Algorithmica manuscript No.
(will be inserted by the editor)

The Hospitals/Residents Problem with Lower Quotas

Koki Hamada · Kazuo Iwama · Shuichi Miyazaki

Received: date / Accepted: date

Abstract The Hospitals/Residents problem is a many-to-one extension of the stable marriage problem. In an instance, each hospital specifies a quota, i.e., an upper bound on the number of positions it provides. It is well-known that in any instance, there exists at least one stable matching, and finding one can be done in polynomial time. In this paper, we consider an extension in which each hospital specifies not only an upper bound but also a *lower* bound on its number of positions. In this setting, there can be instances that admit no stable matching, but the problem of asking if there is a stable matching is solvable in polynomial time. In case there is no stable matching, we consider the problem of finding a matching that is "as stable as possible", namely, a matching with a minimum number of blocking pairs. We show that this problem is hard to approximate within the ratio of $(|H|+|R|)^{1-\epsilon}$ for any positive constant ϵ where H and R are the sets of hospitals and residents, respectively. We then tackle this hardness from two different angles. First, we give an exponential-time

A preliminary version of this paper was presented at the 19th Annual European Symposium on Algorithms, ESA 2011 [20].

K. Hamada
NTT Secure Platform Laboratories, NTT Corporation, 3-9-11, Midori-cho, Musashino-shi, Tokyo 180-8585 Japan.
E-mail: hamada.koki@lab.ntt.co.jp

K. Iwama
Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606–8501, Japan.
E-mail: iwama@kuis.kyoto-u.ac.jp
Supported by JSPS KAKENHI Grant Number 25240002.

S. Miyazaki
Academic Center for Computing and Media Studies, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606–8501, Japan.
E-mail: shuichi@media.kyoto-u.ac.jp
Supported by JSPS KAKENHI Grant Number 24500013.
Corresponding author.

exact algorithm whose running time is $O((|H||R|)^{t+1})$, where t is the number of blocking pairs in an optimal solution. Second, we consider another measure for optimization criteria, i.e., the number of residents who are involved in blocking pairs. We show that this problem is still NP-hard but has a polynomial-time $\sqrt{|R|}$-approximation algorithm.

Keywords The stable marriage problem · The Hospitals/Residents problem · Stable matching · Approximation algorithm

1 Introduction

The *stable marriage problem* is a widely known problem first studied by Gale and Shapley [13]. We are given sets of men and women, and each person's preference list that strictly orders the members of the other sex according to his/her preference. The question is to find a *stable matching*, that is, a matching containing no pair of man and woman who prefer each other to their partners. Such a pair is called a *blocking pair*. Gale and Shapley proved that any instance admits at least one stable matching, and gave an algorithm to find one, known as the *Gale-Shapley algorithm*.

In the same paper [13], they also proposed a many-to-one extension of the stable marriage problem, which is currently known as the *Hospitals/Residents problem* (*HR* for short). In HR, the two sets corresponding to men and women are residents and hospitals. Each hospital specifies its *quota*, which means that it can accept at most this number of residents. Hence in a feasible matching, the number of residents assigned to each hospital is no more than its quota. Most properties of the stable marriage problem carry over to HR, e.g., any instance admits a stable matching, and we can find one by the appropriately modified Gale-Shapley algorithm. As the name of HR suggests, it has real-world applications in assigning residents to hospitals in many countries. Centralized matching schemes for accomplishing this task incorporate algorithms for solving underlying HR instances and include NRMP in the U.S. [16], CaRMS in Canada [8], and SFAS in Scotland [22]. HR also arises in the assignment of students to schools in Singapore [29]. Along with these applications and due to special requirements in reality, several useful extensions have been proposed, such as HR with couples [28, 27, 3, 26], and the Student-Project Allocation problem [2].

In this paper, we study another extension of HR where each hospital declares not only an upper bound but also a *lower* bound on the number of residents it accepts. Consequently, a feasible matching must satisfy the condition that the number of residents assigned to each hospital is between its upper and lower quotas. This restriction seems quite relevant in several situations. For example, the shortage of doctors in hospitals in rural area is a critical issue; it is sometimes natural to guarantee some number of residents for such hospitals in the residents-hospitals matching. Also, when determining supervisors of students in universities, it is quite common to consider that the number of students assigned to each professor should be somehow balanced, which can be achieved again by specifying both upper and lower bounds on the number of students accepted by each professor. We call this problem *HR with Lower Quota* (*HRLQ* for short).

The notion of lower quota was first raised in [18] and followed by [6,21] (see "Related Work" below). In this paper, we are interested in a most natural question, i.e., how to obtain "good" matchings in this new setting. In HRLQ, stable matchings do not always exist. However, it is easy to decide whether or not there is a stable matching for a given instance, since in HR the number of students a specific hospital h receives is identical for any stable matching (this is a part of the well-known *Rural Hospitals Theorem* [14]). Namely, if this number satisfies the upper and lower bound conditions of all the hospitals, it is a feasible (and stable) matching, and otherwise, no stable matching exists. In case there is no stable matching, it is natural to seek for a matching that is "as stable as possible".

Our Contributions. We first consider the problem of minimizing the number of blocking pairs, which is quite popular in the literature (e.g., [25,1,7]). As shown in Sec. 2, it seems that the introduction of the lower quota intrinsically increases the difficulty of the problem. Actually, we show that this problem is NP-hard and cannot be approximated within a factor of $(|H|+|R|)^{1-\varepsilon}$ for any positive constant ε unless P=NP, where H and R denote the sets of hospitals and residents, respectively. This inapproximability result holds even if all the preference lists are complete (i.e., include all the members of the other side), all the hospitals have the same preference list, (e.g., determined by scores of exams and known as the *master list* [23]), and all the hospitals have an upper quota of one. On the positive side, we give a polynomial-time $(|H| + |R|)$–approximation algorithm, which shows that the above inapproximability result is almost tight.

We then tackle this hardness from two different angles. First, we give an exponential-time exact algorithm with running time $O((|H||R|)^{t+1})$, where t is the number of blocking pairs in an optimal solution. Note that this is a polynomial-time algorithm when t is a constant. Second, we consider another measure for optimization criteria, i.e., the number of residents who are involved in blocking pairs. We show that this problem is still NP-hard, but give a quadratic improvement, i.e., we give a polynomial-time $\sqrt{|R|}$-approximation algorithm. We also give an instance showing that our analysis is tight up to a constant factor. Furthermore, we show that if our problem has a constant approximation factor, then the Densest k-Subgraph Problem (DkS) has a constant approximation factor also. Note that the best known approximation factor of DkS has long been $|V|^{1/3}$ [10] in spite of extensive studies, and was recently improved to $|V|^{1/4+\epsilon}$ for an arbitrary positive constant ϵ [5]. The reduction is somewhat tricky; it is done through a third problem, called the Minimum Coverage Problem (MinC), and exploits the best approximation algorithm for DkS. MinC is relatively less studied and only NP-hardness was previously known for its complexity [30]. As a by-product, our proof gives a similar inapproximability result for MinC (Lemma 12), which is of independent interest.

Related Work. Biró, et al. [6] also considered HR with lower quotas. In contrast to our model, which requires the lower quotas of all the hospitals to be satisfied, their model allows some hospitals to be closed, i.e., to receive no residents. They proved that the problem of deciding whether there is a feasible solution is NP-complete. Huang [21] considered *classified stable matchings*, in which each hospital defines a family of subsets of residents and declares upper and lower quotas for each of the subsets. He proved a dichotomy theorem for the problem of deciding the existence of a stable matching; namely, if the subset families satisfy some structural property, then the problem is in P, otherwise, it is NP-complete. Recently, Fleiner and Kamiyama [11] generalized Huang's result to many-to-many case, where not only hospitals' side but also the residents' side can declare upper and lower quotas.

2 Preliminaries

An instance of the *Hospitals/Residents Problem with Lower Quota* (*HRLQ* for short) consists of a set R of residents and a set H of hospitals. Each hospital h has lower and upper quotas, p and q ($p \leq q$), respectively. We sometimes say that the quota of h is $[p, q]$, or h is a $[p, q]$-hospital. For simplicity, we also write the name of a hospital with its quotas, such as $h[p, q]$. Each member (resident or hospital) has a preference list that orders a subset of the members of the other party.

A *matching* is an assignment of residents to hospitals (possibly leaving some residents unassigned), where matched residents and hospitals are in the preference list of each other. Let $M(r)$ be the hospital to which resident r is assigned under a matching M (if it exists), and $M(h)$ be the set of residents assigned to hospital h. A *feasible matching* is a matching such that $p \leq |M(h)| \leq q$ for each hospital $h[p, q]$. We may sometimes call a feasible matching simply a matching when there is no fear of confusion. For a matching M and a hospital $h[p, q]$, we say that h is *full* if $|M(h)| = q$, *under-subscribed* if $|M(h)| < q$, and *empty* if $|M(h)| = 0$.

For a matching M, we say that a pair comprising a resident r and a hospital h who include each other in their lists forms a *blocking pair* for M if the following two conditions are met: (i) r is either unassigned or prefers h to $M(r)$, and (ii) h is under-subscribed or prefers r to one of the residents in $M(h)$. We say that r is a *blocking resident* for M if r is involved in a blocking pair for M.

Minimum-Blocking-Pair Hospitals/Residents Problem with Lower Quota (*Min-BP HRLQ* for short) is the problem of finding a feasible matching with the minimum number of blocking pairs. *Min-BP 1ML-HRLQ* ("1ML" standing for "1 Master List") is the restriction of Min-BP HRLQ so that in a given instance, preference lists of all the hospitals are identical. *0-1 Min-BP HRLQ* is the restriction of Min-BP HRLQ where a quota of each hospital is either $[0, 1]$ or $[1, 1]$. *0-1 Min-BP 1ML-HRLQ* is Min-BP HRLQ with both "1ML" and "0-1" restrictions.

Minimum-Blocking-Resident Hospitals/Residents Problem with Lower Quota (*Min-BR HRLQ* for short) is the problem of finding a feasible matching with the minimum number of blocking residents. *Min-BR 1ML-HRLQ*, *0-1 Min-BR HRLQ*, and *0-1 Min-BR 1ML-HRLQ* are defined similarly.

We assume without loss of generality that the number of residents is at least the sum of the lower quotas of all the hospitals, since otherwise there is no feasible matching. We call this assumption the *Number of Residents assumption* (or the *NR-assumption* for short). Also, in this paper we impose the following restriction, the *Complete List restriction* (or the *CL-restriction* for short), to guarantee existence of a feasible solution: every hospital with a positive lower quota must have a complete preference list, and every resident's list must include all such hospitals. (We remark in Sec. 5 that allowing arbitrarily incomplete preference lists makes the problem extremely hard.)

We say that an algorithm A is an $r(n)$-approximation algorithm for a minimization (maximization, respectively) problem if it satisfies $A(x)/opt(x) \leq r(n)$ ($opt(x)/A(x) \leq r(n)$, respectively) for any instance x of size n, where $opt(x)$ and $A(x)$ are the costs (e.g., the number of blocking pairs in the case of Min-BP HRLQ) of the optimal and the algorithm's solutions, respectively.

As a starting example, consider n residents and $n+1$ hospitals, whose preference lists and quotas are as follows. Here, "$\cdots$" in the residents' preference lists denotes an arbitrary order of the remaining hospitals.

$$
\begin{array}{rllll}
r_1 : & h_1 & h_{n+1} & \cdots & \\
r_2 : & h_1 & h_2 & h_n & \cdots \\
r_3 : & h_2 & h_1 & h_3 & \cdots \\
r_4 : & h_3 & h_1 & h_4 & \cdots \\
\vdots & & & & \\
r_i : & h_{i-1} & h_1 & h_i & \cdots \\
\vdots & & & & \\
r_n : & h_{n-1} & h_1 & h_n & \cdots
\end{array}
\qquad
\begin{array}{rl}
h_1[0,1] : & r_1 \ r_2 \ \cdots \ r_n \\
h_2[1,1] : & r_1 \ r_2 \ \cdots \ r_n \\
\vdots & \\
h_n[1,1] : & r_1 \ r_2 \ \cdots \ r_n \\
h_{n+1}[1,1] : & r_1 \ r_2 \ \cdots \ r_n
\end{array}
$$

Note that we have n $[1, 1]$-hospitals all of which have to be filled by the n residents. Therefore, let us modify the instance by removing the $[0, 1]$-hospital h_1 and apply the Gale-Shapley algorithm (see e.g., [16] for the Gale-Shapley algorithm; in this paper it is always the residents-oriented version, namely, residents make and hospitals receive proposals). Then the resulting matching is $M_1 = \{(r_1, h_{n+1}), (r_2, h_2), (r_3, h_3), \cdots, (r_n, h_n)\}$, which contains at least n blocking pairs (between h_1 and every resident). However, the matching $M_2 = \{(r_1, h_{n+1}), (r_2, h_n), (r_3, h_2), (r_4, h_3), \ldots, (r_n, h_{n-1})\}$ contains only three blocking pairs, namely (r_1, h_1), (r_2, h_1), and (r_2, h_2).

3 Minimum-Blocking-Pair HRLQ

In this section, we consider the problem of minimizing the number of blocking pairs.

3.1 Inapproximability

We first prove a strong inapproximability result for the restricted subclass, as mentioned in Sec. 1.

Theorem 1 *For any positive constant ε, there is no polynomial-time $(|H|+|R|)^{1-\varepsilon}$-approximation algorithm for 0-1 Min-BP 1ML-HRLQ unless P=NP, even if all the preference lists are complete.*

Proof We demonstrate a polynomial-time reduction from the well-known NP-complete problem *Vertex Cover* (*VC* for short) [15]. In VC, we are given a graph $G=(V,E)$ and a positive integer $K \leq |V|$, and asked if there is a subset C of vertices of G such that $|C| \leq K$, which contains at least one endpoint of each edge. Let $I_0=(G_0,K_0)$ be an instance of VC where $G_0=(V_0,E_0)$ and K_0 is a positive integer. Define $n=|V_0|$. For a constant ε, define $c=\lceil \frac{8}{\varepsilon} \rceil$, $B_1=n^c$, and $B_2=n^c-|E_0|$.

We construct the instance I of 0-1 Min-BP 1ML-HRLQ from I_0. The set of residents is $R=C \cup F \cup S$, and the set of hospitals is $H=V \cup T \cup X$. Each set is defined as follows:

$$
\begin{aligned}
C &= \{c_i \mid 1 \leq i \leq K_0\} \\
F &= \{f_i \mid 1 \leq i \leq n-K_0\} \\
S^{i,j} &= \{s^{i,j}_{0,a} \mid 1 \leq a \leq B_2\} \cup \{s^{i,j}_{1,a} \mid 1 \leq a \leq B_2\} \quad ((v_i,v_j) \in E_0, i<j) \\
S &= \bigcup S^{i,j} \\
V &= \{v_i \mid 1 \leq i \leq n\} \\
T^{i,j} &= \{t^{i,j}_{0,a} \mid 1 \leq a \leq B_2\} \cup \{t^{i,j}_{1,a} \mid 1 \leq a \leq B_2\} \quad ((v_i,v_j) \in E_0, i<j) \\
T &= \bigcup T^{i,j} \\
X &= \{x_i \mid 1 \leq i \leq B_1\}
\end{aligned}
$$

Each hospital in X has a quota [0,1], and other hospitals have a quota [1,1]. Note that $|C|+|F|=|V|(=n)$ and $|S|=|T|(=2|E_0|B_2)$. Since any hospital in $V \cup T$ has a quota [1,1], any feasible matching is a one-to-one correspondence between R and $V \cup T$, and every hospital in X must be empty. Note that $|H|=n+2|E_0|B_2+B_1$ and $|R|=n+2|E_0|B_2$; hence $|H|+|R|=2n+4|E_0|B_2+B_1=2n-4|E_0|^2+(4|E_0|+1)n^c<n^2+4n^{c+2}+n^c \leq 6n^{c+2}$, which is polynomial in n.

Next, we construct preference lists. Fig. 1 shows preference lists of residents, where $[[V]]$ (respectively $[[X]]$) denotes a total order of elements in V (respectively X) in an increasing order of indices. The symbol "$\cdots$" denotes an arbitrarily ordered list of all the other hospitals that do not explicitly appear in the list.

$$
\begin{array}{llllll}
c_i & : [[V]] & [[X]] \cdots & & & (1 \le i \le K_0) \\
f_i & : [[V]] & [[X]] \cdots & & & (1 \le i \le n - K_0) \\
s^{i,j}_{0,1} & : t^{i,j}_{0,1} & v_i & t^{i,j}_{1,1} & [[X]] \cdots & ((v_i, v_j) \in E_0, i < j) \\
s^{i,j}_{0,2} & : t^{i,j}_{0,2} & v_i & t^{i,j}_{0,3} & [[X]] \cdots & ((v_i, v_j) \in E_0, i < j) \\
 & \vdots & & & & \\
s^{i,j}_{0,B_2-1} & : t^{i,j}_{0,B_2-1} & v_i & t^{i,j}_{0,B_2} & [[X]] \cdots & ((v_i, v_j) \in E_0, i < j) \\
s^{i,j}_{0,B_2} & : t^{i,j}_{0,B_2} & v_i & t^{i,j}_{0,1} & [[X]] \cdots & ((v_i, v_j) \in E_0, i < j) \\
s^{i,j}_{1,1} & : t^{i,j}_{0,2} & v_j & t^{i,j}_{1,2} & [[X]] \cdots & ((v_i, v_j) \in E_0, i < j) \\
s^{i,j}_{1,2} & : t^{i,j}_{1,2} & v_j & t^{i,j}_{1,3} & [[X]] \cdots & ((v_i, v_j) \in E_0, i < j) \\
 & \vdots & & & & \\
s^{i,j}_{1,B_2-1} & : t^{i,j}_{1,B_2-1} & v_j & t^{i,j}_{1,B_2} & [[X]] \cdots & ((v_i, v_j) \in E_0, i < j) \\
s^{i,j}_{1,B_2} & : t^{i,j}_{1,B_2} & v_j & t^{i,j}_{1,1} & [[X]] \cdots & ((v_i, v_j) \in E_0, i < j)
\end{array}
$$

Fig. 1 Preference lists of residents

Preference lists of hospitals are identical and are obtained from the master list "$[[C]]$ $[[S]]$ $[[F]]$". Here, $[[C]]$ and $[[F]]$ are as before a total order of all the residents in C and F, respectively, in an increasing order of indices. $[[S]]$ is a total order of $[[S^{i,j}]]$ $((v_i, v_j) \in E_0, i < j)$ in any order, where $[[S^{i,j}]] = s^{i,j}_{1,1}\ s^{i,j}_{0,1}\ s^{i,j}_{0,2}\ \cdots\ s^{i,j}_{0,B_2}\ s^{i,j}_{1,2}\ \cdots\ s^{i,j}_{1,B_2}$.

Now the reduction is completed. Before showing the correctness proof, we will see some properties of the reduced instance. For a resident r and a hospital h, if h appears to the right of the $[[X]]$-part of r's list, we call (r, h) a *prohibited pair.*

Lemma 1 *If a matching M contains a prohibited pair, then the number of blocking pairs in M is at least B_1.*

Proof Suppose that a matching M contains a prohibited pair (r, h). By the definition of prohibited pairs, r prefers any hospital $x \in X$ to h. On the other hand, recall that any hospital $x \in X$ is empty in any feasible matching, and hence, under-subscribed. Hence, (r, x) is a blocking pair for every $x \in X$. Since $|X| = B_1$, the proof is completed. $\square$

Now, recall that for each edge $(v_i, v_j) \in E_0$ $(i < j)$, there are the set of residents $S^{i,j}$ and the set of hospitals $T^{i,j}$. We call this pair of sets a $g_{i,j}$-*gadget*, and write it as $g_{i,j} = (S^{i,j}, T^{i,j})$. For each gadget $g_{i,j}$, let us define two perfect matchings between $S^{i,j}$ and $T^{i,j}$ as follows:

$$M^0_{i,j} = \{(s^{i,j}_{0,1}, t^{i,j}_{0,1}), (s^{i,j}_{0,2}, t^{i,j}_{0,2}), \ldots, (s^{i,j}_{0,a}, t^{i,j}_{0,a}), \ldots, (s^{i,j}_{0,B_2-1}, t^{i,j}_{0,B_2-1}), (s^{i,j}_{0,B_2}, t^{i,j}_{0,B_2}), (s^{i,j}_{1,1}, t^{i,j}_{1,2}), (s^{i,j}_{1,2}, t^{i,j}_{1,3}), \ldots, (s^{i,j}_{1,a}, t^{i,j}_{1,a+1}), \ldots, (s^{i,j}_{1,B_2-1}, t^{i,j}_{1,B_2}), (s^{i,j}_{1,B_2}, t^{i,j}_{1,1})\}, \text{ and}$$

$$M^1_{i,j} = \{(s^{i,j}_{0,1}, t^{i,j}_{1,1}), (s^{i,j}_{0,2}, t^{i,j}_{0,3}), \ldots, (s^{i,j}_{0,a}, t^{i,j}_{0,a+1}), \ldots, (s^{i,j}_{0,B_2-1}, t^{i,j}_{0,B_2}), (s^{i,j}_{0,B_2}, t^{i,j}_{0,1}), (s^{i,j}_{1,1}, t^{i,j}_{0,2}), (s^{i,j}_{1,2}, t^{i,j}_{1,2}), \ldots, (s^{i,j}_{1,a}, t^{i,j}_{1,a}), \ldots, (s^{i,j}_{1,B_2-1}, t^{i,j}_{1,B_2-1}), (s^{i,j}_{1,B_2}, t^{i,j}_{1,B_2})\}.$$

Fig. 2 shows $M^0_{i,j}$ and $M^1_{i,j}$ on preference lists of $S^{i,j}$, where the $[[X]]$-part and thereafter are omitted.

$s^{i,j}_{0,1}$	: $t^{i,j}_{0,1}$	v_i	$t^{i,j}_{1,1}$		$s^{i,j}_{0,1}$	: $t^{i,j}_{0,1}$	v_i	$t^{i,j}_{1,1}$
$s^{i,j}_{0,2}$	: $t^{i,j}_{0,2}$	v_i	$t^{i,j}_{0,3}$		$s^{i,j}_{0,2}$	: $t^{i,j}_{0,2}$	v_i	$t^{i,j}_{0,3}$
$\vdots$					$\vdots$			
$s^{i,j}_{0,B_2-1}$	: $t^{i,j}_{0,B_2-1}$	v_i	$t^{i,j}_{0,B_2}$		$s^{i,j}_{0,B_2-1}$	: $t^{i,j}_{0,B_2-1}$	v_i	$t^{i,j}_{0,B_2}$
$s^{i,j}_{0,B_2}$	: $t^{i,j}_{0,B_2}$	v_i	$t^{i,j}_{0,1}$		$s^{i,j}_{0,B_2}$	: $t^{i,j}_{0,B_2}$	v_i	$t^{i,j}_{0,1}$
$s^{i,j}_{1,1}$	: $t^{i,j}_{0,2}$	v_j	$t^{i,j}_{1,2}$		$s^{i,j}_{1,1}$	: $t^{i,j}_{0,2}$	v_j	$t^{i,j}_{1,2}$
$s^{i,j}_{1,2}$	: $t^{i,j}_{1,2}$	v_j	$t^{i,j}_{1,3}$		$s^{i,j}_{1,2}$	: $t^{i,j}_{1,2}$	v_j	$t^{i,j}_{1,3}$
$\vdots$					$\vdots$			
$s^{i,j}_{1,B_2-1}$	: $t^{i,j}_{1,B_2-1}$	v_j	$t^{i,j}_{1,B_2}$		$s^{i,j}_{1,B_2-1}$	: $t^{i,j}_{1,B_2-1}$	v_j	$t^{i,j}_{1,B_2}$
$s^{i,j}_{1,B_2}$	: $t^{i,j}_{1,B_2}$	v_j	$t^{i,j}_{1,1}$		$s^{i,j}_{1,B_2}$	: $t^{i,j}_{1,B_2}$	v_j	$t^{i,j}_{1,1}$

Fig. 2 Matchings $M^0_{i,j}$ (left) and $M^1_{i,j}$ (right)

Lemma 2 *For a gadget $g_{i,j} = (S^{i,j}, T^{i,j})$, $M^0_{i,j}$ and $M^1_{i,j}$ are the only perfect matchings between $S^{i,j}$ and $T^{i,j}$ that do not include a prohibited pair. Furthermore, each of $M^0_{i,j}$ and $M^1_{i,j}$ contains only one blocking pair (r, h) such that $r \in S^{i,j}$ and $h \in T^{i,j}$. (Hereafter, we simply state this as a "blocking pair between $S^{i,j}$ and $T^{i,j}$".)*

Proof Construct a bipartite graph $G_{i,j}$, where each vertex set is $S^{i,j}$ and $T^{i,j}$, and there is an edge between $r (\in S^{i,j})$ and $h (\in T^{i,j})$ if and only if (r, h) is not a prohibited pair. One can see that $G_{i,j}$ is a cycle of length $4B_2$. Hence there are only two perfect matchings between $S^{i,j}$ and $T^{i,j}$, and they are actually $M^0_{i,j}$ and $M^1_{i,j}$. Also, it is easy to check that $M^0_{i,j}$ contains only one blocking pair $(s^{i,j}_{1,1}, t^{i,j}_{0,2})$, and $M^1_{i,j}$ contains only one blocking pair $(s^{i,j}_{0,1}, t^{i,j}_{0,1})$. □

We are now ready to show the gap for inapproximability.

Lemma 3 *If I_0 is a "yes" instance of VC, then I has a solution with at most $n^2 + |E_0|$ blocking pairs.*

Proof Suppose that G_0 has a vertex cover of size at most K_0. If its size is less than K_0, add arbitrary vertices to make the size exactly K_0, which is, of course, still a vertex cover. Let this vertex cover be $V_{0c} (\subseteq V_0)$, and let $V_{0f} = V_0 \setminus V_{0c}$. For convenience, we use V_{0c} and V_{0f} also to denote the sets of corresponding hospitals.

We construct a matching M of I according to V_{0c}. First, match each resident in C with each hospital in V_{0c}, and each resident in F with each hospital in V_{0f}, in an arbitrary way. Since $|C \cup F| = |V| = n$, there are at most n^2 blocking pairs between $C \cup F$ and V.

For each gadget $g_{i,j} = (S^{i,j}, T^{i,j}) \quad ((v_i, v_j) \in E_0, i < j)$, we use one of the two matchings in Lemma 2. Since V_{0c} is a vertex cover, either v_i or v_j is included in V_{0c}. If v_i is in V_{0c}, use $M^1_{i,j}$, otherwise, use $M^0_{i,j}$. It is then easy to see that there is no blocking pair between $S^{i,j}$ and $H \setminus T^{i,j}$ or $R \setminus S^{i,j}$ and $T^{i,j}$. Also, as proved in Lemma 2, there is only one blocking pair between $S^{i,j}$ and $T^{i,j}$ in either case.

Therefore, the number of blocking pairs is at most n^2 between $C \cup F$ and V, and exactly $|E_0|$ within $g_{i,j}$-gadgets, and hence $n^2 + |E_0|$ in total, which completes the proof. ☐

Lemma 4 *If I_0 is a "no" instance of VC, then any solution of I has at least B_1 blocking pairs.*

Proof Suppose that I admits a matching M with less than B_1 blocking pairs. We show that I_0 has a vertex cover of size K_0.

First, recall that any feasible matching must be a one-to-one correspondence between R and $V \cup T$. Also, by Lemma 1, if M contains a prohibited pair then there are at least B_1 blocking pairs, contradicting the assumption. Thus, M does not contain a prohibited pair. Since $|C \cup F| = |V|$ and any resident $r \in C \cup F$ includes only V to the left of the $[[X]]$-part in the preference list, M must include a perfect matching between $C \cup F$ and V.

Next, consider a gadget $g_{i,j} = (S^{i,j}, T^{i,j})$ and observe the preference lists of $S^{i,j}$. Since v_i and v_j are matched with residents in $C \cup F$, for M to contain no prohibited pairs, all residents in $S^{i,j}$ must be matched with hospitals in $T^{i,j}$. By Lemma 2, there are only two possibilities, namely, $M^0_{i,j}$ and $M^1_{i,j}$, and either matching admits one blocking pair within each $g_{i,j}$. Hence there are $|E_0|$ such blocking pairs for all $g_{i,j}$-gadgets.

Suppose that the matching between $S^{i,j}$ and $T^{i,j}$ is $M^0_{i,j}$. Then, if the hospital v_j is matched with a resident in F, there are B_2 blocking pairs between v_j and $s^{i,j}_{1,1}, \ldots, s^{i,j}_{1,B_2}$. Then, we have $|E_0| + B_2 = B_1$ blocking pairs, contradicting the assumption. Hence, v_j must be matched with a resident in C. On the other hand, suppose that the matching for $g_{i,j}$ is $M^1_{i,j}$. If the hospital v_i is matched with a resident in F, again there are B_2 blocking pairs, between

v_i and $s^{i,j}_{0,1}, \ldots, s^{i,j}_{0,B_2}$. Therefore, v_i must be matched with a resident in C. Namely, for each edge (v_i, v_j), either v_i or v_j is matched with a resident in C. Hence, the collection of vertices whose corresponding hospitals are matched with residents in C is a vertex cover of size K_0. This completes the proof. □

Finally, we estimate the gap obtained by Lemmas 3 and 4. As observed previously, $n^c < |H| + |R| \le 6n^{c+2}$. Hence, $B_1/(n^2 + |E_0|) \ge n^c/2n^2 = 8n^{c+2}2^{-4}n^{-4} \ge 8n^{c+2}n^{-8} > (|H| + |R|)^{1-\frac{8}{c}} \ge (|H| + |R|)^{1-\varepsilon}$. Hence a polynomial-time $(|H|+|R|)^{1-\varepsilon}$-approximation algorithm for 0-1 Min-BP 1ML-HRLQ solves VC, implying P=NP. □

3.2 Approximability

The following theorem shows that an almost tight upper bound can be achieved by a simple approximation algorithm for the general class.

Theorem 2 *There is a polynomial-time $(|H| + |R|)$–approximation algorithm for Min-BP HRLQ.*

Proof Before showing an algorithm, we introduce some terms used to describe the algorithm. In a matching M, define a *deficiency of a hospital* $h_i[p_i, q_i]$ to be $\max\{p_i - |M(h_i)|, 0\}$. We say that a hospital $h_i[p_i, q_i]$ has *surplus* if h_i satisfies $|M(h_i)| - p_i > 0$. The following simple algorithm (Algorithm I) achieves the approximation ratio of $|H| + |R|$.

Algorithm I

1: Consider an instance I of Min-BP HRLQ as an instance of HR by ignoring lower quotas. Then apply the Gale-Shapley algorithm to I and obtain a matching M.
2: If there is an unassigned resident in M, output M.
3: Move residents from hospitals with surplus to the hospitals with positive deficiencies in an arbitrary way (but so as not to create new positive deficiency) to fill all the deficiencies. Then output the modified matching.

Obviously, Algorithm I runs in polynomial time. Note that because of the NR-assumption and the CL-restriction, Step 3 is executable, namely, there are sufficiently many residents in hospitals with surplus to fill all the deficiencies.

We first show that if a matching M is returned at Step 2, M is an optimal solution. Let r be a resident unassigned in M. Then r must have been rejected by all the hospitals with a positive lower quota, since r includes all such hospitals in the list because of the CL-restriction. Therefore, any such hospital is full in M, that is, M is a feasible matching. Hence, we obtain a feasible stable matching, which is clearly an optimal solution.

In the following, we assume that all the residents are assigned in M. Let k be the sum of the deficiencies over all the hospitals. Then, k residents are moved. Suppose that resident r is moved from hospital h to another hospital. Then, it is easy to see that a new blocking pair includes either r or h since only they can become worse off. Hence, there arise at most $|H|+|R|$ new blocking pairs per resident movement and there are at most $k(|H|+|R|)$ blocking pairs in total. On the other hand, we show in the following that if there are k deficiencies in M, an optimal solution contains at least k blocking pairs. These observations give an $(|H|+|R|)$-approximation upper bound.

Let M_{opt} be an optimal solution. For convenience, we think that a hospital $h_i[p_i, q_i]$ has q_i distinct positions, each of which can receive at most one resident. Define the bipartite graph $G_{M,M_{opt}} = (V_R, V_H, E)$ as follows: Each vertex in V_R corresponds to a resident in I, and each vertex in V_H to a position (so, $|V_H| = \sum q_i$). If resident r is assigned by M to hospital h, then in $G_{M,M_{opt}}$, we include an edge (called an M-edge) between $r \in V_R$ and some position $p \in V_H$ of h, and similarly, if resident r is assigned by M_{opt} to hospital h, then we include an edge (called an M_{opt}-edge) between r and some position p of h, so that a single vertex p receives at most one M-edge and at most one M_{opt}-edge. Without loss of generality, we may assume that if a resident r is assigned to the same hospital by M and M_{opt}, r is assigned to the same position p. (In this case, we have parallel edges between r and p.) Hence, if a resident is assigned to different positions by M and M_{opt}, then he/she is assigned to different hospitals. Note that each vertex of $G_{M,M_{opt}}$ has degree at most two.

Note that M_{opt} satisfies all the lower quotas, while M has k deficiencies. This means that there are at least k vertices in V_H that are matched in M_{opt} but not in M. It is easy to see that these k vertices are endpoints of k disjoint paths in $G_{M,M_{opt}}$, in which M_{opt}-edges and M-edges appear alternately. By a standard argument (for example, see the proof of Lemma 4.2 of [17]), we can show that each such path contains at least one blocking pair for M or M_{opt}, but all of them are for M_{opt} because M is stable. This completes the proof. □

3.3 Exponential-Time Exact Algorithm

Our goal in this section is to design non-trivial exponential-time algorithms by using the parameter t denoting the optimal cost, i.e., the number of blocking pairs in an optimal solution. Perhaps a natural idea is to set the number c_i of residents ($p_i \le c_i \le q_i$) assigned to each hospital $h_i[p_i, q_i]$, so that the sum of c_i's over all the hospitals is equal to the number of residents. However, there is no obvious way of setting such c_i's rather than exhaustive search, which will result in blow-ups of the computation time even if t is small. Furthermore, even if we would be able to find suitable setting of c_i's, we are still not sure how to assign the residents to hospitals optimally (see the example of Sec. 2).

However, once we guess a set of blocking pairs included in a matching, we can easily test whether it is a correct guess or not by using the Gale-Shapley algorithm and the Rural Hospitals Theorem. Based on this observation, we will show an $O((|H||R|)^{t+1})$-time exact algorithm for Min-BP HRLQ.

Theorem 3 *There is an $O((|H||R|)^{t+1})$-time exact algorithm for Min-BP HRLQ, where t is the number of blocking pairs in an optimal solution of a given instance.*

Remark. This theorem is better than the conference version [20] in two senses: (i) The running time is improved from $O(t^2(|H|(|R| + t))^{t+1})$ to $O((|H||R|)^{t+1})$. (ii) The current algorithm can be applied to general Min-BP HRLQ, while one in the conference version can be applied only to 0-1 Min-BP HRLQ.

Proof For a given integer $k > 0$, the following procedure $A(k)$ finds a solution (i.e., a matching between residents and hospitals) whose cost (i.e., the number of blocking pairs) is at most k if any. Starting from $k = 1$, our algorithm (Algorithm II) runs $A(k)$ until it finds a solution, by increasing the value of k one by one. $A(k)$ is quite simple, for which the following informal discussion suffices.

Let I be a given instance. First, we guess a set B of k blocking pairs. Since there are at most $|H||R|$ pairs, there are at most $(|H||R|)^k$ choices of B. For each $(r, h) \in B$, we remove h from r's preference list (and r from h's list). Let I' be the modified instance. We then apply the Gale-Shapley algorithm to I'. If all the lower quotas are satisfied, then it is a desired solution, otherwise, we fail and proceed to the next guess.

We show that Algorithm II runs correctly. Consider any optimal solution M_{opt} and consider the execution of $A(k)$ for $k = t$ for which our current guess B contains exactly the t blocking pairs of M_{opt}. Then, it is not hard to see that M_{opt} is stable in I' and satisfies all the lower quotas. Then by the Rural Hospitals Theorem, any stable matching for I' satisfies all the lower quotas. Hence if we apply the Gale-Shapley algorithm to I', we find a matching M that satisfies all the lower quotas. Note that M has no blocking pair in I'. Then, M has at most t blocking pairs in the original instance I because, when a removed hospital h is returned back to the preference list of r, only (r, h) can be a new blocking pair.

Finally, we bound the time-complexity of Algorithm II. For each k, we apply the Gale-Shapley algorithm to at most $(|H||R|)^k$ instances, where each execution can be done in time $O(|H||R|)$. Therefore, the time-complexity is $O((|H||R|)^{k+1})$ for each k. Since we find a solution when k is at most t, the whole time-complexity is at most $\Sigma_{k=1}^{t} O((|H||R|)^{k+1}) = O((|H||R|)^{t+1})$. □

4 Minimum-Blocking-Resident HRLQ

In this section, we consider the problem of minimizing the number of blocking residents.

4.1 NP-hardness

We first show a hardness result.

Theorem 4 *Min-BR 1ML-HRLQ is NP-hard even if all the preference lists are complete.*

Proof We will show a polynomial-time reduction from the NP-complete problem CLIQUE [15]. In CLIQUE, we are given a graph $G = (V, E)$ and a positive integer $K \le |V|$, and asked if G contains a complete graph with K vertices as a subgraph.

Let $I_0 = (G_0, K_0)$ be an instance of CLIQUE where $G_0 = (V_0, E_0)$ and $0 < K_0 \le |V_0|$. We will construct an instance I of Min-BR 1ML-HRLQ. Let $n = |V_0|$, $m = |E_0|$, and B be a positive integer such that $B > 2K_0$. Let $R = C \cup E$ be the set of residents and $H = V \cup \{x\}$ be the set of hospitals of I. Each set is defined as $C = \{c_i \mid 1 \le i \le K_0\}$, $E = \{e^k_{i,j} \mid (v_i, v_j) \in E_0, 1 \le k \le B\}$, and $V = \{v_i \mid 1 \le i \le n\}$. (There is a one-to-one correspondence between the set V of hospitals and the set V_0 of vertices, so we use the same symbol v_i to refer to both vertex and the corresponding hospital.)

Corresponding to each edge $(v_i, v_j) \in E_0$, there are B residents $e^k_{i,j} (1 \le k \le B)$. We will call them *residents associated with* (v_i, v_j). Preference lists and quotas are given in Fig. 3. For a set X, "$[X]$" means an arbitrarily (but fixed) ordered list of the members in X, and "$\ldots$" means an arbitrarily ordered list of all the other hospitals that do not appear explicitly in the list. Note that all the preference lists are complete, and all the hospitals have the same preference list.

$$
\begin{array}{lll}
c_i & : [V]\ x & (1 \le i \le K_0) \\
e^k_{i,j} & : v_i \quad v_j \quad x \ \ldots & ((v_i, v_j) \in E_0, 1 \le k \le B) \\
 & & \\
v_i[0,1] & : [C]\ [E] & (1 \le i \le n) \\
x[mB, mB] & : [C]\ [E] &
\end{array}
$$

Fig. 3 Preference lists of residents and preference lists and quotas of hospitals

Lemma 5 *If I_0 is a "yes" instance of CLIQUE, then there is a feasible matching of I having at most $(m - \binom{K_0}{2})B + K_0$ blocking residents.*

Proof Suppose that G_0 has a clique V'_0 of size K_0. We will construct a matching M of I from V'_0. We assign all the residents in C to the hospitals in V'_0 in an arbitrary way, and all the residents in E to the hospital x. Since V'_0 is a clique, $(v_i, v_j) \in E_0$ for any pair of $v_i, v_j \in V'_0 (i \ne j)$. There are B residents $e^k_{i,j}$ $(1 \le k \le B)$ associated with the edge (v_i, v_j). These residents are assigned to the hospital x inferior to the hospitals v_i and v_j in M, but the hospitals v_i and v_j are assigned residents in C, better than $e^k_{i,j}$. Hence all $e^k_{i,j}$ are non-blocking residents. There are $B\binom{K_0}{2}$ such residents $e^k_{i,j}$ and the total number of residents is $mB + K_0$. Hence there are at most $(m - \binom{K_0}{2})B + K_0$ blocking residents in M. $\square$

Lemma 6 *For a matching X of I, let cost(X) be the number of blocking residents of X. For an arbitrary feasible matching M of I, there is a feasible matching M' of I such that (i) M' assigns every resident in C to a hospital in V and (ii) cost$(M') \leq$ cost$(M) + K_0$.*

Proof First, if some residents are unassigned in M, we modify M by assigning them to arbitrary hospitals. This is possible because all the preference lists are complete and the number of residents is at most the sum of the upper quotas. Clearly, this does not increase the cost. Let $C_x = \{c \mid c \in C, M(c) = x\}$ and $E_v = \{e \mid e \in E, M(e) \in V\}$. Then, $|C_x| = |E_v|$ since $|M(x)| = |C_x| + (|E| - |E_v|)$ and $|M(x)| = mB = |E|$ by the lower quota of x. If C_x is empty, we are done because we can let $M' = M$. Hence, suppose that C_x is nonempty. Let M' be a matching obtained by M by exchanging assigned hospitals between C_x and E_v arbitrarily. Then M' is feasible and the following (1)–(3) are easy to verify:

(1) Any resident in $C \setminus C_x$ does not change its assigned hospital, and no hospital in V becomes worse off. Therefore, no new blocking resident arises from $C \setminus C_x$. (2) Any resident r in C_x is a blocking resident in M because r is assigned to x and there is a hospital in V that receives a resident from E_v. Therefore, no new blocking resident arises from C_x. (3) For the same reason as (1), no new blocking resident arises from $E \setminus E_v$.

Hence, only residents in E_v can newly become blocking residents. Since $|E_v| = |C_x| \leq |C| = K_0$, we have that cost$(M') \leq$ cost$(M) + K_0$. □

Lemma 7 *If I_0 is a "no" instance of CLIQUE, then any feasible matching of I contains at least $(m - \binom{K_0}{2} + 1)B - K_0$ blocking residents.*

Proof Suppose that there is a matching M of I that contains less than $(m - \binom{K_0}{2} + 1)B - K_0$ blocking residents. We will show that G_0 contains a clique of size K_0. We first construct a matching M' using Lemma 6. Then M' contains less than $(m - \binom{K_0}{2} + 1)B$ blocking residents, and any resident in C is assigned to a hospital in V. Note that every resident in E is now assigned to x since x's lower quota is $mB = |E|$. Define $V'_0 \subseteq V_0$ be the set of vertices corresponding to the assigned hospitals in V. Clearly, $|V'_0| = K_0$. We claim that V'_0 is a clique.

Recall that there are $mB + K_0$ residents. Since we assume that there are less than $(m - \binom{K_0}{2} + 1)B$ blocking residents, there are more than $K_0 + \binom{K_0}{2}B - B$ non-blocking residents, and since $|C| = K_0$, there are more than $\binom{K_0}{2}B - B$ non-blocking residents in E. Consider the following partition of E into B subsets: $E_k = \{e^k_{i,j} \mid (v_i, v_j) \in E_0\}$ $(1 \leq k \leq B)$. Then the above observation on the number of non-blocking residents in E implies that there is a k such that E_k contains at least $\binom{K_0}{2}$ non-blocking residents. Since every resident in E is assigned to x, only $e^k_{i,j}$ such that both v_i and v_j are in V'_0 can be non-blocking. This means that any pair of vertices in V'_0 causes such a non-blocking resident, implying that V'_0 is a clique. □

Because $B > 2K_0$, we have $(m - \binom{K_0}{2} + 1)B - K_0 > (m - \binom{K_0}{2})B + K_0$. Hence by Lemmas 5 and 7, Min-BR 1ML-HRLQ is NP-hard. □

We can prove the NP-hardness for more restricted case using the following Lemma 8. Since the same reduction will be used in the approximability part (Sec. 4.2), we state the lemma in a stronger form than is needed here.

Lemma 8 *If there is a polynomial-time α–approximation algorithm for 0-1 Min-BR HRLQ, then there is a polynomial-time α–approximation algorithm for Min-BR HRLQ.*

Proof We give a polynomial-time approximation preserving reduction from Min-BR HRLQ to 0-1 Min-BR HRLQ. Let I be an instance of Min-BR HRLQ. We construct an instance I' of 0-1 Min-BR HRLQ in polynomial time: The set of residents of I' is the same as that of I. Corresponding to each hospital $h_i[p_i, q_i]$ of I, I' contains p_i hospitals $h_{i,1}, \cdots, h_{i,p_i}$ with quota $[1, 1]$, and $q_i - p_i$ hospitals $h_{i,p_i+1}, \cdots, h_{i,q_i}$ with quota $[0, 1]$. For any j, the preference list of a hospital $h_{i,j}$ of I' is the same as that of a hospital h_i of I. The preference list of a resident r of I' is constructed from the preference list of the corresponding resident in I by replacing h_i by $h_{i,1} \ \cdots \ h_{i,q_i}$ for each hospital h_i of I. Without loss of generality, we can assume that $q_i \leq |R|$ for each i. Hence I' can be constructed in polynomial time.

From a feasible matching M' for I', it is easy to construct a feasible matching M for I; just adding (r, h_i) to M for each $(r, h_{i,j}) \in M'$. Let *cost*, *cost*$'$, *opt* and *opt*$'$ be the costs of M, M', the optimal costs of I and I', respectively. In order to complete the proof, we must show that $\frac{cost}{opt} \leq \frac{cost'}{opt'}$. To this end, it is enough to show (i) $cost \leq cost'$, and (ii) $opt' \leq opt$. For (i), it is easy to verify that if r is a blocking resident for M, then so is r for M' too. For (ii), we show that from (any) matching X for I, we can construct a matching X' for I' without increasing the cost. Consider a hospital h_j. Let $r_{j,1}, r_{j,2}, \cdots, r_{j,|X(h_j)|}$ be the residents in $X(h_j)$ and suppose that h_j prefers these residents in this order. We construct a matching X' by adding $(r_{j,k}, h_{j,k})$ to X' for all k and j. Again, it is easy to see that X' is feasible for I' and if r is a blocking resident for X', then r is also a blocking resident for X. □

Corollary 1 *0-1 Min-BR 1ML-HRLQ is NP-hard even if all the preference lists are complete.*

Proof Note that the reduction in the proof of Lemma 8 preserves the "1ML" property and the completeness of the preference lists. Then the corollary is immediate from Theorem 4 and Lemma 8. □

4.2 Approximability

For the approximability, we note that Algorithm I in the proof of Theorem 2 does not work. For example, consider the instance introduced in Sec. 2. If we apply the Gale-Shapley algorithm, resident r_i is assigned to h_i for each i, and we need to move r_1 to h_{n+1}. However since h_1 becomes empty, all the residents become blocking residents. On the other hand, the optimal cost is 2 as we have seen there. Thus the approximation ratio becomes as bad as $\Omega(|R|)$.

Theorem 5 *There is a polynomial-time $\sqrt{|R|}$–approximation algorithm for Min-BR HRLQ.*

We know by Lemma 8 that it is enough to attack 0-1 Min-BR HRLQ. Hence we give a $\sqrt{|R|}$–approximation algorithm for 0-1 Min-BR HRLQ (Lemma 10) to prove Theorem 5. In 0-1 Min-BR HRLQ, the number of residents assigned to each hospital is at most one. Hence, for a matching M, we sometimes abuse the notation $M(h)$ to denote the resident assigned to h (if any) although it was originally defined as the *set* of residents assigned to h.

4.2.1 Algorithm

To describe the idea behind our algorithm, recall again Algorithm I presented in the proof of Theorem 2: First, apply the Gale-Shapley algorithm to a given instance I and obtain a matching M. Next, move residents arbitrarily from assigned $[0,1]$-hospitals to empty $[1,1]$-hospitals. Suppose that in the course of the execution of Algorithm I, we move a resident r from a $[0,1]$-hospital h to an empty $[1,1]$-hospital. Then, of course r creates a blocking pair with h, but some other residents may also create blocking pairs with h because h becomes empty. Hence, consider the following modification. First, set the upper quota of a $[0,1]$-hospital h to ∞ and apply the Gale-Shapley algorithm. Then, all residents who "wish" to go to h actually go there. Hence, even if we move all such residents to other hospitals, only the moved residents can become blocking residents. By doing this, we can bound the number of blocking residents by the number (given by the function g introduced below) of those moving residents. In the above example, we extended the upper quota of only one hospital, but in fact, we may need to select two or more hospitals to select sufficiently many residents to be sent to other hospitals so as to make the matching feasible. However, at the same time, this number should be kept minimum to guarantee the quality of the solution.

As mentioned above, we define $g(h,h)$: For an instance I of HR, suppose that we extend the upper quota of hospital h to ∞ and find a stable matching of this new instance. Define $g(h,h)$ as the number of residents who are assigned to h in this stable matching. Recall that this quantity does not depend on the choice of the stable matching by the Rural Hospitals Theorem [14]. Extend $g(h,h)$ to $g(A,B)$ for $A,B \subseteq H$ such that $g(A,B)$ denotes the number of residents assigned to hospitals in A when we change upper quotas of all the hospitals in B to ∞.

We now propose Algorithm III for 0-1 Min-BR HRLQ. The idea is to find a small number of residents (*victims*) to be moved, and construct a feasible matching M^* in which only the victims are blocking. First we apply the Gale-Shapley algorithm to a given instance I while ignoring the lower quotas of I and obtain a matching M_s. The matching M_s is used to find non-empty $[0,1]$-hospitals (denoted $H'_{0,1}$ in the description of Algorithm III) from which the victims will be selected. Next, we estimate the popularity of the hospital h in $H'_{0,1}$ using $g(h,h)$ defined above, and select a certain number of least popular hospitals S from $H'_{0,1}$ (we will later show that $H'_{0,1}$ is large enough to select S). We then apply the Gale-Shapley algorithm again while setting the upper quotas of hospitals in S to ∞ and obtain a matching M_∞. The residents who came to hospitals in S are victims and we move these residents to the empty $[1, 1]$-hospitals to obtain the final solution M^* (we will later show that there are enough number of victims to fill the empty $[1, 1]$-hospitals). We can show that the number of victims is small enough because we have selected less popular hospitals to S.

We will introduce notations used to describe Algorithm III formally. Let I be a given instance. Define $H_{p,q}$ to be the set of $[p, q]$-hospitals of I. Recall from Sec. 3 that the deficiency of a hospital is the shortage of the assigned residents from its lower quota (with respect to the matching obtained by the Gale-Shapley algorithm). Now define the *deficiency of the instance* I as the sum of the deficiencies of all the hospitals of I, and denote it $D(I)$. Since we are considering 0-1 Min-BR HRLQ, $D(I)$ is exactly the number of empty $[1, 1]$-hospitals.

Algorithm III

1: Apply the Gale-Shapley algorithm to I by ignoring the lower quotas. Let M_s be the obtained matching. Compute the deficiency $D(I)$.
2: $H'_{0,1} := \{h \mid M_s(h) \neq \emptyset, h \in H_{0,1}\}$. (If $M_s(h) = \emptyset$, then residents never come to h in the following Steps 3 and 4.)
3: Compute $g(h, h)$ for each $h \in H'_{0,1}$ by using the Gale-Shapley algorithm.
4: From $H'_{0,1}$, select $D(I)$ hospitals with smallest $g(h, h)$ values (ties are broken arbitrarily). Let S be the set of these hospitals. Extend the upper quotas of all hospitals in S to ∞, and run the Gale-Shapley algorithm. Let M_∞ be the obtained matching.
5: In M_∞, move residents who are assigned to hospitals in S arbitrarily to empty hospitals to make the matching feasible. (We first make $[1, 1]$-hospitals full. This is possible because of the NR-assumption and the CL-restriction. If there is a hospital in S still having two or more residents, then send surplus residents arbitrarily to empty $[0, 1]$-hospitals, or simply make them unassigned if there is no $[0, 1]$-hospital to send them to.) Output the resulting matching M^*.

We first prove the following property of the original HR problem.

Lemma 9 *Let I_0 be an instance of HR, and h be any hospital. Let I_1 be a modification of I_0 so that only the upper quota of h is increased by 1. Let M_i be a stable matching of I_i for each $i \in \{0, 1\}$. Then, (i) $|M_0(h)| \leq |M_1(h)|$, and (ii) $\forall h' \in H \setminus \{h\}$, $|M_0(h')| \geq |M_1(h')|$.*

Proof If M_0 is stable for I_1, then we are done, so suppose not. We will construct a stable matching for I_1 by successive modifications starting from M_0. Because M_0 is stable for I_0, if M_0 has blocking pairs for I_1, then all of them involve h. Let r be the resident such that (r, h) is a blocking pair and there is no blocking pair (r', h) such that h prefers r' to r. If we assign r to h (possibly by canceling the previous assignment of r if r was assigned in M_0), all the blocking pairs including h are removed. If no new blocking pairs arise, again we are done. Otherwise, r must be previously assigned to some hospital, say

h', and all the new blocking pairs involve h'. We then choose the resident r', most preferred by h' among all the blocking residents, and assign r' to h'. We continue this operation until there arise no new blocking pairs. This procedure eventually terminates because each iteration improves exactly one resident. By the termination condition, the resulting matching is stable for I_1. Note that by this procedure, only h can gain one more resident, and at most one hospital may lose one resident. By the Rural Hospitals Theorem, the number of residents assigned to each hospital is the same in M_1 and the current matching. This completes the proof. □

Obviously, Algorithm III runs in polynomial time. We show that Algorithm III runs correctly, namely that the output matching M^* satisfies the quotas. To do so, it suffices to show the following conditions

$$|H'_{0,1}| \geq D(I) \tag{1}$$

and

$$|\{r \mid M_\infty(r) \in S\}| \geq |\{h \mid h \in H_{1,1}, M_\infty(h) = \emptyset\}| \tag{2}$$

so that Step 4 and Step 5 are executable, respectively.

For (1), let N_1 be the number of residents assigned to hospitals in $H_{1,1}$ in M_s. Then $|M_s| = |H'_{0,1}| + N_1$ and $D(I) = |H_{1,1}| - N_1$. We can assume that all the residents are assigned in M_s since otherwise, we already have a feasible stable matching (as explained in the proof of Theorem 2) and therefore $|M_s| = |R|$. From these equations, we have $|H'_{0,1}| = D(I) + |R| - |H_{1,1}|$. By the NR-assumption, it follows that $|R| \geq |H_{1,1}|$, from which we have $|H'_{0,1}| \geq D(I)$ as required. For (2), it suffices to show that the number N_2 of residents assigned to $S \cup H_{1,1}$ in M_∞ is at least the number of hospitals in $H_{1,1}$, i.e., $|H_{1,1}|$. Note that empty hospitals in M_s are also empty in M_∞ by Lemma 9. Therefore, the number $\overline{N_2}$ of residents assigned to hospitals in $H \setminus (S \cup H_{1,1})$ in M_∞ is at most the number of hospitals in $H'_{0,1} \setminus S$. Thus $\overline{N_2} \leq |H'_{0,1}| - |S|$ and $N_2 = |R| - \overline{N_2} \geq |R| - (|H'_{0,1}| - |S|)$. By the definition of $D(I)$, we have that $|H'_{0,1}| + |H_{1,1}| = |R| + D(I)$. Thus, $N_2 \geq |R| - (|R| + D(I) - |H_{1,1}| - |S|) = |H_{1,1}|$ (recall that $|S| = D(I)$).

4.2.2 Analysis of the Approximation Ratio

Lemma 10 *The approximation ratio of Algorithm III is at most* $\sqrt{|R|}$.

Proof Let I be a given instance of 0-1 Min-BR HRLQ and let f_{opt} and f_{alg} be the costs of an optimal solution and the solution obtained by Algorithm III, respectively. First, note that any resident r who is assigned to a hospital $h \in H \setminus S$ in M_∞ prefers no hospital in S to h, since otherwise, r and such a hospital (in S) form a blocking pair for M_∞, a contradiction (recall that the upper quota of any hospital in S is ∞). Therefore, even if we move residents from hospitals in S at Step 5, no unmoved resident becomes a blocking resident. Thus only moved residents can be blocking residents and

$$f_{alg} \leq g(S, S). \tag{3}$$

We then give a lower bound on the optimal cost. To do so, recall the proof of Theorem 2, where it is shown that any optimal solution for instance I of Min-BP HRLQ has at least $D(I)$ blocking pairs. It should be noted that those $D(I)$ blocking pairs do not have any common resident. Thus we have

$$f_{opt} \geq D(I). \tag{4}$$

Now here is our key lemma to evaluate the approximation ratio.

Lemma 11 *In Step 3 of Algorithm III, there are at least $D(I)$ different hospitals $h \in H'_{0,1}$ such that $g(h, h) \leq f_{opt}$.*

The proof will be given in a moment. By this lemma, we have $g(h, h) \leq f_{opt}$ for any $h \in S$, since at Step 4 of Algorithm III, we select $D(I)$ hospitals with the smallest $g(h, h)$ values. This implies that

$$\sum_{h \in S} g(h, h) \leq D(I) f_{opt}. \tag{5}$$

Also, by Lemma 9, we have

$$g(h, S) \leq g(h, h) \tag{6}$$

for any $h \in S$. Hence, by (3), (6), (5) and (4), we have

$$f_{alg} \leq g(S, S) = \sum_{h \in S} g(h, S) \leq \sum_{h \in S} g(h, h) \leq D(I) f_{opt} \leq (f_{opt})^2.$$

Therefore, we have that $\sqrt{f_{alg}} \leq f_{opt}$, and hence $\frac{f_{alg}}{f_{opt}} \leq \sqrt{f_{alg}} \leq \sqrt{|R|}$, completing the proof of Lemma 10. $\square$

Proof of Lemma 11. Let h be a hospital satisfying the condition of the lemma. In order to calculate $g(h, h)$ in Step 3, we construct a stable matching, say M_h for the instance $I_\infty(h)$ in which the upper quota of h is changed to ∞. We do not know what kind of matching M_h is, but in the following, we show that there is a stable matching, say M_2, for $I_\infty(h)$ such that $|M_2(h)| \leq f_{opt}$. Matchings M_h and M_2 may be different matchings, but we can guarantee that $|M_h(h)| = |M_2(h)| \leq f_{opt}$ by the Rural Hospitals Theorem. A bit trickily, we construct M_2 from an optimal matching.

Let M_{opt} be an optimal solution of I (which of course we do not know). Let R_b and R_n be the sets of blocking residents and non-blocking residents for M_{opt}, respectively. Then $|R_b| = f_{opt}$ by definition. We modify M_{opt} as follows: Take any resident $r \in R_b$. If r is unassigned, we do nothing. Otherwise, force r to be unassigned. Then there may arise new blocking pairs involving residents in R_n. Let BP_1 be the set of such new blocking pairs. Note that all of the blocking pairs in BP_1 include the hospital h' to which r was assigned. Among the residents involved in BP_1, we select the resident r' who is most preferred by h' and assign r' to h'. Then, all the blocking pairs in BP_1 disappear. However, there may arise new blocking pairs (BP_2) involving residents in R_n, and all the blocking pairs in BP_2 include the hospital h'' to which r' was assigned. In a similar way as the proof of Lemma 9, we continue to move residents until no new blocking resident arises from R_n (but this time, we move only residents in R_n as explained above). We do this for all the residents in R_b, and let M_1 be the resulting matching.

The following (7) and (8) are immediate:

$$\text{There are at least } f_{opt} \text{ unassigned residents in } M_1, \tag{7}$$

since residents in R_b are unassigned in M_1.

$$\text{Residents in } R_n \text{ are non-blocking for } M_1. \tag{8}$$

We prove the following properties:

$$\text{There are at most } f_{opt} \text{ empty } [1,1]\text{-hospitals in } M_1. \tag{9}$$

Define $H' = \{h \mid h \in H'_{0,1} \text{ and } h \text{ is empty in } M_1\}$. Then

$$|H'| \geq D(I). \tag{10}$$

For (9), note that all the $[1,1]$-hospitals are full in M_{opt}. It is easy to see that, in the above procedure for each $r \in R_b$, at most one assigned hospital is made empty. Since $|R_b| = |f_{opt}|$, the number of such hospitals is at most $|f_{opt}|$ and hence the claim holds.

For (10), let H_1 be the set of hospitals assigned in M_1. We have that

$$H' = H'_{0,1} \setminus (H_1 \cap H_{0,1}) \tag{11}$$

by the definition of H', and that

$$|H'_{0,1}| = |R| + D(I) - |H_{1,1}| \tag{12}$$

by the definition of $D(I)$. Also, the above property (7) implies that $|R| - |H_1| \geq f_{opt}$ and (9) implies that $|H_{1,1}| - |H_1 \cap H_{1,1}| \leq f_{opt}$, from which we have that

$$\begin{aligned} |H_1 \cap H_{0,1}| &= |H_1| - |H_1 \cap H_{1,1}| \\ &\leq (|R| - f_{opt}) + (f_{opt} - |H_{1,1}|) \\ &= |R| - |H_{1,1}|. \end{aligned} \tag{13}$$

From (11), (12), and (13), we have $|H'| \geq |H'_{0,1}| - |H_1 \cap H_{0,1}| \geq (|R| + D(I) - |H_{1,1}|) - (|R| - |H_{1,1}|) = D(I)$, as required.

Let h be an arbitrary hospital in H'. We show that $g(h,h) \leq f_{opt}$. Then, this completes the proof of Lemma 11 because $H' \subseteq H'_{0,1}$ and $|H'| \geq D(I)$ (10). Since h is empty in M_1, residents in R_n are still non-blocking for M_1 in $I_\infty(h)$ (whose definition is in the beginning of this proof) by the property (8). Now, choose any resident r from R_b, and apply the Gale-Shapley algorithm to $I_\infty(h)$ starting from M_1. This execution starts from the proposal by r, and at the end, nobody in $R_n \cup \{r\}$ is a blocking resident for $I_\infty(h)$. Since hospitals assigned in M_1 never become empty, and since unassigned residents in R_n never become assigned, h receives at most one resident. If we do this for all the residents in R_b, the resulting matching M_2 is stable for $I_\infty(h)$, and h is assigned at most $|R_b| = f_{opt}$ residents. As mentioned previously, this implies $g(h,h) \leq f_{opt}$. $\square$

4.2.3 Tightness of the Analysis

We give an instance of 0-1 Min-BR HRLQ for which Algorithm III produces a solution of cost $|R| - \sqrt{|R|}$ but the optimal cost is at most $2\sqrt{|R|}$. Namely, the analysis of Lemma 10 is tight up to a constant factor.

Let $R = C \cup D \cup E$ and $H = A \cup B \cup X$, where $C = \{c_i \mid 1 \leq i \leq n\}$, $D = \{d_{i,j} \mid 1 \leq i \leq n, 1 \leq j \leq n-2\}$, $E = \{e_i \mid 1 \leq i \leq n\}$, $A = \{a_i \mid 1 \leq i \leq n\}$, $B = \{b_i \mid 1 \leq i \leq n\}$, and $X = \{x_i \mid 1 \leq i \leq n^2 - n\}$. The preference lists of residents are

$$\begin{array}{lll} c_i & : a_i \ b_i \quad [[X]] \ \cdots & (1 \leq i \leq n) \\ d_{i,j} & : b_i \ [[X]] \ \cdots & (1 \leq i \leq n, 1 \leq j \leq n-2) \\ e_i & : b_i \ [[A]] \ [[X]] \ \cdots & (1 \leq i \leq n) \end{array}$$

and the preference lists and quotas of hospitals are

$$\begin{array}{lll} a_i[0,1] : c_i & \cdots & (1 \leq i \leq n) \\ b_i[0,1] : d_{i,1} & \cdots & (1 \leq i \leq n) \\ x_i[1,1] : \cdots & & (1 \leq k \leq n^2 - n) \end{array}$$

where $[[X]]$ denotes $x_1 \ \cdots \ x_{n^2-n}$ and $[[A]]$ denotes $a_1 \ \cdots \ a_n$. "$\cdots$" denotes an arbitrarily ordered list of the members that do not appear explicitly. Note that all the preference lists are complete. The deficiency of this instance is n. If we set the upper quota of a_i to ∞, then $n+1$ residents $c_i, e_1, e_2, \ldots, e_n$ are assigned to a_i, so $g(a_i, a_i) = n+1$ for all $1 \leq i \leq n$. If we set the upper quota of b_i to ∞, then $n-1$ residents $e_i, d_{i,1}, d_{i,2}, \ldots, d_{i,n-2}$ are assigned to b_i, so $g(b_i, b_i) = n-1$. Thus, Algorithm III constructs $S = \{b_1, \cdots, b_n\}$ at Step 4 and the solution has $n^2 - n = |R| - \sqrt{|R|}$ blocking residents. However, consider the following matching: First, apply the Gale-Shapley algorithm for D and $B \cup X$. Then, assign the residents in $C \cup E$ to the empty hospitals in X arbitrarily. Then, nobody in D can be a blocking resident. Hence the cost is at most $2n = 2\sqrt{|R|}$. Therefore, the approximation ratio is at least $(|R| - \sqrt{|R|})/(2\sqrt{|R|}) = \Omega(\sqrt{|R|})$.

4.3 Inapproximability

For the hardness of Min-BR HRLQ, we have only NP-hardness, but we can give a strong evidence for its inapproximability. The *Densest k-Subgraph Problem* (DkS) is the problem of finding, given a graph G and a positive integer k, an induced subgraph of G with k vertices that contains as many edges as possible. This problem is NP-hard because it is a generalization of Max CLIQUE. Its approximability has been studied intensively but there still remains a large gap between approximability and inapproximability: The best known approximation ratio is $|V|^{1/4+\epsilon}$ [5], while there is no PTAS under reasonable assumptions [9, 24]. The following Theorem 6 shows that approximating Min-BR HRLQ within a constant ratio implies the same for DkS.

Theorem 6 *If Min-BR 1ML-HRLQ has a polynomial-time c-approximation algorithm, then DkS has a polynomial-time $(1+\epsilon)c^4$-approximation algorithm for any positive constant ϵ.*

Proof The proof uses another problem called *Minimum Coverage Problem* (MinC) [30]. In MinC, we are given a family $\mathcal{P}$ of subsets of a base set $\mathcal{U}$ and a positive integer t, and asked to select t sets from $\mathcal{P}$ so that their union is minimized. Theorem 6 can be easily proved by combining the following two lemmas, whose proofs will be given shortly:

Lemma 12 *If MinC admits a polynomial-time c-approximation algorithm, then DkS admits a polynomial-time $(1+\epsilon)c^4$-approximation algorithm for any positive constant ϵ.*

Lemma 13 *If Min-BR 1ML-HRLQ admits a polynomial-time d-approximation algorithm, then MinC admits a polynomial-time $(1+\epsilon)d$-approximation algorithm for any positive constant ϵ.*

Suppose that Min-BR 1ML-HRLQ admits a polynomial-time c-approximation algorithm. Given an arbitrary positive constant ϵ, we choose ϵ' such that $\epsilon' \leq (1+\epsilon)^{\frac{1}{5}} - 1$ in Lemmas 12 and 13. By Lemma 13, MinC admits a polynomial-time $(1+\epsilon')c$-approximation algorithm and then by Lemma 12, DkS admits a polynomial-time $(1+\epsilon')^5 c^4$-approximation algorithm. By the choice of ϵ', we have $(1+\epsilon')^5 c^4 \leq (1+\epsilon)c^4$, and hence the proof of Theorem 6 is completed. □

Proof of Lemma 12. We will construct a polynomial-time $(1+\epsilon)c^4$-approximation algorithm for DkS using a c-approximation algorithm A for MinC. Suppose that we are given a graph $G = (V, E)$ and an integer k as an instance I of DkS. We regard each vertex in V as an element and each edge in E as a set of size two containing its two endpoints, and consider it as an instance of MinC. Recall that in MinC, we are given a positive integer t which specifies the number of sets we must select. We repeatedly apply algorithm A to this instance by increasing the target value of t one by one from one, until A outputs a solution of cost $c(k+1)$ or more for the first time. Let $\tilde{t}$ be the value of t at this point and $\tilde{s}$ be the value output by A. (If A never outputs such a solution even when $t = |E|$, it means that $|V| < c(k+1)$ in the given graph. This is more desirable case for us, as shown below.) Then, $\tilde{s} \geq c(k+1)$ by the above condition, and the optimal value of MinC when the target value is $\tilde{t}$ is at least $k+1$ since A is a c-approximation algorithm. This means that there is no subset of k vertices in G containing $\tilde{t}$ edges; in other words, the optimal value of the DkS instance I is less than $\tilde{t}$.

Note that when the target values in MinC differ by one, the two corresponding optimal values differ by at most two because adding one edge increases the number of vertices by at most two. Therefore, $\tilde{s} \leq c^2(k+1)+c$ since otherwise, $\tilde{s} > c^2(k+1)+c$ and the optimal value of MinC when the target value is $\tilde{t}$ is more than $c(k+1)+1$, namely at least $c(k+1)+2$, because A is a c-approximation algorithm. Then, when the target value is $\tilde{t}-1$, the optimal value of MinC is at least $c(k+1)$ by the above observation, and hence A must have already output a solution of value at least $c(k+1)$, a contradiction.

We now have a subgraph G' of G with $\tilde{s}$ vertices and at least $\tilde{t}$ edges. We then solve DkS approximately for G' (with the same k) using the greedy algorithm given in [4]. We can find a subgraph of G' with k vertices and at least $\frac{k(k-1)}{\tilde{s}(\tilde{s}-1)}\tilde{t}$ edges, which is a $\frac{\tilde{s}(\tilde{s}-1)}{k(k-1)}$-approximate solution of the original problem I (recall that the optimal value of I is less than $\tilde{t}$). Since $\tilde{s} \leq c^2(k+1)+c$ as proved above,

$$\frac{\tilde{s}(\tilde{s}-1)}{k(k-1)} \leq c^4 + \frac{(3k+1)c^4 + 2(k+1)c^3 - kc^2 - c}{k(k-1)}.$$

Note that for any fixed constants c and ϵ, we can find a constant k_0 such that $\frac{(3k+1)c^4+2(k+1)c^3-kc^2-c}{k(k-1)} \leq \epsilon c^4$ for all $k \geq k_0$. Also, note that DkS when k is a constant is solvable in polynomial time. Thus, given a DkS instance, solving optimally when $k < k_0$, and using the above reduction otherwise, is a desirable $(1+\epsilon)c^4$-approximation algorithm.

If A does not output a solution when determining $\tilde{s}$, we know that $|V| < c(k+1)$ as discussed previously. In this case we simply apply the above greedy algorithm to G itself instead of G'. The optimal cost is at most $|E|$ and the algorithm's cost is at least $\frac{k(k-1)}{|V|(|V|-1)}|E|$, so the approximation ratio is at most $\frac{|V|(|V|-1)}{k(k-1)}$. By a similar argument as above, we can show that this is bounded by $(1+\epsilon)c^2$ for any positive ϵ for large enough k. This completes the proof. $\square$

Proof of Lemma 13. We give a polynomial-time reduction from MinC to Min-BR 1ML-HRLQ. Suppose that a given instance I_0 of MinC consists of the base set $\mathcal{U}=\{u_1, u_2, \ldots, u_n\}$, a collection $\mathcal{P}=\{P_1, P_2, \ldots, P_m\}$ of subsets of $\mathcal{U}$, and a positive integer t (the number of subsets to be selected). We construct an instance I of Min-BR 1ML-HRLQ.

Let $R = C \cup U$ be the set of residents and $H = P \cup \{x\}$ be the set of hospitals, where each set is defined as follows: $C = \{c_i \mid 1 \leq i \leq m-t\}$, $U = \{u_i^j \mid 1 \leq i \leq n, 1 \leq j \leq B\}$, and $P = \{p_i \mid 1 \leq i \leq m\}$. Note that $|R| = nB + m - t$. Here, B is a positive integer determined later. Preference lists and quotas are defined in Fig. 4. For each i $(1 \leq i \leq n)$, residents u_i^j $(1 \leq j \leq B)$ correspond to the element u_i of the base set $\mathcal{U}$ of MinC. Each $[0,1]$-hospital p_i corresponds to the subset P_i of MinC instance I_0. For each resident u_i^j, the set $P(i)$ contains the hospital p_k if and only if the element u_i is contained in the set P_k in I_0. For a set S, "$[S]$" denotes an arbitrarily ordered list of the members in S. Note that all the preference lists of hospitals are identical. It is easy to see that the reduction can be performed in polynomial time.

c_i	$: [P]$	x		$(1 \leq i \leq m-t)$
u_i^j	$: [P(i)]$	x	$[P \setminus P(i)]$	$(1 \leq i \leq n, 1 \leq j \leq B)$
$p_i[0,1]$	$: [C]$	$[U]$		$(1 \leq i \leq m)$
$x[nB, nB]$	$: [C]$	$[U]$		

Fig. 4 Preference lists of residents and hospitals

Let $opt(I_0)$ and $opt(I)$ be the optimal costs of I_0 and I, respectively. In the following, we show that (i) $opt(I) \leq B \cdot opt(I_0) + (m-t)$, and (ii) from a solution of I of cost a, we can construct a solution of I_0 of cost at most $(a+m-t)/B$ in polynomial time.

Hence, if there is a polynomial-time d-approximation algorithm for Min-BR 1ML-HRLQ, namely, if $\frac{a}{opt(I)} \leq d$, then we can obtain

$$\begin{aligned}\frac{(a+m-t)/B}{opt(I_0)} &\leq d + \frac{(d+1)(m-t)}{B \cdot opt(I_0)}\\ &\leq d + \frac{2md}{B \cdot opt(I_0)}\\ &\leq (1 + \frac{2m}{B})d.\end{aligned}$$

Now, if we take $B \geq \frac{2}{\epsilon}m$, then $(1+\frac{2m}{B})d \leq (1+\epsilon)d$, as desired.

We first prove (i). Let $\mathcal{P}^*$ be an optimal solution (a subset of size t) for I_0. We will construct a solution M of I as follows: Let $M(u_i^j) = x$ for all i and j. Assign residents in C to hospitals corresponding to subsets in $\mathcal{P} \setminus \mathcal{P}^*$ in an arbitrary way. For each $P_j \in \mathcal{P}^*$, let the hospital p_j be empty. Consider a resident u_i^j and consider a subset P_k of I_0 that contains the element u_i. Note that u_i^j prefers the hospital p_k to x. If $P_k \notin \mathcal{P}^*$, then p_k receives a resident better than u_i^j in M and hence (u_i^j, p_k) is not a blocking pair. If $P_k \in \mathcal{P}^*$, then p_k is empty in M and hence (u_i^j, p_k) is a blocking pair. Hence, $\mathcal{P}^*$ does not include any P_k that contains u_i (in other words, the element u_i does not contribute to the cost of $\mathcal{P}^*$) if and only if u_i^j is not a blocking resident. There are $(m-t)+nB$ residents and among them $B(n - opt(I_0))$ are non-blocking as observed. Thus the number of blocking residents for M is at most $(m-t)+nB-B(n-opt(I_0)) = B \cdot opt(I_0) + (m-t)$, which completes the proof of (i).

We then prove (ii). Consider a feasible matching M of cost a. We may assume without loss of generality that all the residents are assigned in M because if not, we can assign unassigned residents to under-subscribed hospitals arbitrarily without increasing the cost. Let $C_x = \{c \mid c \in C, M(c) = x\}$ and $U_p =$

$\{u \mid u \in U, M(u) \in P\}$. Then, $|C_x| = |U_p|$ since $|M(x)| = |C_x| + (|U| - |U_p|)$ and $|M(x)| = nB = |U|$ by the lower quota of x.

Let M' be a matching obtained by M by exchanging assigned hospitals between C_x and U_p arbitrarily. The following (1)–(3) are easy to verify: (1) Any resident in $C \setminus C_x$ does not change its assigned hospital, and no hospital in P becomes worse off. Therefore, no new blocking resident arises from $C \setminus C_x$. (2) Any resident r in C_x is a blocking resident in M because r is assigned to x and there is a hospital in P that receives a resident from U_p. Therefore, no new blocking resident arises from C_x. (3) For the same reason as (1), no new blocking resident arises from $U \setminus U_p$. Hence, only residents in U_p can newly become blocking residents. Since $|U_p| = |C_x| \leq |C| = m - t$, the number of blocking residents for M' is at most $a + (m - t)$.

Construct a solution $\mathcal{P}'$ of I_0 from M' such that $\mathcal{P}' = \{P_i \mid$ hospital p_i is empty in $M'\}$. Clearly, $|\mathcal{P}'| = t$. We show that the cost of $\mathcal{P}'$ is at most $(a+m-t)/B$. Partition U into B subsets $U_j = \{u_i^j \mid 1 \leq i \leq n\}$ $(1 \leq j \leq B)$. Then there is an integer j such that U_j contains at most $(a + m - t)/B$ blocking residents. If u_i^j is non-blocking, all the hospitals superior to x for u_i^j are assigned in M', and hence by the construction of $\mathcal{P}'$, no subset containing u_i is selected in $\mathcal{P}'$, i.e., the element u_i does not contribute to the cost of $\mathcal{P}'$. Hence, only elements u_i whose corresponding residents u_i^j are blocking can contribute to the cost of $\mathcal{P}'$. Therefore, the cost of $\mathcal{P}'$ is at most $(a+m-t)/B$. □

5 Concluding Remarks

An obvious future research is to obtain lower bounds on the approximation factor for Min-BR HRLQ (we even do not know its APX-hardness at this moment). Since the problem is harder than DkS, it should be a reasonable challenge.

As for Min-BP HRLQ, it is interesting to consider a decision variant, namely, the problem of asking whether an optimal solution contains at most k blocking pairs for a given integer k. In Theorem 1, we have shown that the problem of determining whether the optimal cost is at most n^δ or at least $n^{1-\delta}$ is NP-hard for any constant $\delta (> 0)$, where $n = |H| + |R|$. This implies that the decision problem is NP-hard if $k = O(n^\delta)$ for any δ. On the other hand, Theorem 3 implies that the problem is solvable in polynomial time when k is a constant. It is interesting to consider the complexity of the problem when k is between them, e.g., $k =$polylog(n). Another direction is to develop an FPT algorithm (parameterized by the optimal cost t) for Min-BP HRLQ, improving Theorem 3.

Finally, we remark on the possibility of generalization of instances: In this paper, we guarantee existence of feasible matchings by the CL-restriction (Sec. 2). However, even if we allow arbitrarily incomplete lists (and even ties), it is decidable in polynomial time if the given instance admits a feasible matching [12]. Thus, it might be interesting to seek approximate solutions for instances without the CL-restriction. Unfortunately, however, we can easily imply its $|R|^{1-\epsilon}$-approximation hardness in the following way.

Consider the problem of finding a maximum cardinality matching with the fewest blocking pairs, given a stable marriage instance with incomplete preference lists (call it *Min-BP SMI* for short). Its approximation hardness of $n^{1-\epsilon}$ for any positive constant ϵ is already known [7, 19], where n is the number of men in an input. The reduction given in [19], whose idea was taken from [7], constructs an instance of Min-BP SMI having a perfect matching and creates a large gap on the number of blocking pairs between "yes" instances and "no" instances. We can verify that this gap holds also for the number of men involved in blocking pairs. If we regard instances produced by this reduction as ones of Min-BR HRLQ, by considering men and women as residents and hospitals, respectively, and setting the quotas to $[1, 1]$ for all the hospitals, then we can show $|R|^{1-\epsilon}$-approximation hardness of 0-1 Min-BR HRLQ.

References

1. D. J. Abraham, P. Biró and D. F. Manlove, " "Almost stable" matchings in the roommates problem," *Proc. WAOA 2005*, LNCS 3879, pp. 1–14, 2005.
2. D. J. Abraham, R. W. Irving and D. F. Manlove, "Two algorithms for the Student-Project Allocation problem," *J. Discrete Algorithms*, Vol.5(1), pp. 73–90, 2007.
3. B. Aldershof and O. M. Carducci, "Stable matchings with couples," *Discrete Applied Mathematics*, Vol. 68, pp. 203–207, 1996.
4. Y. Asahiro, K. Iwama, H. Tamaki and T. Tokuyama, "Greedily finding a dense subgraph," *J. Algorithms*, Vol. 34(2), pp. 203–221, 2000.
5. A. Bhaskara, M. Charikar, E. Chlamtac, U. Feige, and A. Vijayaraghavan, "Detecting high log-densities – an $O(n^{1/4})$ approximation for densest k-subgraph," *Proc. STOC 2010*, pp. 201–210, 2010.
6. P. Biró, T. Fleiner, R. W. Irving, and D. F. Manlove, "The College Admissions problem with lower and common quotas," *Theoretical Computer Science*, Vol. 411(34-36), pp. 3136–3153, 2010.
7. P. Biró, D. F. Manlove, and S. Mittal, "Size versus stability in the marriage problem," *Theoretical Computer Science*, Vol. 411(16-18), pp. 1828–1841, 2010.
8. Canadian Resident Matching Service (CaRMS), `http://www.carms.ca/`
9. U. Feige, "Relations between average case complexity and approximation complexity," *Proc. STOC 2002*, pp. 534–543, 2002.
10. U. Feige, G. Kortsarz, and D. Peleg, "The dense k-subgraph problem," *Algorithmica*, Vol. 29, pp. 410–421, 2001.
11. T. Fleiner and N. Kamiyama, "A matroid approach to stable matchings with lower quotas," *Proc. SODA 2012*, pp. 135-142, 2012.
12. H. N. Gabow, "An efficient reduction technique for degree-constrained subgraph and bidirected network flow problems," *Proc. STOC 83*, pp. 448–456, 1983.
13. D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *Amer. Math. Monthly*, Vol. 69, pp. 9–15, 1962.
14. D. Gale and M. Sotomayor, "Some remarks on the stable matching problem," *Discrete Applied Mathematics*, Vol. 11, pp. 223–232, 1985.
15. M. R. Garey and D. S. Johnson, "Computers and Intractability: A Guide to the Theory of NP-Completeness," W. H. Freeman & Co., New York, NY, USA, 1979.
16. D. Gusfield and R. W. Irving, "The Stable Marriage Problem: Structure and Algorithms," MIT Press, Boston, MA, 1989.
17. M. M. Halldórsson, K. Iwama, S. Miyazaki and H. Yanagisawa, "Improved approximation results of the stable marriage problem," *ACM Transactions on Algorithms* 3 (3) Article No. 30, 2007.

18. K. Hamada, K. Iwama, and S. Miyazaki, "The hospitals/residents problem with quota lower bounds," *Proc. MATCH-UP* (satellite workshop of ICALP 2008), pp. 55–66, 2008.
19. K. Hamada, K. Iwama, and S. Miyazaki, "An improved approximation lower bound for finding almost stable maximum matchings," *Information Processing Letters*, Vol. 109(18), pp. 1036–1040, 2009.
20. K. Hamada, K. Iwama, and S. Miyazaki, "The hospitals/residents problem with quota lower bounds," *Proc. ESA 2011*, LNCS 6942, pp. 180–191, 2011.
21. C.-C. Huang, "Classified stable matching," *Proc. SODA 2010*, pp. 1235-1253, 2010.
22. R. W. Irving, D. F. Manlove and S. Scott, "The hospital/residents problem with ties," *Proc. SWAT 2000*, LNCS 1851, pp. 259–271, 2000.
23. R. W. Irving, D. F. Manlove and S. Scott, "The stable marriage problem with master preference lists," *Discrete Applied Math.*, Vol. 156(15), pp. 2959-2977, 2008.
24. S. Khot, "Ruling out PTAS for graph min-bisection, densest subgraph and bipartite clique," *Proc. FOCS 2004*, pp. 136–145, 2004.
25. S. Khuller, S. G. Mitchell and V. V. Vazirani, "On-Line algorithms for weighted bipartite matching and stable marriages," *Theoretical Computer Science*, Vol. 127(2), pp. 255–267, 1994.
26. E. J. McDermid and D. F. Manlove, "Keeping partners together: algorithmic results for the hospitals/residents problem with couples," *Journal of Combinatorial Optimization*, Vol. 19(3), pp. 279-303, 2010.
27. E. Ronn, "NP-complete stable matching problems," *J. Algorithms*, Vol. 11, pp. 285–304, 1990.
28. A. E. Roth, "The evolution of the labor market for medical interns and residents: a case study in game theory", *J. Political Economy*, Vol. 92(6), pp. 991–1016, 1984.
29. C. P. Teo, J. V. Sethuraman and W.P. Tan, "Gale-Shapley stable marriage problem revisited: Strategic issues and applications," *Management Science*, Vol. 47(9), pp. 1252–1267, 2001.
30. S. A. Vinterbo, "A stab at approximating minimum subadditive join," *Proc. WADS 2007*, LNCS 4619, pp. 214–225, 2007.