be constructed by adding to the edge set of G_0, the edges incident on $A'' = \mathsf{nbr}(Z) \setminus \mathsf{nbr}_H(Y \cap L_3)$ that are present in both G' and H (see Figure 6). This is due to the fact that G' has no edge in $A' \times Z$.

We claim that all posts in $(X \cap L_2) \cup (Y \cap L_3)$ are odd/unreachable in G_1. This is because Claim 6 tells us that each post in this set is odd/unreachable in G_0, and due to the absence of $A' \times Z$ edges in G', the graph G_1 has no *new* edge (new when compared to G_0) incident on the set A' of applicants in G_0. Hence all posts in $(X \cap L_2) \cup (Y \cap L_3)$ remain odd/unreachable in G_1.

Claim 6 also tells us that all edges in G' that are missing in H are incident on posts in $(X \cap L_2) \cup (Y \cap L_3)$. We know that all these posts are odd/unreachable in G_1, hence G' has no *new* edge (new when compared to G_1) on posts that are *even* in G_1. Thus the size of a maximum matching in G' equals the size of a maximum matching in G_1. This is at most the size of a maximum matching in H, since G_1 is a subgraph of H. Hence if H has no A-complete matching, then neither does G'. ∎

Proof of Claim 6. We will now show that all posts in $(X \cap L_2) \cup (Y \cap L_3)$ are odd/unreachable in G_0. Let a be an applicant with degree 2 in the graph G_0 and let

b_1 and b_2 be the two neighbors of a, where b_1 is the more preferred neighbor of a. We claim either (i) $b_1 \in X \cap L_2$ and $b_2 \in Y \cap L_3$ or (ii) $b_1 \in L_1$ and $b_2 \in Y \cap L_2$. This is due to the following reasons.

- There is no applicant in G_0 with edges to both a post in L_1 and post in $X \cap L_2$. If there was such an applicant a, then a would have two neighbors in the set X, which is forbidden in H. Recall that any edge in G_0 is an edge in H as well.
- There is no applicant in G_0 with edges to both a post in $X \cap L_2$ and a post in $Y \cap L_2$. If there was such an applicant a, then a would have two neighbors in the set L_2, which is forbidden in G'. Recall that any edge in G_0 is an edge in G' as well.
- There is no applicant in G_0 with edges to both a post in $Y \cap L_2$ and a post in $Y \cap L_3$. If there was such an applicant a, then a would have two neighbors in the set Y, which is forbidden in H.
- There is no applicant a in G_0 with edges to both a post in L_1 and a post in $Y \cap L_3$. This is because G' cannot contain such a pair of edges as it is only applicants in $A \setminus \mathsf{nbr}(L_3)$ that are adjacent to posts in L_1.

Thus in the graph G_0, vertices in $(X \cap L_2) \cup (Y \cap L_3)$ and those in $L_1 \cup (Y \cap L_2)$ belong to different connected components. Note that all dummy posts belong to $Y \cap L_2$. So none of these posts belongs to any connected component in G_0 that contains vertices in $(X \cap L_2) \cup (Y \cap L_3)$. Consider the subgraph H' of H, obtained by restricting the set of posts in H to *real* posts in $X \cup Y$. All real posts in $X \cup Y$ are odd/unreachable in H'. Since $(X \cap L_2) \cup (Y \cap L_3)$ consists of real posts, all these posts are odd/unreachable in H'.

We now claim that all posts in $(X \cap L_2) \cup (Y \cap L_3)$ remain odd/unreachable in G_0. In the first place, every edge (a_0, b_0) in H' incident on any vertex $b_0 \in Y \cap L_3$ is present in G'' as well. This is because $a_0 \in A'$ and if $b_0 \in Y \cap L_3$ is the most preferred post in Y for applicant a_0, then the rank of b_0 in a_0's preference list is r_{a_0} and thus b_0 is also a_0's most preferred post in L_3, so the edge (a_0, b_0) belongs to the graph G''. Similarly every edge (a_1, b_1) in H' incident on any post $b_1 \in X \cap L_2$ is present in G'' as well—this is because $a_1 \in A'$ and b_1 has to be $f(a_1)$ for the edge (a_1, b_1) to exist in H'. Thus b_1 is also a_1's most preferred post in L_2. Hence all edges in H' incident on posts in $(X \cap L_2) \cup (Y \cap L_3)$ are present in G_0.

Let b be any post in $(X \cap L_2) \cup (Y \cap L_3)$. The connected component in G_0 that contains b can be obtained by taking the connected component containing b in H' and deleting all vertices in $L_1 \cup (Y \cap L_2)$ from this component. Since no edge incident on b has been deleted here and because b is odd/unreachable in the starting component, it follows that b is odd/unreachable in G_0.

We will now show the second part of Claim 6: Every edge (a, b) in G' that is missing in H satisfies $b \in (X \cap L_2) \cup (Y \cap L_3)$. We partitioned the set B of posts into five sets (refer to Figure 6). These are L_1, $X \cap L_2$, $Y \cap L_2$, $Y \cap L_3$, and Z. We will now show that every edge in G' that is incident on $L_1 \cup (Y \cap L_2) \cup Z$ is present in H also.

- Any edge (a, b) in G' where $b \in L_1$ is such that $f(a) = b$ and $a \in A \setminus \mathsf{nbr}(L_3)$. Since $L_3 \supseteq Z$ (by Lemma 13), this means $a \in A \setminus \mathsf{nbr}(Z)$. Thus H also contains the edge (a, b).
- Any edge (a, b) in G' where $b \in Y \cap L_2$ is such that b is a's most preferred post in L_2 and the rank of b in a's preference list is $\le r_a$. Note that $Y \setminus L_2 = (Y \cap L_3) \subseteq B \setminus F$ (by Lemma 10). Thus the rank of a's most preferred post

in $Y \setminus L_2$ is $\geq r_a$ and hence no post in $Y \setminus L_2$ can be ranked better than b. So the post b is, in fact, a's most preferred post in Y. Thus the edge (a, b) belongs to H as well.

– Any edge (a, b) in G' where $b \in Z$ is such that b is a's most preferred post in L_3. Since $L_3 \supseteq Z$, this means b is a's most preferred post in Z. Thus H also contains the edge (a, b).

Thus every edge (a, b) in G' that is missing in H satisfies $b \in (X \cap L_2) \cup (Y \cap L_3)$. ◻

Theorem 15, along with Lemma 14, finishes the proof of the necessary part of Theorem 5, and this completes the proof of correctness of our algorithm.

3.4. Finding a max-size popular matching in G. If G admits a popular matching, then our algorithm currently finds some popular matching in G, not necessarily a *max-size* popular matching. We could easily modify our algorithm so that it always returns a max-size popular matching in G. This requires the following minor changes in our algorithm:

1. Compute the graph H exactly as done in phases (I)–(III) of our algorithm and find a maximum matching M in the subgraph of H restricted to real posts in $X \cup Y$ and their neighbors.
2. (i) Add the edges of H on posts in Z and augment M.
   (ii) Add the edges of H on dummy posts and augment M further.

Step 1 above is identical to our original algorithm, the minor change in our revised algorithm is in step 2. While our original algorithm did not show any distinction between the posts in Z and the dummy posts while augmenting M, here we give precedence to the posts in Z. Thus M is a maximum matching in H that matches all real posts in $X \cup Y$ (as before) and subject to this, as many posts in Z as possible, or equivalently, as *few* dummy posts as possible.

Suppose G admits a popular matching. Then we know that M is an A-complete matching in H. But this includes edges incident on dummy posts as well. Let M' denote the matching M after deleting all edges incident on dummy posts from M. Theorem 16 below shows that M' is a max-size popular matching in G.

THEOREM 16. *Let M^* be a popular matching in $G = (A \cup B, E)$. Then $|M^*| \leq |M'|$.*

Proof. Let $\tilde{H}$ denote the graph H without the edges incident on dummy posts. This is the graph obtained at the end of step 2(i) in our revised algorithm above. It follows from the construction of M that M' is a max-size matching in $\tilde{H}$.

Corresponding to the popular matching M^*, we construct the graph G' as done in the early part of this section. This is the graph G' before augmenting it with dummy posts and let us denote this graph by $\tilde{G'}$. We know from Lemma 12 that M^* belongs to $\tilde{G'}$, thus the size of a maximum matching in $\tilde{G'}$ is at least k, where $|M^*| = k$.

Consider the graph $\tilde{G}_1$ whose edge set is the intersection of the edge sets of $\tilde{G'}$ and $\tilde{H}$. Analogous to Claim 6, we have the following claim here.

CLAIM 7. *All posts in $(X \cap L_2) \cup (Y \cap L_3)$ are odd/unreachable in $\tilde{G}_1$. Moreover, every edge (a, b) in $\tilde{G'}$ that is missing in $\tilde{H}$ satisfies $b \in (X \cap L_2) \cup (Y \cap L_3)$.*

We will first finish the proof of Theorem 16 and then we will prove Claim 7. Claim 7 tells us that $\tilde{G'}$ has no *new* edge (new when compared to $\tilde{G}_1$) on posts that are *even* in $\tilde{G}_1$. Thus the size of a maximum matching in $\tilde{G'}$ equals the size of a maximum matching in $\tilde{G}_1$. We have already seen that the size of a maximum matching in $\tilde{G'}$ is at least k, thus the size of a maximum matching in $\tilde{G}_1$ is also at least k. Since $\tilde{G}_1$ is a subgraph of $\tilde{H}$, the size of a maximum matching in $\tilde{H}$ is also at least k. Thus $|M'| \geq k$. ∎

Proof of Claim 7. The edge set of the graph $\tilde{G}_1$ is obtained by the intersection of the edge sets of $\tilde{G}'$ and $\tilde{H}$. Equivalently, $\tilde{G}_1$ is obtained by deleting the edges incident on dummy posts from the graph G_1 (defined in the proof of Theorem 15). It was shown in the proof of Theorem 15 that all posts in $(X \cap L_2) \cup (Y \cap L_3)$ are odd/unreachable in G_1.

All the edges in G_1 that are missing in $\tilde{G}_1$ are incident on dummy posts and these posts are in $Y \cap L_2$. Since all posts in $(X \cap L_2) \cup (Y \cap L_3)$ are odd/unreachable in G_1 and the edges that got deleted from G_1 are incident on posts in $Y \cap L_2$, it follows that all posts in $(X \cap L_2) \cup (Y \cap L_3)$ remain odd/unreachable in the resulting graph $\tilde{G}_1$.

Regarding the second part of the claim, every edge of $\tilde{G}'$ that is missing in $\tilde{H}$ is an edge of G' that is missing in H. We know from Claim 6 that every such edge (a, b) satisfies $b \in (X \cap L_2) \cup (Y \cap L_3)$. This finishes the proof of Claim 7. ∎

Size of a max-size popular matching. We showed at the end of section 3.2 that every popular matching in G has size at least $2/3 \cdot |M_{\max}|$, where $M_{\max}$ is a max-size matching in G. We will now show that there are instances where a max-size popular matching need not be much larger than $2/3 \cdot |M_{\max}|$. Consider the following instance $G = (A \cup B, E)$ on $6k$ vertices (see Figure 7)—the vertices can be partitioned into k groups here: The first group consists of $a_0^1, a_1^1, a_2^1, b_0^1, b_1^1, b_2^1$ whose preferences are identical to the vertices $a_0, a_1, a_2, b_0, b_1, b_2$ in the instance in Figure 3 (see Figure 7 also); the tth group, for $2 \leq t \leq k$, consists of six vertices $a_0^t, a_1^t, a_2^t, b_0^t, b_1^t, b_2^t$. The preferences of the vertices in the tth group, for $2 \leq t \leq k$, are as follows:
- b_1^t is the top post of all the 3 applicants a_0^t, a_1^t, a_2^t,
- b_2^t is the second ranked post of both a_1^t and a_2^t,
- b_0^t is the *fourth* ranked post of a_2^t,
- b_0^1 (which is a post in the *first* group) is the third ranked post of a_2^t.

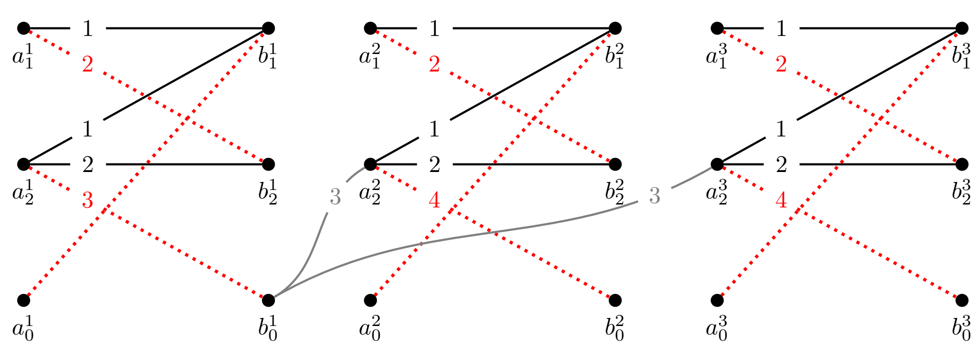

FIG. 7. *The instance G corresponding to $k = 3$, i.e., on $6k = 18$ vertices. Note that b_0^1 is the common third ranked post of all applicants a_2^t for $1 \leq t \leq k$.*

The instance G admits a perfect matching $\{(a_0^t, b_1^t), (a_1^t, b_2^t), (a_2^t, b_0^t) : 1 \leq t \leq k\}$, thus $|M_{\max}| = 3k$. The instance G also admits a popular matching $\{(a_0^1, b_1^1), (a_1^1, b_2^1), (a_2^1, b_0^1)\} \cup \{(a_1^t, b_1^t), (a_2^t, b_2^t) : 2 \leq t \leq k\}$ of size $2k+1$. Note that no popular matching can match the post b_0^t, for any $t \in \{2, \ldots, k\}$. Suppose M is a matching that matches b_0^t, i.e., the edge $(a_2^t, b_0^t) \in M$ for some t, where $2 \leq t \leq k$. It is easy to show that M is not popular.

We can assume that b_0^1 is matched in M, otherwise we get a more popular matching by promoting a_2^t to b_0^1. Similarly, we can assume that b_1^t and b_2^t are matched in M for all $1 \le t \le k$. Consider the following alternating path ρ with respect to M:

$$\rho = b_0^t - a_2^t - b_0^1 - a_2^s - b_2^s - a_1^s - b_1^s - a_0^s,$$

where a_2^s is the vertex matched to b_0^1 in M. This means the post b_2^s is matched to its other neighbor a_1^s and so b_1^s has to be matched to a_0^s. The path ρ has three consecutive nonmatching edges (a_2^t, b_0^1), (a_2^s, b_2^s), and (a_1^s, b_1^s) that are labeled $+1$. Hence $M \oplus \rho \succ M$, thus M is not popular. So the size of any popular matching in G is at most $3 + 2(k-1) = 2k+1$, which is $(2/3 \cdot |M_{\max}|) + 1$.

3.5. Running time of our algorithm. We now analyze the running time of our algorithm. Observe that we can maintain the most preferred posts in X, Y, and Z for all applicants over all iterations in $O(m)$ time, where m is the number of edges in G. To begin with, the most preferred non-f-post for all applicants can be determined in $O(m)$ time. Thereafter, whenever a post b moves from X to Y (similarly, from Y to Z), we charge b a cost of the degree of b to pay for checking if any of its neighbors now has b as its most preferred post in Y (resp., Z).

Let n be the number of vertices in G. The number of iterations is $O(n)$ and the most expensive step in each iteration is finding a maximum matching in a subgraph where each vertex in A has degree at most 2. This step can easily be performed in $O(n)$ time. Thus the running time of our algorithm is $O(n^2)$. Hence we can deduce Theorem 2 stated in section 1.

There are instances on $O(n)$ vertices and $O(n)$ edges where our algorithm takes $\Theta(n^2)$ time. Consider the example given in Figure 8: Here there are $2n+1$ applicants and $2n+2$ posts and the number of edges is $5n+2$. For each $1 \le i \le n$, we have $f(a_i) = f(a_i') = f_i$ and s_i is the most preferred non-f-post for both a_i and a_i'. For a_0, we have $f(a_0) = f_0$ and a_0's most preferred non-f-post is s_0.

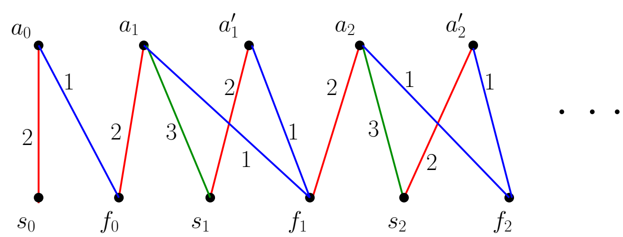

FIG. 8. *The preferences of applicants are indicated on the edges. Our algorithm runs for $n+1$ iterations here.*

In the starting graph H_1, there is exactly one post that is even in Y_1: This is s_0 and so s_0 moves from Y_1 to Z_1. In the second iteration, f_0 has no applicant in $A \setminus \mathsf{nbr}(Z)$ that regards it as a top post and this causes the demotion of f_0 from X_2 to Y_2. Now the post f_0 is the most preferred post in Y_2 for a_1 and this makes s_1 even in Y_2 and causes s_1 to move from Y_2 to Z_2.

This makes both a_1 and a_1' belong to $\mathsf{nbr}(Z)$, and hence f_1 gets isolated in H_3 and so f_1 moves from X_3 to Y_3. Now f_1 becomes the most preferred post in Y_3 for a_2 and this causes s_2 to move from Y_3 to Z_3 and so on. Thus our algorithm runs for $n+1$ iterations. This instance admits popular matchings; for instance, $\{(a_0, f_0), (a_i, f_i), (a_i', s_i) : 1 \le i \le n\}$ is a popular matching here.

4. Our NP-hardness result. Given a matching M in $G = (A \cup B, E)$, it was shown in [3] that M can be tested for popularity in $O(\sqrt{|V|} \cdot |E|)$ time (even in the presence of ties), where $|V| = |A \cup B|$. Thus the 2-sided popular matching problem in G with 1-sided ties is in the complexity class NP. We now show the NP-hardness of the 2-sided popular matching problem in G with 1-sided ties using the (2,2)-E3-SAT problem.

Recall that the (2,2)-E3-SAT problem takes as its input a Boolean formula $\mathcal{I}$ in CNF, where each clause contains three literals and every variable appears exactly twice in unnegated form and exactly twice in negated form in the clauses. The problem is to determine if $\mathcal{I}$ is satisfiable or not and this problem is NP-complete [2].

Constructing a popular matching instance $G = (A \cup B, E)$ from $\mathcal{I}$. Let $\mathcal{I}$ have m clauses and n variables. The instance G constructed consists of n variable gadgets, m clause gadgets, and some interconnecting edges between these; see Figure 9. A *variable gadget* representing variable v_j, for $1 \le j \le n$, is a 4-cycle on vertices $a_{j_1}, b_{j_1}, a_{j_2}$, and b_{j_2}, where $a_{j_1}, a_{j_2} \in A$ and $b_{j_1}, b_{j_2} \in B$. A *clause gadget* representing clause C_i, for $1 \le i \le m$, is a subdivision graph of a claw. Its edges are divided into three classes: $c_i \in B$ is at the center, the neighbors of c_i are $x_{i_1}, x_{i_2}, x_{i_3} \in A$, and finally, each of $x_{i_1}, x_{i_2}, x_{i_3}$ is adjacent to its respective copy in $\mathcal{Y}_i = \{y_{i_1}, y_{i_2}, y_{i_3}\}$, where $\mathcal{Y}_i \subseteq B$.

A vertex in $\mathcal{Y}_i$ represents an appearance of a variable. For instance, y_{3_1} is the first literal of the third clause. Each of the vertices in $\mathcal{Y}_i$ is connected to a vertex in the variable gadget via an *interconnecting edge*. Vertex y_{i_k} is connected to the gadget standing for variable j if the kth literal of the ith clause is either v_j or $\neg v_j$. If it is v_j, then the interconnecting edge ends at a_{j_1}, else at a_{j_2}. The preferences of this instance can be seen in Figure 9. The constructed graph trivially satisfies both conditions claimed in section 1, i.e., every vertex in A has a strict preference list of length 2 or 4 and every vertex in B has either a strict preference list of length 2 or a single tie of length 2 or 3 as a preference list.

Constructing a truth assignment in $\mathcal{I}$, given a popular matching M in G. The graph G is as described above. Claim 8 states that any popular matching M in G has a certain structure.

CLAIM 8. *Any popular matching M in G has to obey the following properties.*
- *M avoids all interconnecting edges.*
- *M is one of the two perfect matchings on any variable gadget; i.e., for each j, the edges of M, restricted to the gadget corresponding to variable v_j, are either* (i) *(a_{j_1}, b_{j_1}) and (a_{j_2}, b_{j_2}), or* (ii) *(a_{j_1}, b_{j_2}) and (a_{j_2}, b_{j_1}).*
- *M leaves exactly one vertex per clause i unmatched and this unmatched vertex y_{i_k} is adjacent to an a_{j_t} that is matched to b_{j_1}.*

Proof. Label each edge (a, b) in $G \setminus M$ by the pair (α, β), where $\alpha \in \{\pm 1\}$ is a's vote for b versus $M(a)$ and $\beta \in \{\pm 1, 0\}$ is b's vote for a versus $M(b)$. Our first observation is that every c_i, for $1 \le i \le m$, and every b_{j_1}, for $1 \le j \le n$, must be matched in M. That is because these vertices are the top choices for each of their neighbors, hence if one of them is left unmatched, then there would be an edge labeled $(+1, +1)$ incident to an unmatched vertex. This contradicts the popularity of M.

Having established that c_i is matched for all $1 \le i \le m$, we can assume without loss of generality that $(c_i, x_{i_3}) \in M$ for a chosen clause gadget i. Also, the edges (x_{i_1}, y_{i_1}) and (x_{i_2}, y_{i_2}) must be in M, because they are the top ranked edges of y_{i_1} and y_{i_2}, respectively. Let us now investigate an arbitrary variable gadget j for some $1 \le j \le n$. Again, without loss of generality we can assume that $(a_{j_1}, b_{j_1}) \in M$. We now claim that $(a_{j_2}, b_{j_2}) \in M$ as well.

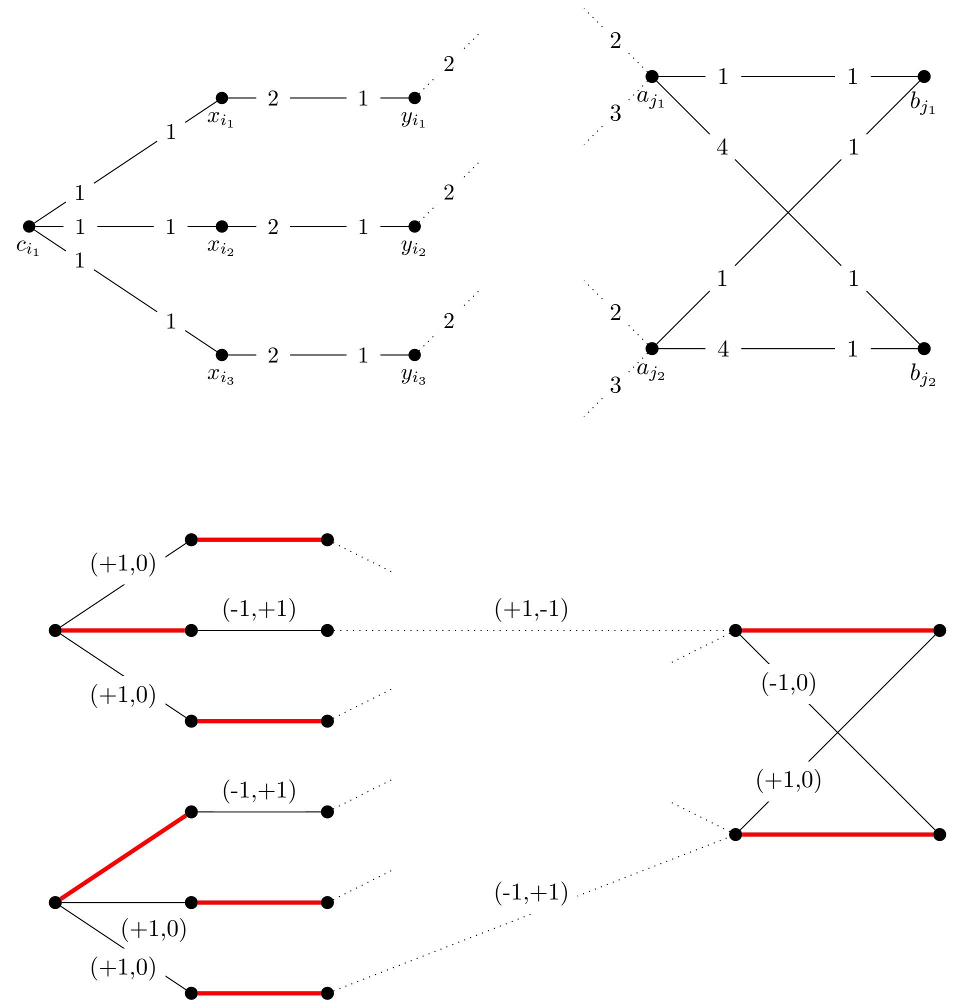

FIG. 9. *A clause gadget, a variable gadget, and the structure of the entire construction with a variable that appears at the second place in the first clause in unnegated form and at the third place in the second clause in negated form. The thick red matching corresponds to a true variable. (Figure appears in color online.)*

Suppose $(a_{j_2}, b_{j_2}) \notin M$. Since M is a maximal matching, $(a_{j_2}, y_{i_k}) \in M$ for some i_k. Based on the above described structure of the clause gadgets, the edges $(x_{i_k}, c_i), (x_{i_{k+1}}, y_{i_{k+1}})$, and $(x_{i_{k+2}}, y_{i_{k+2}})$ are in M, where the subscripts are taken modulo 3. Consider the following augmenting path p with respect to M:

$$\rho = b_{j_2} - a_{j_1} - b_{j_1} - a_{j_2} - y_{i_k} - x_{i_k} - c_i - x_{i_{k+1}} - y_{i_{k+1}}.$$

We have $M \oplus p \succ M$, which contradicts the popularity of M. Thus $(a_{j_2}, b_{j_2}) \in M$.

An analogous argument proves that if $(a_{j_2}, b_{j_1}) \in M$ for some j, then (a_{j_1}, b_{j_2}) has to be in M. The last observation we make is that if y_{i_k} is unmatched in M, then its interconnecting edge leads to an a_{j_t} that is matched to b_{j_1}. Otherwise, (y_{i_k}, a_{j_t}) would be labeled $(+1, +1)$ with one vertex unmatched, a contradiction again to the popularity of M. This finishes the proof of Claim 8. ∎

We assign true to all variables v_j such that $M \supseteq \{(a_{j_1}, b_{j_1}), (a_{j_2}, b_{j_2})\}$ and false to all variables v_j such that $M \supseteq \{(a_{j_1}, b_{j_2}), (a_{j_2}, b_{j_1})\}$.

So the truth value of every variable is uniquely defined and all we need to show is that every clause has a true literal. Assume that in clause C_i, all three literals are false. The clause gadget has an unmatched vertex y_{i_k} that is adjacent to an a_{j_t}. If the literal is false, then a_{j_t} prefers y_{i_k} to $M(a_{j_t}) = b_{j_2}$ and this becomes an edge labeled $(+1, +1)$ with an unmatched end vertex—this contradicts the popularity of M. Hence, in every clause there is at least one true literal, and so this is a satisfying assignment.

Constructing a popular matching in G, given a truth assignment in $\mathcal{I}$. Here we first construct a matching M in the graph G as described below and then show that it is popular. Initially $M = \emptyset$. For each j, where $1 \le j \le n$, if $v_j = \mathsf{true}$ in the assignment, then add (a_{j_1}, b_{j_1}) and (a_{j_2}, b_{j_2}) to M, else add (a_{j_1}, b_{j_2}) and (a_{j_2}, b_{j_1}) to M. For each i, where $1 \le i \le m$, in the gadget corresponding to clause C_i, any true literal is chosen (say, the kth literal) and y_{i_k}, representing its appearance, is left unmatched. Moreover, $(x_{i_k}, c_i), (x_{i_{k+1}}, y_{i_{k+1}})$, and $(x_{i_{k+2}}, y_{i_{k+2}})$ (where the subscripts are taken modulo 3) are added to M. No interconnecting edge appears in M. This finishes the description of M.

LEMMA 17. *The matching M is popular in G.*

Proof. Suppose M is not popular. Then there is another matching M' that is more popular than M. This can only happen if $M \oplus M'$ contains a component ρ such that the number of vertices in ρ that prefer M' to M is more than those that prefer M to M'. To achieve this, the matching M' should contain at least one edge labeled either $(+1, +1)$ or $(+1, 0)$, where we use edge labels (α, β) as described in the proof of Claim 8. We now analyze the cases based on the occurrences of such "positive" edges.

Since we started with a truth assignment, no interconnecting edge can be labeled $(+1, +1)$. In fact, it is straightforward to check that no edge here can be labeled $(+1, +1)$. We now check for the occurrences of edges labeled $(+1, 0)$. These can occur at two places: the edge (a_{j_t}, b_{j_1}) for any $1 \le j \le n$ and the edge (x_{i_k}, c_i) for any $1 \le i \le m$.

Case 1. Suppose (a_{j_2}, b_{j_1}) is labeled $(+1, 0)$. This happens if v_j is true in the truth assignment. We start the augmenting path ρ at (a_{j_2}, b_{j_1}). Augmenting along the 4-cycle is not sufficient to break popularity; therefore, a_{j_1} must be matched along one of its interconnecting edges, say (a_{j_1}, y_{i_k}).

- If y_{i_k} is unmatched, consider the path $\rho = b_{j_2}$-a_{j_2}-b_{j_1}-a_{j_1}-y_{i_k}. There are two vertices (a_{j_1} and b_{j_2}) that prefer M to $M \oplus \rho$ and two vertices (a_{j_2} and y_{i_k}) that prefer $M \oplus \rho$ to M.
- If y_{i_k} is matched, then extend the path ρ until the unmatched vertex of the ith variable gadget (call this y_{i_t}). The path ρ is described below:

$$\rho = b_{j_2} - a_{j_2} - b_{j_1} - a_{j_1} - y_{i_k} - x_{i_k} - c_i - x_{i_t} - y_{i_t}.$$

  So 4 vertices, i.e. b_{j_2}, a_{j_1}, y_{i_k}, and x_{i_t}, prefer M to $M \oplus \rho$ while three vertices, i.e., a_{j_2}, x_{i_k}, and y_{i_t}, prefer $M \oplus \rho$ to M.

Case 2. Now suppose (x_{i_k}, c_i) is labeled $(+1, 0)$. Let us assume that this edge is (x_{i_3}, c_i) and suppose $(x_{i_1}, c_i) \in M$. Consider the alternating path $\rho = y_{i_1}$-x_{i_1}-c_i-x_{i_3}-y_{i_3}. In the matching $M \oplus \rho$, the vertices x_{i_3} and y_{i_1} are better-off while x_{i_1} and y_{i_3} are worse-off, i.e., they prefer M to $M \oplus \rho$. In order to collect one more vertex that prefers $M \oplus \rho$, let us extend this alternating path ρ to include (a_{j_k}, y_{i_3}), the interconnecting edge of y_{i_3}. The vertex y_{i_3} still prefers M to $M \oplus \rho$ since y_{i_3} was paired in M to its top ranked neighbor.

Without loss of generality, let us assume that this interconnecting edge is (a_{j_2}, y_{i_3}). We have two cases here: either $\{(a_{j_1}, b_{j_1}), (a_{j_2}, b_{j_2})\} \subseteq M$ or $\{(a_{j_1}, b_{j_2}), (a_{j_2}, b_{j_1})\} \subseteq M$.

- In the first case, the path ρ gets extended to $\cdots$-a_{j_2}-b_{j_2}. So a_{j_2} prefers $M \oplus \rho$ to M. However, b_{j_2} is left unmatched in $M \oplus \rho$, so b_{j_2} prefers M to $M \oplus \rho$.
- In the second case, the path ρ gets extended to $\cdots$-a_{j_2}-b_{j_1}-a_{j_1}-b_{j_2}. So a_{j_1} prefers $M \oplus \rho$ to M. However, both a_{j_2} and b_{j_2} prefer M to $M \oplus \rho$.

We have analyzed all the cases where edges can be labeled $(+1, 0)$ and we showed that there is no alternating cycle or path ρ containing an edge labeled $(+1, 0)$ such that $M \oplus \rho \succ M$. Thus M is popular. ∎

This finishes the proof of Theorem 3 stated in section 1.

Conclusions and open problems. We gave an $O(n^2)$ algorithm for the popular matching problem in $G = (A \cup B, E)$ where vertices in A have strict preference lists while each vertex in B puts all its neighbors into a single tie and $n = |A \cup B|$. Our algorithm needs the preference lists of vertices in A to be strict and the complexity of the popular matching problem when ties are allowed in the preference lists of vertices in A is currently unknown.

When each $b \in B$ either has a single tie of length at most 3 or a strict preference list (and each $a \in A$ has a strict preference list), we showed that the popular matching problem becomes NP-hard. The complexity of the same problem with ties of length at most 2 instead of 3 is open.

Acknowledgments. We would like to thank the two anonymous reviewers for their helpful comments and suggestions. We would also like to thank the second reviewer for asking us about finding a max-size popular matching.

REFERENCES

[1] D.J. Abraham, R.W. Irving, T. Kavitha, and K. Mehlhorn, *Popular matchings*, SIAM J. Comput., 37 (2007), pp. 1030–1045, https://doi.org/10.1137/06067328X.

[2] P. Berman, M. Karpinski, and A.D. Scott, *Approximation Hardness of Short Symmetric Instances of MAX-3SAT*, Electronic Colloquium on Computational Complexity Report, number 49, 2003.

[3] P. Biró, R.W. Irving, and D.F. Manlove, *Popular matchings in the marriage and roommates problems*, in Proceedings of the 7th International Conference on Algorithms and Complexity (CIAC), Lecture Notes in Comput. Sci. 6078, Springer, Berlin, 2010, pp. 97–108.

[4] Á. Cseh, C.-C. Huang, and T. Kavitha, *Popular matchings with two-sided preferences and one-sided ties*, in Proceedings of the 42nd International Colloquium on Automata, Languages, and Programming (ICALP): Part I, Lecture Notes in Comput. Sci. 9134, Springer, Heidelberg, 2015, pp. 367–379.

[5] Á. Cseh and T. Kavitha, *Popular edges and dominant matchings*, Math. Program., (2017), https://doi.org/10.1007/s10107-017-1183-y.

[6] A.L. Dulmage and N.S. Mendelsohn, *Coverings of bipartite graphs*, Canad. J. Math., 10 (1958), pp. 517–534.

[7] P. Gärdenfors, *Match making: Assignments based on bilateral preferences*, Behavioural Science, 20 (1975), pp. 166–173.

[8] R.L. Graham, M. Grötschel, and L. Lovász, eds., *The Handbook of Combinatorics*, Matchings and Extensions, by W.R. Pulleyblank, North-Holland, Amsterdam, 1995, pp. 179–232.

[9] M. Hirakawa, Y. Yamauchi, S. Kijima, and M. Yamashita, *On the structure of popular matchings in the stable marriage problem - who can join a popular matching?*, in the 3rd International Workshop on Matching under Preferences (MATCH-UP), 2015, p. 124.

[10] C.-C. Huang and T. Kavitha, *Popular matchings in the stable marriage problem*, Inform. and Comput., 222 (2013), pp. 180–194.

[11] C.-C. Huang and T. Kavitha, *Popularity, mixed matchings, and self-duality*, in Proceedings of the 28th ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, Philadelphia, 2017, pp. 2294–2310.

[12] T. Kavitha, *A size-popularity tradeoff in the stable marriage problem*, SIAM J. Comput., 43 (2014), pp. 52–71, https://doi.org/10.1137/120902562.

[13] T. Kavitha, *Popular half-integral matchings*, in Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming (ICALP), LIPIcs. Leibniz Int. Proc. Inform. 55, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016, 22.

[14] T. Kavitha, J. Mestre, and M. Nasre, *Popular mixed matchings*, Theoret. Comput. Sci., 412 (2011), pp. 2679–2690.

[15] T. Kavitha and M. Nasre, *Note: Optimal popular matchings*, Discrete Appl. Math., 157 (2009), pp. 3181–3186.

[16] M. Mahdian, *Random popular matchings*, in Proceedings of the 7th ACM Conference on Electronic Commerce, 2006, pp. 238–242.

[17] C.T.S. Sng and D.F. Manlove, *Popular matchings in the weighted capacitated house allocation problem*, J. Discrete Algorithms, 8 (2010), pp. 102–116.

[18] R.M. McCutchen, *The least-unpopularity-factor and least-unpopularity-margin criteria for matching problems with one-sided preferences*, in Proceedings of the 8th Latin American Theoretical Informatics Symposium (LATIN), Lecture Notes in Comput. Sci. 4957, Springer, Berlin, 2008, pp. 593–604.

[19] E. McDermid and R.W. Irving, *Popular matchings: Structure and algorithms*, J. Comb. Optim., 22 (2011), pp. 339–358.

[20] J. Mestre, *Weighted popular matchings*, ACM Trans. Algorithms, 10 (2014), 2.

[21] K. Paluch, *Popular and clan-popular b-matchings*, Theoret. Comput. Sci., 544 (2014), pp. 3–13.