

Redesigning the Israeli Medical Internship Match

Noga Alon Tel-Aviv University	Slava Bronfman Bar-Ilan University
Avinatan Hassidim Bar-Ilan University	Assaf Romm Harvard University

March 7, 2015

Abstract

The final step in getting an Israeli M.D. is performing a year-long internship in one of the hospitals in Israel. Internships are decided upon by a lottery, which is known as “The Internship Lottery”. In 2014 we redesigned the lottery, replacing it with a more efficient one. This paper presents the market, the redesign process and the new mechanism which is now in use. There are two main lessons that we have learned from this market. The first is the “Do No Harm” principle, which states that (almost) all participants should prefer the new mechanism to the old one. The second is that new approaches need to be used when dealing with two-body problems in object assignment. We focus on the second lesson, and study two-body problems in the context of the assignment problem. We show that decomposing stochastic assignment matrices to deterministic allocations is NP-hard in the presence of couples, and present a polynomial time algorithm with the optimal worst case guarantee. We also study the performance of our algorithm on real-world and on simulated data.

1 Introduction

Prior to receiving their medical degrees Israeli medical graduates (and foreign trained doctors) must participate in a year-long internship program (not

to be confused with the residency period following their graduation). The internship is carried out in one of 23 possible hospitals in the country, which are allocated interns relative to the size of their patient population (with an advantage to peripheral hospitals), with the smallest allocation being of 4 interns. The internship is perceived as a tax that must be paid: there is (almost) no correlation between the hospital in which internship was performed to the hospital in which residency is performed, interns work nights and weekend to receive below minimum wage, and are not getting any tuition during the internship.¹

By and large, interns are not assigned to hospitals on the basis of merit, since the government wants to spread the more talented interns everywhere.² Since merit is not a criterion, it makes sense to use some form of lottery to assign the interns. For many years, the lottery which was chosen was Random Serial Dictatorship, with some house rules.

In this paper, we describe the market and the different populations of students, discuss the old lottery and present our new design for the lottery, which was used first in 2014 (and is also in use this year). We focus this paper on the lessons we learned, and on techniques that can be used in designing future markets. More specifically, this paper deals with

1. the “Do No Harm” principle, and the constraint that under some metric no student will be worse off in the new match,
2. the trade-off between efficiency and truthfulness, and
3. two-body problems in the context of the assignment problem.

In addition, we describe the market, the participants, and share the available data.

1.1 Background on the internship market

Each year, two different cohorts of students enter independent lotteries which determine where each student performs her internship. The number of interns allocated to each hospital is determined independently for each cohort.

¹One justification for this tax is that medical studies are highly subsidized by the government, with a tuition cost of around 2500\$ per year. In return, the government uses the interns as cheap labor in hospitals (all hospitals are operated by the government).

²There is an exception to this rule - five interns with a PhD get to choose where they want to be assigned.

The first cohort is composed of students who are in their final year of their medical school in Israel. This cohort is the more interesting and diverse one, and contains four populations:

1. a handful of students (usually with a PhD) who get to choose their internship. From a market design perspective, they can be treated as reducing the capacity.
2. Couples who wish to be in the same hospital. Unlike the (American) National Residency Matching Program (NRMP), there is no notion of preferences over pairs of hospitals. This makes sense: in the NRMP one spouse may want to be a radiologist, and the other may want to be a psychiatrist (so they need to rank pairs of programs), whereas in the current match everyone wants to (well, doesn't want to, but has to) be an intern. Around 10% of the interns are involved in a couple, and most couples are not married, but are just friends who want to share an apartment if they are stuck in some village where they do not know anybody. Unlike the NRMP, couples are *guaranteed* to be interns in the same place.
3. Students who have kids. This population is characterized by not wanting to transfer the family. Hence, they are usually willing to take any hospital which is a driving distance from home, and are less sensitive to hospital's quality.
4. The rest of the students.

In the last decade, the lottery that was used was Random Serial Dictatorship, with a few small modifications:

1. The five students who get to pick, choose where they want to go.
2. A couple is considered as one student for the matter of choosing the permutation. Alternatively, each couple gets two consecutive spots. If the couple's top choice among hospitals which have vacancies is A , but A has just one free spot, they are not allowed to go there and will keep going down their preferences.³

³It is not clear what happens if all hospitals have just one vacancy when a couple is chosen. Maybe it just did not happen in recent years. The medical students believed that couples will always stay together, and this was one of their requirements.

3. Every year there is a vote about what to do with students with kids (sometimes an advantage is given only to students with two or more kids). The options are to treat them as regular students, to let them take part in the lottery but promise them a seat in their district (the country is divided into five districts for that matter), or to promise them a seat in their district, without participation in the lottery (so they lose the chance to get good hospitals). When the last option is being used, most parents do not declare themselves parents, as they claim that this is not really a benefit.

The second cohort consists of graduates of foreign schools, who want to become Israeli doctors. This only applies to students who studied in certain countries and have no experience, and most of this cohort is composed of Israelis who were not accepted to medicine in Israel, and studied in Jordan or in Europe with the intention of going back to Israel after graduation.

For the foreign cohort, the internship lottery is run by the Ministry of Health (MoH). There are no votes or appeals, and rules are simpler. For the Israeli cohort, the MoH just decides on the capacities, and delegates the rest of the lottery process to a committee of students, elected by the student body. It is common that the committee puts important decisions (such as the parental benefits) to vote by the entire student body.

2 The redesign process

We were first approached in 2010 by several medical students who asked for help in redesigning their market. They had two things which worried them:

1. giving fair benefits to parents. They felt that parents should be treated differently, but it should not be at the expense of other students, and were not sure how to do this, and
2. they wanted to improve the efficiency of the system.

We took part in the internship committee of 2011, to understand their demands better. They added a couple of technical requirements:

1. *Do No Harm*: The redesign process should not hurt any population of students. Everyone should be (weakly) better off in the new design (at least in expectation).

2. In particular, students should have the option of matching together, just like they did up until then. A couple should get a guarantee that they will be matched together.

The 2011 class was also generous enough to share their data with us. Our initial idea was to use some form of Probabilistic Serial [5]. However, the results were very similar to RSD, and we wanted to improve efficiency. Hence, we decided to use a *rank efficient* mechanism, which essentially amounts to maximizing some linear function on the number of students who get their i 'th rank, for every i .

To cope with the DNH principle, we ran surveys, in which we gave students two possible vectors of probabilities (for being assigned to each hospital), and asked them which is better. We took the data, and tried to find a simple utility model for it. The model we ended up using is that there are m hospitals, and you have probability p to get to rank i , this gives you $p(m - i + 1)^2$ points of happiness, and a profile with more happiness points is better. While one could find a better fit to the model using more parameters, we were happy that there are no constants that we need to compute. We also defined a different happiness function for parents, but it was not used (see below).

Given that we have a definition of happiness, we first compute for every intern what would be her expected happiness under RSD. Then we choose allocation probabilities to maximize total happiness, conditioned on capacities and on the DNH principle – that every agent is happier under the current algorithm. Given the probability matrix, we decompose it to a convex sum of assignment matrices (more on that in section 3), and choose a matrix at random (according to the weights).

Switching to the new mechanism We proposed the new mechanism to the class of 2012, but were rejected. They did not want to have anything to do with computers, and were afraid of bugs and backdoors.

For the 2013 class we approached the MoH. We understood that the MoH is less conservative than the student body, as it has more to gain from new ideas: if the student body is given a new idea who would make it better off for with probability 0.4 and worse off with probability 0.6, then they should reject it. However, under these terms it is in the MoH's best interest to try the new idea. If the new idea is a failure than one class pays the price, but if it is a success then future generations will use it as well. In the MoH we

found moral support, more data, but no willingness to try the new system. Their main argument was that they do not want to interfere in something which is traditionally run by the students, and are not willing to use the foreign trained students as guinea pigs.

For the 2014 class we went back to the student internship committee. At this stage we had a large quantity of historical data to present about both algorithms, and also had a new idea: we will not run the lottery with a computer, but rather hand them the decomposition of the matrix into a weighted list of assignments, and let them choose the assignment. This way they can compute their own marginal distributions, and verify that there is no backdoor.⁴ This persuaded the committee to put this for a vote (along with the parental benefits for that year), with the condition that the new algorithm will be used only if it wins an absolute majority. 80% of the medical graduates who participated in the poll (55% of the students) voted in favor of the new approach. Following the successful vote, we indeed implemented the algorithm and it was used to assign interns on 2014. In addition, it was used for the foreign cohort that year. A subsequent vote of the following cohort of interns, that took place on January 2015, certified the continued use of our algorithm for 2015, and the foreign trained interns will keep using it until further notice. In order to increase transparency we explained the student body the rules of the new match [6].

We have ran a survey in the class of 2014, presenting students with their marginal distribution under RSD and under the new algorithm, and received supportive feedback. However, we were asked by the student committee to distribute the survey only to committee members, as we ran it after the lottery was over and they wanted to let the dead rest. This turned out to be very unfortunate, as the class of 2015 was mad at us for not handing each student in 2014 his or her marginal distribution. We have agreement from this year's committee to perform the survey with the entire body of students.

In the years 2014 and 2015, the students voted against giving any parental benefits. We think that the main source for this vote is not our algorithm, but a new school of medicine which accepts students who already earned an undergraduate degree (usually in Biology) and want to undergo retraining and perform a career change. The graduates of this school constitute the

⁴To be more exact, they can compute their marginal distributions to see that they are better off, and the distributions of the members of the committee to see that no one is cheating.

majority of parents in the market, and therefore it became an issue of “helping students in the other school” which attracts less sympathy than “helping the parents in my class”.

Truthfulness and efficiency Our main concern with the new mechanism is that we sacrifice truthfulness (at least to some degree) to achieve efficiency. This is probably a necessity: [13] showed that in large random market all asymptotically efficient, symmetric, and asymptotically strategy-proof mechanisms are equivalent to Random Serial Dictatorship. While in theory it could be that our market is not large enough, or that there is something special in the valuation profiles we face, it is unlikely that this is the case. Hence, any improvement in the welfare will lead to a non-truthful mechanism, such that some agents will be able to deviate and make a non-negligible gain.

When choosing between truthfulness and efficiency, we need to remember that truthfulness is a means, while efficiency is an end. We see two main arguments for truthful mechanisms:

- Truthful mechanisms reduce cognitive burden for the participants.
- When bidding is truthful, the designer can evaluate the welfare. When bidding is not truthful, such evaluation is not possible.

We think that for this market, one can mitigate the disadvantages of a non-truthful algorithm, and the gain in welfare is big enough to warrant some degree of untruthfulness. Specifically, we believe that bidding today is truthful (as much as anything can be truthful in ranking 25 priorities), and that if in the future a large body of students would bid untruthfully, then discussions in forums and on Facebook revolving around strategic bidding would emerge. We know that in the past, it was a common (and illegal) practice to sell and buy internship positions after the lottery.⁵ Negotiating these deals was done through social media, and it was not too difficult for us to find it. Discussing bidding strategies should happen through the same channels, and be less hidden (since there is nothing wrong about it).

⁵In theory, there should be no trade following a lottery if monetary transfers are not allowed. In practice, internship begins at least ten months (and sometime eighteen months) after the lottery is conducted, so it makes sense that someone who would want a desired hospital in city A would want to move to city B for some personal reason. Moreover, as mentioned, illegal monetary transfers were conducted. The students would approach the MoH, asking to trade places, hiding any financial agreement, and would be granted the permission to trade. The market price of a good internship used to be 2500\$.

3 Couples in the assignment problem

This section will deal with the students internship committee’s second condition, and how we were able to comply with it. In the past, any two students were allowed to declare that they are a “couple” and be matched together (that is, to the same hospital) by receiving only one joint turn in the Random Serial Dictatorship. The committee required us to give couples the same option under the new algorithm. Similar to several other algorithms for assignment of indivisible objects (e.g., [10, 5]), our algorithm “trades” probabilities among participants (based on their preferences), and reaches a stochastic matrix that has to be decomposed into a convex combination of valid assignments. The last step uses the famous Birkhoff-von Neumann decomposition process [4, 19]. The fact that couples cannot be split imposes (hard) complementarity constraints, which invalidates some of the assignments, and consequently the decomposition process may become harder or even impossible. On the bright side, only stochastic matrices in which both members of the couple get exactly the same probability for each hospital are candidates for decomposition. Is this little advantage enough to make the problem solvable? And if not, what could be done?

In this paper we first show that not only is the problem often not solvable when there are couples, but also that it is NP-hard to determine whether a given stochastic matrix with couples can be decomposed (Theorem 1). This result extends trivially to environments with more general complementarities between groups of participants.

Our second main contribution is in bypassing this impossibility result by providing a polynomial time approximation algorithm that outputs a convex combination of valid assignments that is similar to the target stochastic matrix. We show that the approximation is tight and behaves like $2/\underline{q}$, where \underline{q} is the capacity of the smallest hospital participating in the match. In the last section of the paper we consider several extensions of the algorithm that can be useful for other applications. We also use anonymized preferences data provided by the MoH to subsample and test our algorithm performance. We find that our algorithm often performs significantly better on actual data (compared to the theoretical bounds we establish).

3.1 Our Results

1. It is NP-hard to determine whether a stochastic matrix with couples can be decomposed into a convex combination of deterministic assignments with couples.
2. It is possible to approximate such a decomposition in polynomial time. The approximation is tight and its quality is reciprocal to the smallest hospital' capacity.
3. Subsampling real preferences data from the Israeli Medical Internship Match for which the approximation was developed shows excellent performance of the approximation algorithm.

3.2 Related Work

The assignment problem with couples is a specific case of multi-unit allocation of indivisible objects without transfers (for example, course allocation), and we indeed intend for our treatment to convey a message regarding the more general case. [7] provides a deterministic algorithm to find an *ex-post* efficient allocation that represent an approximately competitive equilibrium from approximately equal incomes. [8] suggest a decomposition in the spirit of Birkhoff and von-Neumann, but complementarities are not allowed. A contemporary working paper by [14] is the closest to ours, and it solves the problem of allocating bundles of indivisible objects by assuming (like we do) that bundles are limited in size, and that the capacity constraints are “soft”. While we insist on capacities being met, allowing deviations as those of Nguyen et al. and then correcting them in a smart way will give a result similar to our approximation result (Theorem 3, see also Section 5). [2] suggest a different kind of approximation, by ignoring the couples constraint with a small probability.

The motivation for decomposing stochastic assignment matrices comes from mechanisms that efficiently allocate probabilities in the interim stage. Notable contributions for the case of single-item assignments are [10] who proposes a competitive equilibrium from equal incomes approach, [5] who suggest the Probabilistic Serial mechanism which is ordinally efficient (i.e., interim efficient given ordinal preferences and not cardinal preferences), and [9] who studies rank-efficient mechanisms (similar in spirit to the one implemented by us for the Israeli Medical Internship Match). We note that

several authors have dealt with the case of object assignment when objects are the endowment of some agents, in which case the top-trading cycles (or a variation thereof) is often the mechanism of choice (see, e.g., [17, 1]).

Finally, the affect of couples in matching problems has also been studied in the context of two-sided matching. See, for example, the works by [12] and [3]. A treatment more related to our current approach is provided by [15].

3.3 Model and Notation

An assignment problem with couples is a tuple (S, C, H, M) . S is a finite set of single interns, C a finite set of couples of interns, with each $c \in C$ being a set of two interns $c = \{c_1, c_2\}$. We denote by $I = S \cup (\bigcup_{c \in C} c)$ the set of all interns. H is a finite set of hospitals. $M \in [0, 1]^{I \times H}$ is *the target matrix* and it satisfies

1. $\forall i \in I : \sum_{h \in H} M_{i,h} = 1$, and
2. $\forall (c_1, c_2) \in C, h \in H : M_{c_1,h} = M_{c_2,h}$.

We let $q_h = \sum_{i \in I} M_{i,h}$ be the capacity of hospital h , and $\underline{q} = \min_h q_h$ be the capacity of the smallest hospital. Let \mathcal{P} be the domain of all assignment problems with couples. We specify two special sub-domains of \mathcal{P} : $\mathcal{P}^{C=\emptyset}$ is the set of problems without couples (i.e., $C = \emptyset$), and $\mathcal{P}^{S \geq C}$ is the set of problems in which singles receive more weight than couples in *every* hospital (i.e., $\forall h \in H : \sum_{s \in S} M_{s,h} \geq 2 \sum_{c \in C} M_{c_1,h}$).

A matrix $M' \in [0, 1]^{I \times H}$ that satisfies conditions (1) and (2), and $\forall h \in H : \sum_{i \in I} M'_{i,h} = q_h$ is called *a stochastic assignment matrix (with respect to P)*. If all the elements of a stochastic assignment matrix (with respect to P) are in $\{0, 1\}$, then the matrix is called *a deterministic assignment matrix (with respect to P)*. A stochastic assignment matrix M' is *decomposable* if it can be represented as a convex combination of deterministic assignment matrices, i.e., $\{(\lambda^k, M^k)\}_{k=1}^K$ such that for each k , $\lambda^k \in (0, 1)$ and M^k is a deterministic assignment matrix, $\sum_{k=1}^K \lambda^k = 1$, and $M' = \sum_{k=1}^K \lambda^k M^k$. A straightforward extension of the Birkhoff-von Neumann theorem shows that on the sub-domain $\mathcal{P}^{C=\emptyset}$ all target matrices are decomposable. In Section 3.4 we show that on the general domain, \mathcal{P} , it is NP-hard to verify whether a target matrix is decomposable.

We say that a convex combination of deterministic assignment matrices $\{(\lambda^k, M^k)\}_{k=1}^K$ ϵ -approximates a target matrix M if

$$\left\| \left| M - \sum_{k=1}^K \lambda^k M^k \right| \cdot \vec{\mathbf{1}}_{[I]} \right\|_{\infty} < \epsilon,$$

where $\vec{\mathbf{1}}_{[I]}$ is a vector of ones of size I , so we approximate the maximum value over the L_1 norm of each row.

A decomposition algorithm \mathcal{A} takes a problem $P \in \mathcal{P}$ and outputs a convex combination of deterministic assignment matrices (with respect to P). We say that \mathcal{A} provides an f -approximation on domain \mathcal{P}' if for every $P \in \mathcal{P}'$, $\mathcal{A}(P)$ $f(P)$ -approximates M . Our aim in Section 3.5 is to provide a lower bound for the optimal approximation on \mathcal{P} , and an upper bound for the optimal approximation on $\mathcal{P}^{S \geq C}$.

3.4 NP-hardness result

Theorem 1. *On the domain \mathcal{P} , determining whether the target matrix is decomposable is in NPC.*

Proof. We will reduce 3-edge coloring of cubic graphs to the interns-hospitals assignment problem, ([11], proved that 3-edge coloring of cubic graphs is in NPC).

Let $G = (V, E)$ be a cubic graph with $|V| = n$ vertices and $m = 3n/2$ edges. For each edge e let us have 3 hospitals $A(e), B(e), C(e)$, each of capacity 2. For each vertex v and each color $i \in \{1, 2, 3\}$ let us have a hospital (v, i) with capacity 1. Thus altogether we have $3m$ hospitals of capacity 2 and $3n$ hospitals of capacity 1, total hospital capacity is $3n + 6m = 3n + 6 \cdot 3n/2 = 12n$.

Now for each edge $e = (u, v)$ we have one couple that wants to be either in $A(e)$ or in $B(e)$ or in $C(e)$, each with probability $1/3$ and we also have 6 single interns as follows:

- First single wants either $A(e)$ with probability $2/3$ or $(u, 1)$ with probability $1/3$.
- Second wants either $A(e)$ with probability $2/3$ or $(v, 1)$ with probability $1/3$.

- Third wants either $B(e)$ with probability $2/3$ or $(u, 2)$ with probability $1/3$.
- Forth wants either $B(e)$ with probability $2/3$ or $(v, 2)$ with probability $1/3$.
- Fifth wants either $C(e)$ with probability $2/3$ or $(u, 3)$ with probability $1/3$.
- Sixth wants either $C(e)$ with probability $2/3$ or $(v, 3)$ with probability $1/3$.

Note that altogether we have $(2 + 6)m = 8 \cdot 3n/2 = 12n$ interns. Hence, any assignment of all of them uses all capacity of all hospitals.

Suppose first that G is class 1 (that is, it is 3-edge colorable). Fix a proper 3-edge coloring by colors 1, 2, 3. Each such coloring corresponds to a full assignment of all interns as follows:

If edge $e = (u, v)$ is colored 1 then the couple is in $A(e)$, First single intern is in $(u, 1)$, second in $(v, 1)$, third and forth in $B(e)$, fifth and sixth in $C(e)$. If e is colored 2 and 3 we proceed symmetrically in the obvious way. It's clear that this is a full assignment.

Now give this assignment weight $1/3$, and give weight $1/3$ to each of the two other assignments obtained from it by cyclically shifting the colors of all edges. This gives a decomposition of our matrix of probabilities.

Conversely, if there is a decomposition, then any single assignment in its support must be a full assignment (assigning all interns and saturating all hospitals). But this means that for each edge e , among the 8 interns corresponding to this edge, 6 including the couple are assigned to the hospitals $A(e), B(e), C(e)$ and there is a unique $i \in \{1, 2, 3\}$ so that two of the single interns are assigned to (u, i) and (v, i) and thus G is 3-edge colorable, as needed. \square

3.5 Approximation

3.5.1 Lower bound

Theorem 2. *If \mathcal{A} provides an f -approximation on $\mathcal{P}^{S \geq C}$, then for any $n \in \mathbb{N}$ there exists $P \in \mathcal{P}^{S \geq C}$ such that $|I| \geq n$ and $f(P) \geq \frac{2}{q+2}$.*

Proof. Given $n \in \mathbb{N}$, let $q = 4\lceil \frac{n}{8} \rceil$. Let $S = \{s_1, \dots, s_{\frac{1}{2}q+1}, s'_1, \dots, s'_{\frac{1}{2}q+1}\}$, $C = \{c_*, c_1, \dots, c_{\frac{1}{4}q-1}, c'_1, \dots, c'_{\frac{1}{4}q-1}\}$, and $H = \{h, h'\}$. Consider the target matrix $M \in \mathcal{P}^{S \geq C}$ described in Table 1. Under this stochastic assignment matrix each single intern in $\{s_1, \dots, s_{\frac{1}{2}q+1}\}$ and each couple in $\{c_1, \dots, c_{\frac{1}{4}q-1}\}$ gets a probability 1 to be assigned to h , each single intern in $\{s'_1, \dots, s'_{\frac{1}{2}q+1}\}$ and each couple in $\{c'_1, \dots, c'_{\frac{1}{4}q-1}\}$ gets a probability of 1 to be assigned to h' , and the couple c_* gets an equal probability to be assigned to either h or h' . Note that the capacity of both hospitals is q .

Table 1: Target matrix for proof of Theorem 2

	h	h'
s_1	1	0
\vdots	1	0
$s_{\frac{1}{2}q+1}$	1	0
s'_1	0	1
\vdots	0	1
$s'_{\frac{1}{2}q+1}$	0	1
$c_{*,1}$	0.5	0.5
$c_{*,2}$	0.5	0.5
$c_{1,1}$	1	0
$c_{1,2}$	1	0
\vdots	1	0
$c_{\frac{1}{4}q-1,1}$	1	0
$c_{\frac{1}{4}q-1,2}$	1	0
$c'_{1,1}$	0	1
$c'_{1,2}$	0	1
\vdots	0	1
$c'_{\frac{1}{4}q-1,1}$	0	1
$c'_{\frac{1}{4}q-1,2}$	0	1

In any convex combination that approximates M , one of the hospitals will be assigned at least $\frac{1}{4}q$ of the couples with a weight of at least $\frac{1}{2}$. This

means that with probability $\frac{1}{2}$ the left over capacity for single interns is at most $\frac{1}{2}q$, where as the single interns require $\frac{1}{2}q + 1$. The optimal way to minimize the deviation from the singles probability would be to divide the deviation equally among all the relevant single interns. This means that each individual intern in the relevant hospital will get at most $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{\frac{1}{2}q}{\frac{1}{2}q+1}$ instead of getting 1, and to complete the individual's total probability to 1, the single will get $1 - (\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{\frac{1}{2}q}{\frac{1}{2}q+1})$ to the other hospital (instead of 0). Therefore the approximation cannot be better than:

$$2 \left(1 - \frac{1}{2} \cdot 1 - \frac{1}{2} \cdot \frac{\frac{1}{2}q}{\frac{1}{2}q+1} \right) = \frac{2}{q+2} \left(= \frac{2}{\underline{q}+2} \right)$$

□

One may argue that the example provided in the proof for Theorem 2 seems quite pathological. Indeed, the target matrix allocates many interns either h or h' with probability 1. There could be scenarios in which it would be very reasonable to consider only target matrices in which each intern's probability of reaching any particular hospital is bounded by some expression related to the number of hospitals (see also the concluding discussion). Nevertheless, in our main application the chosen algorithm involves trading probability between interns, and so it often outputs target matrices that are close to being pathological in exactly the same sense. This motivates our approximation metric and the consideration of extreme cases. More importantly, a similar bound can be achieved even if probabilities are restricted to being small, although the problems are not in the domain $\mathcal{P}^{S \geq C}$ (see Appendix A). Finding a lower bound in the domain $\mathcal{P}^{S \geq C}$ when probabilities are constrained to be "small" remains an open problem.

3.5.2 Suggested algorithm for upper bound

We present an approximation algorithm for decomposing any matrix $M \in \mathcal{P}$. The algorithm can be roughly divided into two main stages: the first stage deals only with couples and assigns them to hospitals according to their probability in M , and the second stage 'fixes' the probabilities of the single interns in M to match the assignment of the couples from the first stage. In the second stage the over-demanded capacity taken by the couples in any specific hospital is deducted from singles' demand according to a division which preserves the weight of each single intern in the singles' demand.

The algorithm starts by “enlarging” each hospital capacity to the sum of the couples probabilities to be assigned to h divided by two and rounded up to the nearest integer. The couples now behave like singles, and for each hospital we add a new single to take the “extra” capacity (the new single’s demand is completed by a dummy hospital, h_\emptyset). Doing so for all hospitals gives us a new stochastic assignment matrix containing only “singles” (with each single representing an original couple), and it can be decomposed using the Birkhoff-von Neumann theorem, to get a convex combination of deterministic assignment matrices that represent the allocation of the couples. In some of these assignment the couples take slightly more capacity than their expected share (but only up to the nearest multiple of two).

In the second stage of the algorithm, we look at the residuals of the first-stage assignment. For each couples assignment and for each hospital we check whether the couples exceeded their expected share in that hospital, and if so, we let all the singles reduce their demand by their respective weights. The missing demand is directed at a dummy hospital, h_\emptyset . We again have a stochastic assignment matrix containing only singles that can be decomposed using the Birkhoff-von Neumann theorem.⁶ The output of the decomposition assigns some of the singles to the dummy hospital, so we move them to vacant positions arbitrarily. The algorithm then combines back the first-stage and the second-stage assignments to get a valid deterministic assignment matrix with respect to the original problem.

The full algorithm’s pseudo-code can be found in Algorithm 1.

Theorem 3. *Algorithm 1 provides an \bar{f} -approximation on $\mathcal{P}^{S \geq C}$, where $\bar{f}(P) = 2/\underline{q}$.*

Lemma 4. *Algorithm 1 runs in polynomial time.*

Proof. The size of the convex combination $\{(\lambda^k, M^k)\}_{k=1}^K$ related to the couples is polynomial, also we use Birkhoff’s proof which provides a constructive algorithm for the implementation of the decomposition [16] in polynomial time, so the whole algorithm is polynomial. \square

of theorem 3. In the first part of the algorithm (lines 1 through 10) the couples are allocated exactly their demanded capacity. This is done by taking

⁶Strictly speaking, the demand needs to be rounded up to the nearest integer, and this can be done again by adding dummy interns for each hospital, and putting all the leftover demand in the dummy hospital.

Algorithm 1 Approximation

Input: A stochastic assignment matrix $M^{I \times H} \in \mathcal{P}$

- 1: $S' = C \cup H$, $H' = H \cup \{h_\emptyset\}$
 - 2: Create a new matrix $M' \in [0, 1]^{|S'| \times |H'|}$ (initialize with zeros)
 - 3: **for all** $h \in H$ **do**
 - 4: **for all** $c \in C$ **do**
 - 5: $M'_{c,h} = M_{c_1,h}$
 - 6: **end for**
 - 7: $M'_{h,h} = \lceil \sum_{c \in C} M'_{c,h} \rceil - \sum_{c \in C} M'_{c,h}$
 - 8: $M'_{h,h_\emptyset} = 1 - M'_{h,h}$
 - 9: **end for** ▷ We get that $(S', \emptyset, H', M') = P' \in \mathcal{P}^{C=\emptyset}$
 - 10: Decompose M' into a convex combination of deterministic assignment matrices $\{(\lambda^k, M^k)\}_{k=1}^K$ (with respect to P')
 - 11: Create an empty set of allocations $\psi = \{\}$
 - 12: **for** $k = 1$ to K **do**
 - 13: $\tilde{H} = H \cup \{h_\emptyset\}$
 - 14: Create a new matrix $\tilde{M} \in [0, 1]^{|S| \times |\tilde{H}|}$ (initialize with zeros)
 - 15: **for all** $s \in S$ **do**
 - 16: **for all** $h \in H$ **do**
 - 17: **if** $\sum_{c \in C} M_{c,h}^k > \sum_{c \in C} M'_{c,h}$ **then** ▷ If couples exceeded their quota
 - 18: $\tilde{M}_{s,h} = M_{s,h} - \frac{M_{s,h}}{\sum_{s' \in S} M_{s',h}} \cdot 2 (\sum_{c \in C} M_{c,h}^k - \sum_{c \in C} M'_{c,h})$
 - 19: **else**
 - 20: $\tilde{M}_{s,h} = M_{s,h}$
 - 21: **end if**
 - 22: **end for**
 - 23: $\tilde{M}_{s,h_\emptyset} = \sum_{h \in H} M_{s,h} - \sum_{h \in H} \tilde{M}_{s,h}$
 - 24: **end for** ▷ We get that $(S, \emptyset, \tilde{H}, \tilde{M}) = \tilde{P} \in \mathcal{P}^{C=\emptyset}$
 - 25: Decompose \tilde{M} into a convex combination of deterministic assignment matrices $\{(\hat{\lambda}^l, \hat{M}^l)\}_{l=1}^L$ (with respect to \tilde{P})
 - 26: **for** $l = 1$ to L **do**
 - 27: Stitch M^k and \hat{M}^l into a valid deterministic assignment matrix $M^{k,l}$ for P . Couples get the hospitals they are assigned under M^k , singles not assigned to h_\emptyset get the hospitals they are assigned under \hat{M}^l , and singles assigned to h_\emptyset get the rest of the vacant positions (arbitrarily)
 - 28: Add $(\lambda^k \cdot \hat{\lambda}^l, M^{k,l})$ to ψ
 - 29: **end for**
 - 30: **end for** 16
 - 31: Output ψ
-

only the couples, treating them as single agents, enlarging the capacities related to the couples so that they would be an integer, and add a dummy hospital to take all the left-over probabilities of the dummy agents. Then it is possible to use the Birkhoff-von Neumann decomposition.

We note that since each hospital has only one related dummy agent, then the capacity taken by the couples is always either the full (rounded up) capacity of the hospital according to M' , or one less than that. If it is the first case, the singles can be allocated their entire demand for this hospital in the next step. However, if the couples exceeded their fractional capacity, we “fix” the demand of the single interns, and each intern i loses $2(\lceil q'_h \rceil - q'_h)$ times his weight in the sum of probabilities of singles, where $q'_h = \sum_{c \in C} M_{c_1, h}$.

Let y_h be the total weight given the matrices of the second kind (for a certain hospital h). Then the following holds:

$$(1 - y) \cdot \lfloor q' \rfloor + y \cdot \lceil q'_h \rceil = q'_h \quad (1)$$

If we denote by $x_h = q'_h - \lfloor q'_h \rfloor$ the fractional part of the couples’ total probability we get $y_h = x_h/2$. Now each time the second case materializes, the singles lose $2 - x_h$. The total loss is given by $(2 - x_h) \cdot \frac{x_h}{2}$, which is maximized at $x_h = 1$. Since we assumed that each hospital contains more singles interns than interns who are part of a couple, each single s losses at most $M_{s, h} \cdot \frac{1}{2} / \frac{q_h}{2} = M_{s, h} / q_h$.

Across all hospitals, each single intern s losses at this point at most $\sum_{h \in H} M_{s, h} / q_h \leq (\sum_{h \in H} M_{s, h}) / \underline{q} = 1 / \underline{q}$. However, in the last phase of the algorithm (line 27), singles receive back their “lost probability” in arbitrary hospitals, and so each single can deviate up to $1 / \underline{q}$ further from her endowment, resulting in a total of at most $2 / \underline{q}$. \square

4 Experimental results

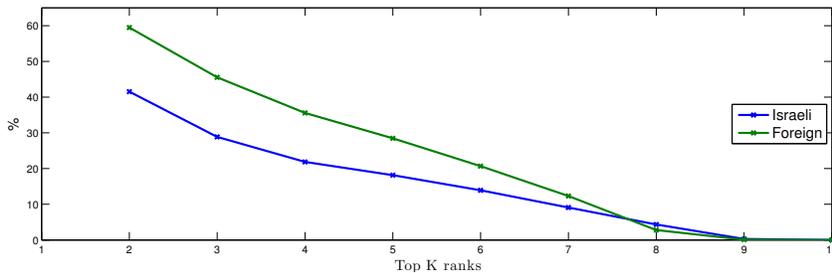
In this section we characterize the interns’ preferences and provide simulation results using the Israeli Medical Internship Match data provided to us by the Israeli MoH. The data contained the preferences of interns starting from 1995. However, due to significant changes in the hospitals and the internship conditions in them, some of the earlier data cannot be treated as coming from a similar distribution to the more recent data. We therefore focused only on preferences from the years 2010-2014. During those years there were 23 possible hospitals for each intern to rate, and interns *must* rank all

hospitals. The number of interns varied from one year to another, but it is always around 500, and the number of internship positions is always equal to the number of interns. Preferences of foreign trained doctors are also available to us, but only until the year 2012.

4.1 Description of interns’ preferences

Looking at the interns’ preferences, there are several interesting points to be made. The first is that preferences are geographically driven. For example, ranking the hospitals according to the number of students who listed this hospital as their top priority (this is the ranking published by the MoH), we see that in the top five hospitals there are two hospitals from the center, one from the north, one from the south, and one from the Jerusalem area (these are four of the districts in Israel). However, for any given student, such a geographically dispersed ranking is very unlikely. Indeed, except for Ichilov hospital (a hospital in the center of Israel which represents the top option for most of the population), it is common that all the top choices of a student come from the same geographic area.⁷ To measure this effect, Figure 1 shows the percentage of interns whose top k choices all came from the same area, for $k = 2$ to 10 (this graph ignores Ichilov hospital). Note that if choices were random, one would expect an exponential decay, where here we see a linear decay.

Figure 1: Geographical orientation at the top of the list



Another interesting point is the differences between the Israeli interns (interns who studied medicine in Israel) and the foreign trained doctors. For-

⁷It is known that one can always swap Ichilov hospital later for something else. Hence some students pick it first, because they want the flexibility – the lottery is conducted between ten and eighteen months before the internship starts.

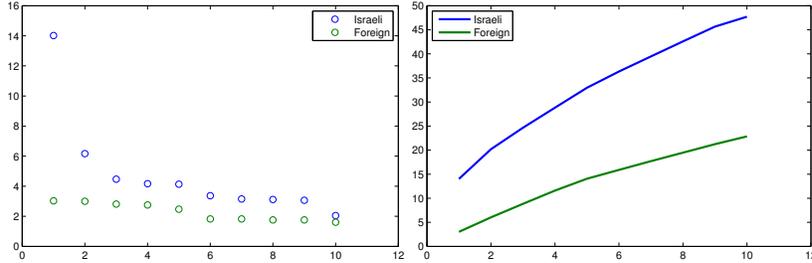
eign trained graduates have even stronger geographic preferences than Israeli ones. This helps explain another difference between these two populations' preferences: their performance in the old RSD mechanism. Under that mechanism the average rank of hospital that the Israeli interns received was 4.594, whereas the foreign trained doctors (who are matched separately) received an average rank of 2.549. Similarly, under the new mechanism the average rank of Israeli interns was improved to 3.686, and the foreign trained doctors got 2.042. Our hypothesis was that the foreign trained doctors' preferences therefore exhibit more heterogeneity in some sense.⁸

To verify that foreign trained doctors' preferences are more heterogeneous, we grouped preferences according to the top three hospitals, sorted by frequency, and plotted the percentage of students who chose the most common triplet, one of the two most common triplets, and so on. For example, if the most common triplet for the foreign trained doctors is A, B, C , and the second most common is B, C, A , we are interested in the fraction of students whose top three choices are A, B, C , and at the fraction of students whose top three choices are either A, B, C or B, C, A . The results are presented in Figure 2. The right panel shows that the cumulative distribution of local interns is much higher than the cumulative distribution of the foreign trained doctors. The top ten triplets cover almost 50% of the Israeli interns, but only about 20% of the foreign trained doctors. Amazingly, even the density of each of the top 10 triplets (which are not necessarily the same among the two populations) is higher for the local interns (as depicted by the left panel). This indicates that their preferences are indeed much more homogeneous, at least at the top of the distribution. We note that Figure 1 suggests that foreign trained doctors do exhibit a higher tendency to group hospitals by area. The only reasonable conclusion is that while foreign trained doctors care a lot about geography, there are less prone to select necessarily the center of Israel as their preferred location, or that they tend to have more heterogeneity in ranking hospitals within each area.

We also tried to study the differences between the preferences of singles and couples, but did not reach any significant conclusions. Furthermore, while we did not perform a thorough analysis, it is likely to assume that preferences depend also on the medical school in which an intern completed her medical studies, for both personal reasons (already lives in the same city)

⁸The common conspiracy theory is that the MoH gives them more seats at the preferable hospitals. We have verified the numbers and the conspiracy theory is incorrect.

Figure 2: Distribution of top triplets (left panel: density, right panel: cumulative)



and professional reasons (knows the hospital better from her time at medical school). This by itself might not have an impact on designing a mechanism, but it may be of interest for understanding better the distribution of preferences and what creates heterogeneity.

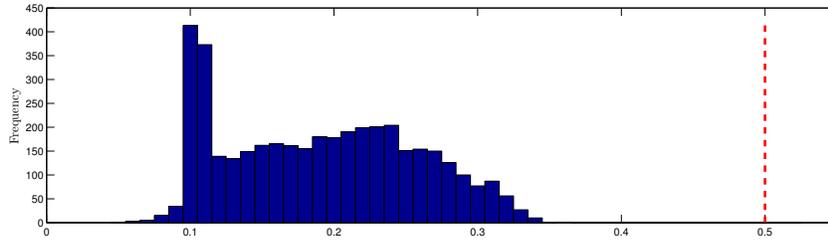
4.2 Simulation results for the algorithm

In the simulations we used a subsampling method, i.e., we drew preference profiles from the union of all data points, where we distinguish between preferences that were submitted by single interns and preferences that were submitted by couples. This allowed us to create multiple “possible markets”, each with $|I| = 496$ (which was the actual number of interns in 2014) and with 24 couples. For each point in the figure we used 5,000 different market draws.

To create Figures 3 and 4, we drew markets (in the sense explained above), then used our assignment mechanism to generate the matrix M , to which Algorithm 1 was applied. While Theorem 3 only ensures approximation quality of $2/\underline{q}$, we can clearly see in Figure 3 that although in our case $\underline{q} = 4$, the average performance of our algorithm is far better than $\frac{1}{2}$ (the red vertical line). One reason for the difference between the theoretical prediction and the data is that our analysis assumes that the percentage of singles’ weights is equal across all hospitals (see also the concluding discussion). The second and much more important reason is that even after applying the probability allocation mechanism, pathological examples similar to the one presented in Theorem 2 are relatively rare. A third reason is that the arbitrary way in which singles were allocated at the end of the algorithm may in fact improve

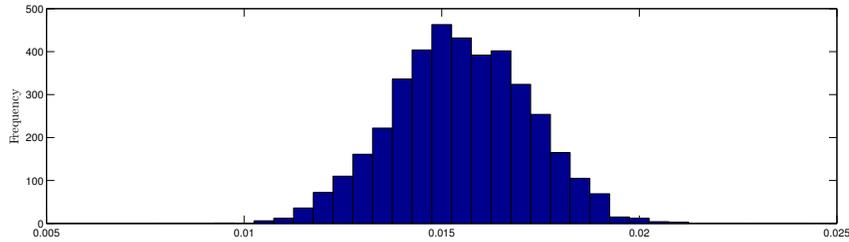
the performance.⁹ We note that the spike at the left side of the distribution occurs exactly because of a small hospital to which couples are likely to be allocated because of their (non-random) preferences (indeed, this spike disappears completely in Figure 5 which uses random allocation of probabilities).

Figure 3: The histogram shows the distribution of the maximum value over the L_1 norm of each intern in the absolute distance between the original assignment matrix and the approximated one. The red dashed vertical line represents the theoretical upper bound.



For completeness, Figure 4 depicts the distribution of $\frac{|M - \sum_{k=1}^K \lambda^k M^k| \cdot \vec{1}_{[I]}}{|I|}$, which is the average L_1 norm. This metric was not analyzed theoretically, but it is of much interest to the social planner. It roughly means that while the intern who got her probability vector changed the most suffered an average of at about 15% change, the average intern's probability vector was only changed by less than 2%.

Figure 4: The histogram shows the distribution of the average value over the L_1 norm of each intern in the absolute distance between the original assignment matrix and the approximated one.



⁹One may consider replacing the arbitrary assignment with a more sophisticated method to slightly improve the results. We did not pursue this direction any further.

In order to separate the effect of the LP mechanism used to generate probability matrices, we also produce similar figures for matrices produced randomly using the iterative algorithm of [18] (but only select those within our sub-domain of interest, $\mathcal{P}^{S \geq C}$), with the same market size as used for Figures 3 and 4. Figure 5 indeed reveals that using random matrices reduces the maximal impact by a factor of about 5. However, it is evident from Figure 6 that the average intern’s probability vector is now more susceptible to changes following the application of our approximation algorithm.

Figure 5: The histogram shows the distribution of the maximum value over the L_1 norm of each intern in the absolute distance between the original assignment matrix and the approximated one, for randomly generated values.

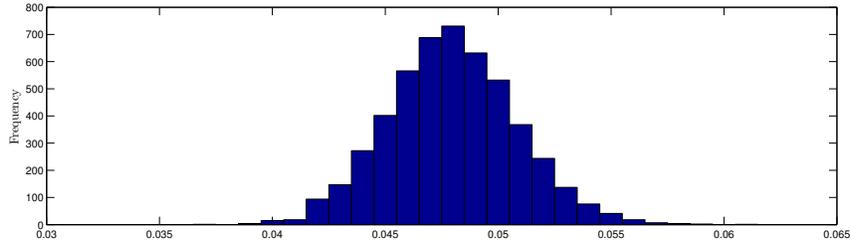
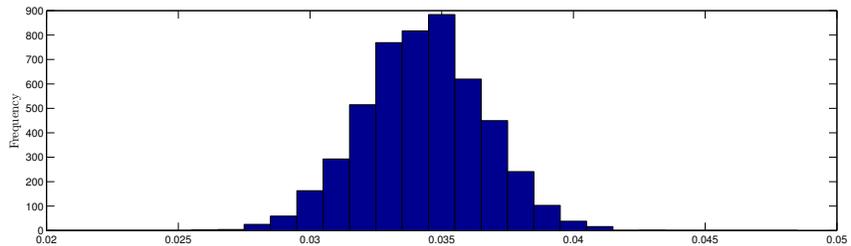


Figure 6: The histogram shows the distribution of the average value over the L_1 norm of each intern in the absolute distance between the original assignment matrix and the approximated one, for randomly generated values.



Finally, we also tried to see what would have happened to the performance of our algorithm if couples’ preferences were more capacity-driven compared to their true distribution. For this test, we used the same distribution for singles’ preferences, but each couples’ preference was created by drawing hospitals one by one, where every time the draw is from those hospitals not

drawn yet, and with the capacity of the hospital being the weight in the random draw. This made couples like big hospitals better than they like small hospitals, and thus the minimal weight of singles in all hospitals went up, and our algorithm’s performance was improved (see Figures 7 and 8).

Figure 7: The histogram shows the distribution of the maximum value over the L_1 norm of each intern in the absolute distance between the original assignment matrix and the approximated one, for capacity-driven couples.

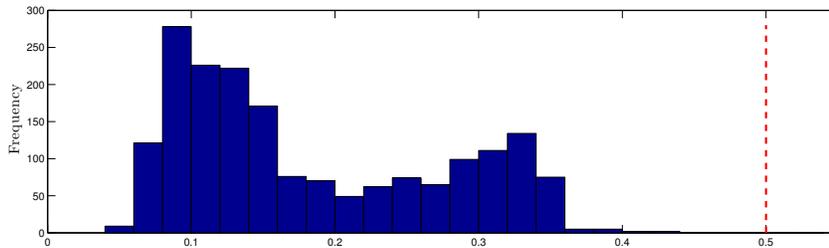


Figure 8: The histogram shows the distribution of the average value over the L_1 norm of each intern in the absolute distance between the original assignment matrix and the approximated one, for capacity-driven couples.

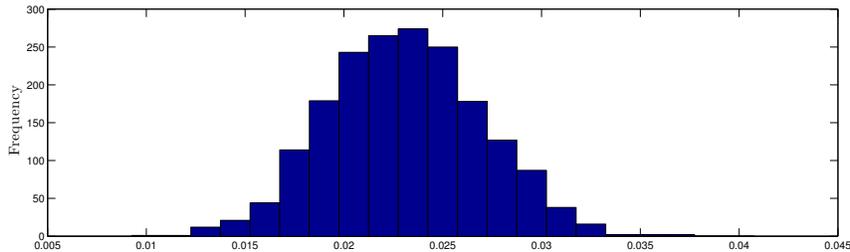


Figure 7 presents a funny double-peaked distribution. The reason for this is not lack of experiments, but rather whether or not a couple has a chance of getting to a hospital with capacity 4. Since the LP is an affine maximizer, if a couple has a chance to get to this small hospital, then this probability will be non negligible (in the order of at least 0.3 or so). If the couple has a chance to get there then rounding the probabilities of the intern there will give the intern with the highest discrepancy in the decomposition. On the other hand, if no couple ever gets to that hospital the decomposition there

is always perfect, and we need to worry about the second smallest hospital (which already has twice the number of interns than the smallest hospital).

5 Extensions

5.1 Lower percentage of singles in each hospital

For the sake of simplicity we focused on the domain $\mathcal{P}^{S \geq C}$ in which there are more single interns than interns who are part of a couple in each hospital. It is easy to generalize our results to the domain of $\mathcal{P}^{S \geq \alpha C}$, in which in every hospital the total weight allocated to single interns is at least α times the total weight allocated to couples ($\forall h \in H : \sum_{s \in S} M_{s,h} \geq \alpha \cdot 2 \sum_{c \in C} M_{c_1,h}$), for some positive α .

The example given in Theorem 2 can be modified to have $\frac{2\alpha}{1+\alpha}q + 2$ singles and $\frac{1}{1+\alpha}q - 1$ couples, and the resulting bound will be $\frac{1}{\left(\frac{\alpha}{1+\alpha}\right)^{\underline{q}+1}}$. The approximation algorithm stays exactly the same, and it provides an approximation of $\frac{1}{\left(\frac{\alpha}{1+\alpha}\right)^{\underline{q}}}$.

On a similar note, we remark that presenting the approximation as depending on the minimal capacity was a matter of choice. A slightly more accurate bound can be formulated in terms of the minimal singles' demand across hospital, i.e., $\min_{h \in H} \sum_{s \in S} M_{s,h} \left(\geq \frac{\alpha}{1+\alpha}q\right)$. This bound works much better when for some reason couples focus their demand on larger hospitals.

5.2 Groups larger than two

In the Israeli Medical Internship Match, interns were only allowed to register either as singles or as couples. However, other applications may require larger groups to be assigned together. To extend our solution to larger groups it is possible to replace the first stage of the algorithm (the one which rounds up the capacities demanded by the couples and allocates the couples) by the allocation method of [14] (without the singles). Their algorithm ensures an allocation that deviates from the initial allocation by at most $k - 1$ seats in each hospital, where k is the size of the largest group. We can then take the remaining capacities and split them among the singles in a similar manner, and the analysis of the upper bound will remain similar (the coefficient however will change according to k). Obviously, our two extensions can be combined.

6 Conclusion

In this paper we described a recent application of knowledge and research in market design to the problem of allocating interns to internship positions in Israel. Our data exhibited several very interesting characteristics that lead us, for example, to recommend (or at least suggest for consideration) merging the two cohorts (of local interns and foreign trained doctors) and assigning them using the same mechanism, since there are likely significant gains to trade. A naïve attempt that we made in putting combining these two populations resulted in an improvement for both populations under the new mechanism, and an improvement only for the local interns under RSD (with the foreign trained doctors receiving a worse expected rank). However, it is possible that the MoH will decide that the DNH principle should apply here as well and that the baseline that should be taken is not RSD on the merged population, but rather two independent runs of RSD (one for each population) and running the linear program on top of it.

We expect that decomposing stochastic matrices under small complementarity constraints will rise in other applications. Consider for example a lottery which assigns students to courses. A student needs to care about her probability of getting each course, but would also like the guarantee that two courses will not overlap. While the algorithm we presented here has a worst case approximation ratio $2/\underline{q}$ (where \underline{q} is the minimal capacity in the problem), it behaves much better on simulated and on real data. The reason for this better behavior is that couples do not necessarily concentrate in the small hospitals (and indeed making couples prefer big hospitals improves the performance).

6.1 Open questions

1. What is a more principled way of balancing strategy-proofness with efficiency? We did not pay any extra price for strategy-proofness (in addition to the price for the DNH principle).
2. The students ended up giving no advantages to parents in 2014 and 2015. How would the algorithm behave given a population which consists also of parents? Is there a more principled way to cope with parents?
3. For the assignment problem with couples, what is the lower bound for

approximation within the domain $\mathcal{P}^{S \geq C}$ and with elements of M being “small”?

4. For the assignment problem with couples, our approximation algorithm and analysis assumed that couples must be matched to the same hospital. Can a similar approximation be found when couples can be in different hospitals within the same city?

References

- [1] ABDULKADIROĞLU, A. AND SÖNMEZ, T. 2003. School choice: A mechanism design approach. *The American Economic Review* 93, 3, 729–747.
- [2] AKBARPOUR, M. AND NIKZAD, A. 2014. Approximate random allocation mechanisms. available at SSRN 2422777.
- [3] ASHLAGI, I., BRAVERMAN, M., AND HASSIDIM, A. 2014. Stability in large matching markets with complementarities. *Operations Research*. forthcoming.
- [4] BIRKHOFF, G. 1946. Three observations on linear algebra. *Revi. Univ. Nac. Tucuman, ser A* 5, 147–151.
- [5] BOGOMOLNAIA, A. AND MOULIN, H. 2001. A new solution to the random assignment problem. *Journal of Economic Theory* 100, 2, 295–328.
- [6] BRONFMAN, S., HASSIDIM, A., AFEK, A., ROMM, A., SHERBERK, R., HASSIDIM, A., AND MASSLER, A. 2015. Assigning israeli medical graduates to internships. *Israel Journal of Health Policy Research*.
- [7] BUDISH, E. 2011. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy* 119, 6, 1061–1103.
- [8] BUDISH, E., CHE, Y.-K., KOJIMA, F., AND MILGROM, P. 2013. Designing random allocation mechanisms: Theory and applications. *The American Economic Review* 103, 2, 585–623.
- [9] FEATHERSTONE, C. R. 2014. Rank efficiency: Investigating a widespread ordinal welfare criterion. mimeo.

- [10] HYLLAND, A. AND ZECKHAUSER, R. 1979. The efficient allocation of individuals to positions. *The Journal of Political Economy*, 293–314.
- [11] I.HOLYER. 1981. The np-completeness of some edge-partition problems. *SIAM J. Comput.* 10, 4, 713–717.
- [12] KOJIMA, F., PATHAK, P. A., AND ROTH, A. E. 2013. Matching with couples: Stability and incentives in large markets. *The Quarterly Journal of Economics* 128, 4, 1585–1632.
- [13] LIU, Q. AND PYCIA, M. 2013. Ordinal efficiency, fairness, and incentives in large markets. *Fairness, and Incentives in Large Markets (September 2013)*.
- [14] NGUYEN, T., PEIVANDI, A., AND VOHRA, R. 2014. One-sided matching with limited complementarities. PIER Working Paper.
- [15] NGUYEN, T. AND VOHRA, R. 2014. Near feasible stable matchings with complementarities. PIER Working Paper.
- [16] PLUMMER, M. D. AND LOVÁSZ, L. 1986. *Matching theory*. Elsevier.
- [17] SHAPLEY, L. AND SCARF, H. 1974. On cores and indivisibility. *Journal of mathematical economics* 1, 1, 23–37.
- [18] SINKHORN, R. AND KNOPP, P. 1967. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics* 21, 2, 343–348.
- [19] VON NEUMANN, J. 1953. A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the Theory of Games* 2, 5–12.

A Lower Bound for small probabilities

Define an example with

$$\begin{aligned}
 H &= \{h_1, \dots, h_{2m}, h'_1, \dots, h'_{2m}\}, \\
 S &= \{s_1, \dots, s_{2m(2k+1)}\}, \text{ and} \\
 C &= \{c_1, \dots, c_{(2k+1)m}\},
 \end{aligned}$$

where m and k are integers. Consider the target matrix $M \in \mathcal{P}^{S \geq C}$ described in Table II. Under this stochastic assignment matrix each single intern in S gets a probability $\frac{1}{2m}$ to be assigned to each hospital in $\{h'_1, \dots, h'_{2m}\}$, and each couple in C gets a probability of $\frac{1}{2m}$ to be assigned to each hospital in $\{h_1, \dots, h_{2m}\}$. One can verify that the capacity of every hospital is $2k + 1$.

In each deterministic assignment, the couples occupy at most $4mk$ slots in the hospitals in $\{h_1, \dots, h_{2m}\}$, hence the singletons occupy at least $2m$ slots in $\{h_1, \dots, h_{2m}\}$. This is true for every assignment, hence true for their convex combination as well, meaning that some single intern spends at least $\frac{2m}{(2m)(2k+1)} = \frac{1}{2k+1}$ of his probability in the hospitals in $\{h_1, \dots, h_{2m}\}$. Thus, the approximation cannot be better than $\frac{2}{2k+1} = \left(\frac{2}{q}\right)$.

Table 2: Target matrix for small probabilities example

	h_1	\dots	h_{2m}	h'_1	\dots	h'_{2m}
s_1	0	\dots	0	$1/(2m)$	\dots	$1/(2m)$
\vdots	0	\dots	0	$1/(2m)$	\dots	$1/(2m)$
$s_{2m(2k+1)}$	0	\dots	0	$1/(2m)$	\dots	$1/(2m)$
$c_{1,1}$	$1/(2m)$	\dots	$1/(2m)$	0	\dots	0
$c_{1,2}$	$1/(2m)$	\dots	$1/(2m)$	0	\dots	0
\vdots	$1/(2m)$	\dots	$1/(2m)$	0	\dots	0
$c_{m(2k+1),1}$	$1/(2m)$	\dots	$1/(2m)$	0	\dots	0
$c_{m(2k+1),2}$	$1/(2m)$	\dots	$1/(2m)$	0	\dots	0