Individual rationality and participation in large scale, multi-hospital kidney exchange

Itai Ashlagi and Alvin E. Roth*

January 14, 2011

Abstract

As multi-hospital kidney exchange clearinghouses have grown, the set of players has grown from patients and surgeons to include hospitals. Hospitals have the option of enrolling only their hard-to-match patient-donor pairs, while conducting easily arranged exchanges internally. This behavior has already started to be observed.

We show that the cost of making it individually rational for hospitals to participate fully is low in almost every large exchange pool (although the worst-case cost is very high), while the cost of failing to guarantee individually rational allocations could be large, in terms of lost transplants. We also identify an incentive compatible mechanism.

1 Introduction

When kidney exchange was just beginning, most exchanges were conducted in single hospitals, or in closely connected networks of hospitals like the fourteen New England transplant centers organized by the New England Program for Kidney Exchange (Roth et al. (2005a)). But today exchanges often involve multiple hospitals that may have relatively little repeated interaction outside of kidney exchange. The present paper is meant to help establish a theoretical framework to study the kinds of problems that can be anticipated as the United States moves in the direction of nationally organized exchange, as it has begun to do in 2010.

*Ashlagi: iashlagi@mit.edu. Roth: al_roth@harvard.edu. We have had valuable conversations about this paper with Itay Fainmesser, Duncan Gilchrist, Jacob Leshno, and Mike Rees, and have benefited from comments by participants at the NBER Market Design conference and the Harvard-MIT Economic Theory seminar.

In particular, this paper concerns the growing problem of giving hospitals the incentive to participate fully, in order to achieve the gains that kidney exchange on a large scale makes possible. Our results suggest that, if care is taken in how kidney exchange mechanisms are organized, the problems of participation may be less troubling in large exchange programs than they are starting to be in multi-hospital exchanges as presently organized.

1.1 Background

Kidney transplantation is the treatment of choice for end stage renal disease, but there are many more people in need of kidneys than there are kidneys available. Kidneys for transplantation can come from deceased donors, or from live donors (since healthy people have two kidneys and can remain healthy with one). However not everyone who is healthy enough to donate a kidney and wishes to do so can donate a kidney to his or her intended recipient, since a successful transplant requires that donor and recipient be compatible, in blood and tissue types. This raises the possibility of *kidney exchange*, in which two or more incompatible patient-donor pairs exchange kidneys, with each patient in the exchange receiving a compatible kidney from another patient's donor.[1]

Note that it is illegal for organs for transplantation to be bought or sold in the United States and throughout much of the world (see Roth (2007) and Lieder and Roth (2010)). Kidney exchange thus represents an attempt to organize a barter economy on a large scale, with the aid of a computer-assisted clearinghouse.[2]

The first kidney exchange in the United States was carried out in 2000 at the Rhode Island Hospital, between two of the hospital's own incompatible patient-donor pairs.[3] Roth et al. (2004) made an initial proposal for organizing kidney exchange on a large scale, which included the ability to integrate cycles and chains, and considered the incentives that well designed allocation mechanisms would give to participating patients and their surgeons to reveal relevant information about patients. The surgical infrastructure available in 2004 meant that only pairwise exchanges (between exactly two incompatible patient donor pairs) could initially be considered, and Roth et al. (2005b) proposed a mechanism for accomplishing this,

[1] In addition to such cyclic exchanges, chains are also possible, which involve not only incompatible patient donor pairs, and begin with a deceased donor or an undirected donor (one without a particular intended recipient), and end with a patient with high priority on the deceased donor waiting list, or with a donor who will donate at a future time.

[2] Recall that Jevons (1876) proposed that precisely the difficulties of organizing barter economies–in particular, the difficulty of satisfying the "double coincidence" of wants involved in simultaneous exchange without money–had led to the invention of money.

[3] For an account of this and other early events in kidney exchange see Roth (2010), "The first kidney exchange in the U.S., and other accounts of early progress," http://marketdesigner.blogspot.com/2010/04/first-kidney-exchange-in-us-and-other.html

again paying close attention to the incentives for patients and their surgeons to participate straightforwardly. As kidney exchanges organized around these principles gained experience, Saidman et al. (2006) and Roth et al. (2007b) showed that efficiency gains could be achieved by incorporating chains and larger exchanges that required only relatively modest additional surgical infrastructure, and today there is growing use of larger exchanges and longer chains, particularly following the publication of Rees et al. (2009).

Roth et al. (2005a) describe the formation of the New England Program for Kidney Exchange (NEPKE) under the direction of Dr. Frank Delmonico, and these proposals were also instrumental in helping organize the Alliance for Paired Donation (APD) under the direction of Dr Mike Rees.[4] In 2010 a National Kidney Paired Donation Pilot Program became operational, still on a very small scale.[5]

During the initial startup period, there was some evidence that attention to the incentives of patients and their surgeons to reveal information was important. But as infrastructure has developed, the information contained in blood tests has come to be conducted and reported in a more standard manner (sometimes at a centralized testing facility), reducing some of the choice about what information to report, with what accuracy. So some strategic issues have become less important over time (and indeed the current practice at both APD and NEPKE does not deal with the provision of information that derives from blood tests as an incentive issue). However, as kidney exchange has become more widespread, and as multi-hospital exchange consortia have been formed and a national exchange is being explored, the "players" are not just (and perhaps not even) patients and their surgeons, but hospitals (or directors of transplant centers). And as kidney exchange is practiced on a wider scale, free riding has become possible, with hospitals having the option of participating in one or more kidney exchange networks but also of withholding some of their patient-donor pairs, or some of their non-directed donors, and enrolling those of their patient-donor pairs who are hardest to match, while conducting more easily arranged exchanges internally. Some of this behavior is already observable.

The present paper considers the 'kidney exchange game' with hospitals as the players, to clarify the issues currently facing hospitals in existing multi-hospital exchange consortia, and those that would face hospitals in a large-scale national kidney exchange program.

[4]Today, in addition to those two large kidney exchange clearinghouses, kidney exchange is practiced by a growing number of hospitals and formal and informal consortia (see Roth (2008)). Computer scientists have become involved, and an algorithm of Abraham, Blum, and Sandholm (2007) designed to handle large populations was briefly used in the APD, and algorithms of that sort may form the basis of a national exchange.

[5]The national pilot program ran two pilot matches in October and December of 2010. Under its initial guidelines, only exchanges are considered, not chains.

1.2 Individual rationality

Hospitals participate in a multi-center exchange by reporting a list of incompatible patient-donor pairs to a central clearinghouse, and a matching mechanism chooses which exchanges to carry out.[6] At the same time, some hospitals conduct exchanges only internally among their own patients, and even hospitals participating in multi-center exchange programs may conduct some internal exchanges, and may participate in more than one exchange program.

To see why it is important to have hospitals participate in centralized exchange, we ran a simulation to compare the number of transplants that can be done when each hospital conducts only *internal exchanges* (consisting of pairs only from the same hospital) with the number of transplants a centralized mechanism can potentially produce given that it has access to all incompatible pairs.[7] The efficiency gains from centralization grow as the number of (moderate sized) hospitals increases: for two hospitals having around 11 pairs each, centralized kidney exchange can potentially increase transplants by about 50% compared to the internal exchanges that could be accomplished, but this rises to an increase of almost 300% when we consider 22 hospitals with an average of 11 pairs each (see also Toulis and Parkes (2010) who analytically quantify the benefit from a centralized clearinghouse for organizing 2-way exchange).

However when kidney exchange clearinghouses try to maximize the (weighted) number of transplants without attention to whether those transplants are internal to a hospital, it may not be individually rational for a hospital to contribute those pairs it can match internally (cf. Roth 2008).[8] For example, consider a hospital a with two pairs, *a1* and *a2*, that it can

[6]NEPKE initially organized the fourteen transplant centers in New England (cf. Roth et al. (2005a)) and now includes several others, and the APD now counts as members several dozen hospitals around the country (with varying degrees of participation).

[7]We briefly explain here the Monte-Carlo simulations.

To generate incompatible pairs we use a method similar to Saidman et al. (2006). First we create a patient and donor with blood-types drawn from the national distributions as reported by Roth et al. (2007b). Blood type compatibility is not sufficient for transplantation. Each patient is also assigned a percentage reactive antibody (PRA) level also drawn from a distribution as in Roth et al. (2007b). Patient PRA is interpreted as the probability of a positive crossmatch (tissue type incompatibility) with a random donor. If the generated pair is compatible, i.e. if they are both blood type compatible and have a negative crossmatch, they are discarded (this captures the fact that compatible pairs go directly to transplantation). Otherwise the population generation continues until each hospital accumulates a certain number of incompatible pairs. In our simulations the number of incompatible pairs for each hospital is drawn from a discrete uniform distribution on $[8, 14]$. For each generated population we ran 500 trials.

When allowing 3-way exchanges, finding an allocation that maximizes the number of matches is an NP hard problem (see Abraham et al. (2007) and Biro et al. (2009)). The compatibility graph is generally sparse enough however that the problem is tractable in reasonably sized populations.

[8]Some weighted matching algorithms currently in use put some weight on internal exchanges, but this

match internally. Suppose it enters those two pairs in a centralized exchange. It may be that the weighted number of transplants is maximized by including *a1* in an exchange but not *a2*, in which case only one of hospital a's patients will be transplanted, when it could have performed two transplants on its own.

This is becoming a visible problem, as membership in a kidney exchange network does not mean that a hospital does not also do some internal exchanges. Mike Rees, the director of the APD, writes (personal communication):

> "...competing matches at home centers is becoming a real problem. Unless it is mandated, I'm not sure we will be able to create a national system. I think we need to model this concept to convince people of the value of playing together".

This paper attempts to understand the problem raised by the APD director. We will see that when the number of hospitals and incompatible pairs is small, it may be costly (in terms of lost transplants) for a centralized clearinghouse to guarantee hospitals individual rationality, compared to how many transplants could be accomplished if all pairs were submitted to a centralized exchange despite no guarantee of individual rationality. However in large markets we will show that this cost becomes very low, and we begin to explore incentive compatible mechanisms for achieving full participation by hospitals as efficiently as possible.

2 Kidney Exchange and Individual Rationality

2.1 Exchange pools

An exchange pool consists of a set of patient-donor pairs. A patient p and a donor d are **compatible** if patient p can receive the kidney of donor d and **incompatible** otherwise. It is assumed that every pair in the pool is incompatible.[9] Thus a pair is a tuple $v = (p, d)$ in which donor d is willing to donate his kidney to patient p but p and d are incompatible. We further assume that each donor and each patient belong to a single pair.

An exchange pool V induces a **compatibility graph** $D(V) = D(V, E(V))$ which captures the compatibilities between donors and patients as follows: the set of nodes is V, and for every pair of nodes $u, v \in V$, (u, v) is an edge in the graph if and only if the donor of

does not solve the problem, since it neither guarantees a hospital the exchanges it could conduct internally, nor does it guarantee that the pairs that could be internally exchanged will be used efficiently if submitted to the central clearinghouse.

[9]Pairs that are compatible would presently go directly to transplantation and not join the exchange pool (but see e.g. Roth et al. (2005a) on the advantages of changing this policy).

node u is compatible with the patient of node v. We will use the terms nodes and pairs interchangeably.

An exchange can now be described through a cycle in the graph. Thus an **exchange** in V is a cycle in $D(V)$, i.e. a list $v_1, v_2, \ldots, v_k$ for some $k \geq 2$ such that for every $i, 1 \leq i < k$, $(v_i, v_{i+1}) \in E(V)$ and $(v_k, v_1) \in E(V)$. The size of an exchange is the number of nodes in the cycle. An **allocation** in V is a set of distinct exchanges in $D(V)$ such that each node belongs to at most one exchange. Since in practice the size of an exchange is limited (mostly due to logistical constraints), we assume there is an exogenous maximum size limit $k > 0$ for any exchange. Thus if $k = 3$ only exchanges of size 2 and 3 can be conducted.[10]

Let M be an allocation in V. We say that node v is **matched** by M if there exists an exchange in M that includes v. For any set of nodes $V' \subseteq V$ let $M(V')$ be the set of all nodes in V' that are matched (or "covered") by M.

We will be interested in finding efficient allocations, that have as many transplants as possible. Two types of efficiency will be considered. M is called **k-efficient** if it matches the maximum number of transplants possible for exchanges of size no more than k, i.e. there exists no other allocation M' consisting of exchanges of size no more than k such that $|M'(V)| > |M(V)|$.[11] M is called **k-maximal** if there exists no such allocation M' such that $M'(V) \supsetneq M(V)$. A matching will be called **efficient** (or **maximal)** if it is k-efficient (or k-maximal) for unbounded k, i.e. for no limit on how many transplants can be included in an exchange. Note that every k-efficient allocation is also k-maximal. The converse is not true. However for $k = 2$, both types of efficiency coincide, since the collection of sets of simultaneously matched nodes in allocations forms a matroid (see Edmonds et al. (1971)).

A **Kidney Exchange Program** (or simply a Kidney Exchange) consists of a set of n hospitals $H_n = \{h_1, \ldots, h_n\}$ and a set of incompatible pairs V_h for each hospital $h \in H_n$. We let $V_{H_n} = \cup_{h \in H_n} V_h$. The compatibility graph induced by V_{H_n} is called the **underlying graph**. We will take the hospitals (e.g. the director of transplantation at each hospital) as the active decision makers in the Kidney Exchange, whose choices are which incompatible pairs to reveal to the Exchange. We will approximate the preferences of hospitals as being concerned only with their own patients. Mostly we will assume hospitals are concerned only with the *number* of their patients who receive transplants, although we do not rule out hospitals having preferences over which of their patients are transplanted.

An exchange and that matches only pairs from the same hospital is called **internal**.

[10] In the APD and NEPKE k was originally set to 2, was increased to 3, and now optimization is conducted over even larger exchanges and chains, and the pilot national program considers exchanges up to size 3. Exchanges are generally conducted simultaneously, so an exchange of size k requires 2k operating rooms and surgical teams for the k nephrectomies (kidney removals) and k transplants.

[11] In graph theory a 2-efficient allocation is referred to as a maximum matching.

Hospital h can match a set of pairs $B_h \subseteq V_h$ **internally** if there exists an allocation in V_h such that all nodes in B_h are matched.

2.2 Participation constraints: individual rationality for hospitals

The kidney exchange setting invites discussions of various types of individual rationality (IR). In this paper an allocation is not individually rational if some hospital can internally match more pairs than the number of its pairs matched in the allocation. Formally, an allocation M in V_{H_n} is not **individually rational** if there exists a hospital h and an allocation M_h in V_h such that $|M(V_h)| < |M_h(V_h)|$.[12]

To illustrate this, consider the compatibility graph in Figure 1, where nodes a_1 and a_2 belong to hospital a and b_1 and b_2 belong to hospital b. The only individually rational allocation is the one that matches a_1 and a_2.

Remark: Throughout this paper, undirected edges represent two directed edges, one in each direction.

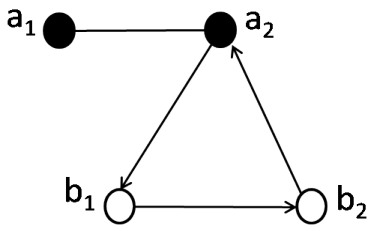

Figure 1: No 3-efficient allocation is individually rational.

In the next section we study worst case efficiency loss from choosing IR allocations.

3 IR and Efficiency: Worst case results for compatibility graphs

By choosing the individually rational allocation in Figure 1 we obtain 2 transplants whereas the efficient allocation provides 3. The following result shows that a maximum individually rational allocation can be very costly in the worst case:

[12]Other formulations of individual rationality may be appropriate under some circumstances, such as requiring not merely that a hospital be allocated the same number of transplants that it can achieve on its own, but that it can be guaranteed to transplant a set that includes all the individuals it could match on its own.

Theorem 3.1. *For every $k \geq 3$, there exists a compatibility graph such that no k-maximal allocation which is also individually rational matches more than $\frac{1}{k-1}$ of the number of nodes matched by a k-efficient allocation. Furthermore in every compatibility graph the size of a k-maximal allocation is at least $\frac{1}{k-1}$ times the size of a k-efficient allocation.*

Proof. Let V be a set of nodes and let M be a k-efficient allocation and M' be a k-maximal individually rational allocation in V. Since M' is k-maximal, every exchange in M must intersect an exchange in M' (otherwise a disjoint exchange could be added to M', contradicting maximality). Fix an exchange c with size $2 \leq l \leq k$ in M'. The maximum number of nodes that might be covered by M and not M' would be achieved if for each such exchange c, M contains $l-1$ exchanges each of size k, which each intersect exactly one node of c (and M'). (Note that if all l nodes of c were in such exchanges then M' wouldn't be maximal.) For each such exchange c, M matches $(l-1)k$ nodes and M' matches l nodes, so the ratio is $l/(l-1)k$, which is minimized at $1/(k-1)$ when $l = k$, giving the desired bound.

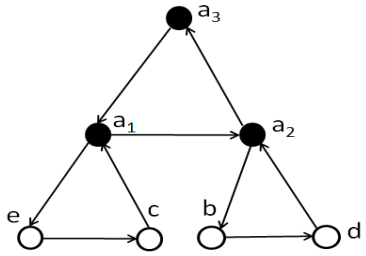

Figure 2: Worst case efficiency loss from choosing an individually rational allocation ($k = 3$).

To see that the bound is tight, observe that the construction used to find the bound achieves it: fix some hospital a with k vertices, and suppose that a has a single internal exchange consisting of all of its pairs (see Figure 2 for an illustration for $k = 3$). The bound $\frac{1}{k-1}$ is obtained by letting the k-efficient allocation in the underlying graph consists of exactly $k-1$ exchanges each of size k, at which a single pair of a is part of each such exchange. That is, the efficient allocation matches all but one of hospital a's pairs, each in exchanges of size k with $k-1$ pairs from other hospitals. ∎

By requiring only a maximal allocation (rather than an efficient one), one can also obtain individual rationality:

Proposition 3.2. *For every $k \geq 2$, and every compatibility graph there exists a k-maximal allocation that is individually rational.*

The proof of Lemma 3.2 is by construction using the following simple augmenting algorithm which is based on the augmenting matching algorithm of Edmonds (1965); Such an

algorithm begins by finding a k-efficient allocation in V_h for every hospital h, and then it repeatedly searches for an allocation that only increases the total number of matched pairs without unmatching any pair that was previously matched (although possibly rematching such pairs using different edges), until such an allocation cannot be found.

As noted above, for the case $k = 2$ every 2-maximal allocation is also 2-efficient. Therefore by applying the augmenting algorithm:

Proposition 3.3. *There exists an individually rational allocation with exchanges of size at most 2 that is also 2-efficient in every compatibility graph.*

So Theorem 3.1 shows that for k>2 there is a very high potential cost of individual rationality, but it gives a worst-case result. However, it appears that the *expected* efficiency loss from requiring individual rationality can be very small. Indeed our simulations show that **if all incompatible pairs are in the same exchange pool, the average number of patients who do not get a kidney due to requiring IR is less than 1** (see Table 1). But as we shall see in Section 8 the cost of *failing* to guarantee individual rationality could be large if that causes hospitals to match their own internal pairs.

No. of hospitals	2	4	6	8	10	12	14	16	18	20	22
IR,k=3	6.8	18.37	35.42	49.3	63.68	81.43	97.82	109.01	121.81	144.09	160.74
Efficient,k=3	6.89	18.67	35.97	49.75	64.34	81.83	98.07	109.41	122.1	144.35	161.07

Table 1: Number of transplants achieved using maximum size individually rational allocations vs. using efficient (and not necessarily individually rational) allocations.

In the next sections we will prove that the efficiency loss from choosing an IR allocation of maximum size is small in large compatibility graphs, supporting the simulations results.

4 Random Exchange Pools

To discuss the Bayesian setting it is useful to consider random compatibility graphs. Each person in the population has one of 4 blood types A,B, AB. and O, according to whether her blood contains the proteins A, B, both A and B, or neither. The probability that a random person's blood type is X is given by $\mu_{\mathrm{X}} > 0$. We will assume that $\mu_O > \mu_{\mathrm{A}} > \mu_{\mathrm{B}} > \mu_{\mathrm{AB}}$ (as in the U.S. population).[13] For any two blood types X and Y, we write Y $\rhd$ X if a donor of

[13]In practice $\mu_{\mathrm{O}} = 0.48$, $\mu_{\mathrm{A}} = 0.34$, $\mu_{\mathrm{B}} = 0.14$ and $\mu_{\mathrm{AB}} = 0.04$.

blood type Y and a patient with blood type X are blood type compatible, which occurs if X includes whatever blood proteins A and B are contained in Y.[14]

A patient-donor pair has pair type (or just type, whenever it is clear from the context) X-Y if the patient has blood type X and the donor has blood type Y. The set of pair types will be denoted by $\mathcal{P}$. In order for a donor and a patient to be compatible they need to be **both** blood type compatible and tissue-type compatible. To test tissue type compatibility a *crossmatch* test is performed. Each patient has a level of *percentage reactive antibodies* (PRA) which determines the likelihood that the patient will be compatible with a random donor. The lower the PRA of a patient, the more likely the patient is compatible with a random donor. In this paper we simplify the PRA characteristics and assume there exist two levels of PRA, L and $H (L < H)$; the probability that a patient p with PRA L (H) and a donor are tissue type *in*compatible is given by γ_L (γ_H). Furthermore the probability that a random patient has PRA L is given by $\upsilon > 0$. Let $\bar{\gamma}$ denote the expected PRA level of a random patient, that is $\bar{\gamma} = \upsilon\gamma_L + (1 - \upsilon)\gamma_H$.

Definition 4.1 (Random Compatibility Graph)**.** *A* ***random (directed) compatibility graph*** *of size m, denoted $D(m)$, consists of m incompatible patient-donor pairs, and a random edge is generated between every donor and each patient compatible with that donor. Hence, such a graph is generated in two phases:*

1. *Each node/incompatible pair in the graph is randomized as follows. A patient p and a potential donor d are created with blood types drawn independently according to the probability distribution $\mu = (\mu_X)_{X\in\{A,B,AB,O\}}$. The PRA of patient p, denoted by $\gamma(p)$, is also randomized (L with probability υ and H with probability $1 - \upsilon$).*

   A number z is drawn uniformly from $[0, 1]$ and (p, d) forms a new node if and only if p and d are blood type incompatible or p and d are blood type compatible but $z \leq \gamma(p)$ (so p and d are tissue type incompatible).

2. *For any two pairs $v_1 = (p_1, d_1)$ and $v_2 = (p_2, d_2)$, there is an edge from v_1 to v_2 if and only if d_1 and p_2 are ABO compatible and also tissue type compatible (d_1 is tissue type compatible with p_2 with probability $1 - \gamma(p_2)$).*

[14]Thus type O patients can receive kidneys only from type O donors, while type O donors can give kidneys to patients of any blood type. Note that since only *incompatible* pairs are present in the kidney exchange pool, donors of blood type O will be underrepresented, since most such donors will be compatible with their intended recipients; the only incompatible pairs with an O donor will be tissue-type incompatible. (Roth et al. (2005a) showed that a significant increase in the number of kidney exchanges could be achieved by allowing compatible pairs to participate, but this has not become common practice.)

We will often denote a random compatibility graph by $D(H_n)$, thus $D(H_n) = D(m)$ where m is the total number of pairs in all hospitals belonging to H_n.

The posterior probability that an incompatible pair (p, d) is of type X-Y will be denoted by $\mu_{\text{X-Y}}$. Define $\frac{1}{\rho}$ to be the probability that a random pair (p, d) is incompatible. Thus if X and Y are two blood types such that Y $\rhd$ X then $\mu_{\text{X-Y}} = \rho\mu_{\text{X}}\mu_{\text{Y}}\bar{\gamma}$ and otherwise $\mu_{\text{X-Y}} = \rho\mu_{\text{X}}\mu_{\text{Y}}$.

We are going to model kidney exchange with many patient-donor pairs (in many hospitals) as a large random compatibility graph and use results and methods from random graph theory.

4.1 Random graphs - background

We briefly describe here some results that will provide intuition and be building blocks in our proofs. A random graph $G(m, p)$ is a graph with m nodes and between each two different nodes an *undirected* edge exists with probability p (p is a non-increasing function of m). A bipartite random graph $G(m, m, p)$ consists of two disjoint sets of nodes V and W, each of size m, and an *undirected* edge between any two nodes $v \in V$ and $w \in W$ exists with probability p (no two nodes within the same set V or W have an edge between them). It will be useful to think of an undirected edge as two directed edges, one in each direction.

Throughout the paper by saying just a "random graph" we will not refer to a specific type, but a graph that is generated by any of the graph generating processes defined in this paper (e.g., $D(m), G(m, p)$, and $G(m, m, p)$).

For any graph theoretic property Q there is a probability that a random graph G satisfies Q, denoted by $\Pr(G \models Q)$. The property Q is monotone if this probability is monotone in p.

A **matching** in an undirected graph is a set of edges for which no two edges have a node in common. A matching is **nearly perfect** if it matches (contains) all but at most one nodes in the graph, and **perfect** if it matches all nodes. Note that the existence of a (nearly) perfect matching in an undirected graph is a monotone property.

Erdos and Renyi provided a threshold function $r(m) = \frac{\ln m}{m}$ for the existence of a perfect matching in $G(m, p(m))$. We state here a corollary of their result.

Erdos-Renyi Theorem: *Let Q be the property that there exists a nearly perfect matching. For any constant p*

1. $\Pr(G(m, p) \models Q) = 1 - o(1)$.[15][16]

[15] For any two functions f and g we write $f = o(g)$ if the limit of the ratio $\frac{f(n)}{g(n)}$ tends to zero when n tends to infinity.

[16] The Erdos-Renyi theorem showed stronger results, which asserts that $r(m) = \frac{\ln m}{m}$ is a threshold function

2. $\Pr(G(m,m,p) \models Q) = 1 - o(1)$.

Remark on the convergence rate: The probability of a perfect matching in $G(m,p)$ and $G(m,m,p)$ converges to 1 at an exponential rate for any constant p. More precisely, as shown in Janson et al. (2000),

$$\Pr(G(m,m,p) \models Q) = 1 - O(me^{-mp}) = 1 - o(2^{-mp}),$$

and clearly a perfect matching in $G(m,p)$ exists with at least the same convergence rate. We will often just write $1-o(1)$ in our results and proofs, however the reader should bear in mind that each time we write $1-o(1)$ it can be replaced by an exponential rate, so we are dealing with fairly rapid convergence, and we will see this in the accompanying simulations.

For simplicity we adopt the following formalism from random graph theory: *if the probability that a given property Q is satisfied in a random graph G tends to 1 when m tends to ∞, we say that Q holds in* ***almost every (large)*** *G.*

In the next section we study efficiency in large random compatibility graphs. We let γ_L and γ_H (the probability of tissue type incompatibility for patients with low or high PRA) be non-decreasing functions of m, with the important special case in which both are constants.

5 Efficient Allocations in Large Random Compatibility Graphs

The relative number of pairs of various types will be useful in studying efficient allocations.

Lemma 5.1. *In almost every large $D(m)$:*

1. *For all $X \in \{A, B, AB\}$ the number of O-X pairs is larger than the number of X-O pairs.*
2. *For all $X \in \{A, B\}$ the number of X-AB pairs is larger than the number of AB-X pairs.*
3. *The absolute difference between the number of A-B pairs and B-A pairs is $o(m)$. Consequently this difference is smaller than the number of pairs of any other pair type.*[17]

for the existence of a perfect matching; that is, if $p = p(m)$ is such that $r(m) = o(p(m))$ then the probability a nearly perfect matching exists converges to 1, and if $p(m) = o(r(m))$, the probability a nearly perfect matching exists converges to 0.

[17] Terasaki et al. (1998) claim that the frequency of A-B pairs (0.05) is larger than B-A pairs (0.03) but they do not give any data or other explanation to support their claim. Our result just asserts that the absolute difference is "small".

Lemma 5.1 whose proof appears in the appendix, motivates the following partition of patient-donor pair types $\mathcal{P}$ (see also Roth et al. (2007b) and Ünver (2010)): Let

$$\mathcal{P}^{\mathcal{O}} = \{\text{X-Y} \in \mathcal{P} : \text{Y} \rhd \text{X} \quad \text{and} \quad \text{X} \neq \text{Y}\}$$

be the set of **overdemanded** types.

Let

$$\mathcal{P}^{\mathcal{U}} = \{\text{X-Y} \in \mathcal{P} : \text{X} \rhd \text{Y} \quad \text{and} \quad \text{X} \neq \text{Y}\}$$

be the set of **underdemanded** types.

Let

$$\mathcal{P}^{\mathcal{S}} = \{\text{X-X} \in \mathcal{P}\}$$

be the set of **selfdemanded** types, and finally let $\mathcal{P}^{\mathcal{R}}$ be the set of **reciprocally demanded** types which consists of types A-B and B-A.

Intuitively, an over-demanded pair is offering a kidney in greater demand than the one they are seeking. For example a patient whose blood type is A and a donor whose blood type is O form an overdemanded pair. Underdemanded types have the reverse property: they are seeking a kidney that is in greater demand than the one they are offering in exchange. A donor and patient in a pair with a selfdemanded type have the same blood type.

We will make the following assumptions which are compatible with blood type frequencies and with observed tissue-type sensitivity frequencies. Zenios et al. (2001) reported that for non-related blood type donors and recipients $\bar{\gamma} = 0.11$.[18]

Assumption A [Non-highly-sensitized patients] $\bar{\gamma} < \frac{1}{2}$.[19]

Assumption B [Blood type frequencies] $\mu_{\text{O}} < 1.5\mu_{\text{A}}$.[20]

Proposition 5.2. *Almost every large $D(m)$ has an efficient allocation that requires exchanges of no more than size 3 with the following properties:*

1. *Every selfdemanded pair X-X is matched in a 2-way or a 3-way exchange with other selfdemanded pairs (no more than one 3-way exchange is needed, in the case of an odd number of X-X pairs).*

[18] One can extend our results to a larger tissue-type incompatibility probability, but the most highly sensitized patients are a topic for another day, since the large graph approximations we use here do not adequately model the situation facing a very highly sensitized patient in a finite market.

[19] This assumption is also used for avoiding case-by-case analysis; one can provide similar results for the opposite inequality. However the limit results we obtain here for large compatibility graphs are less of a good approximation to the situation facing *very* high PRA patients in the finite graphs we see in practical applications than they are for the situation facing the large majority of patients who are not extremely highly sensitized. We will return to this, and the open questions it raises, in the conclusion.

[20] We will use this assumption to construct the efficient allocation. However even if this assumption does not hold, using a similar method of proof one can construct a very similar allocation. The details of the efficient allocation would slightly change, but not our results about individually rational allocations.

2. *Either every B-A pair is matched in a 2-way exchange with an A-B pair or every A-B pair is matched in a two way exchange with a B-A pair.*

3. *Let $X, Y \in \{A, B\}$ and $X \neq Y$. If there are more Y-X than X-Y then every Y-X pair that is not matched to an X-Y pair is matched in a 3-way exchange with an O-Y pair and an X-O pair.*

4. *Every AB-O pair is matched in a 3-way exchange with an O-A pair and an A-AB pair.*

5. *Every overdemanded pair X-O ($X \neq O$) that is not matched as above is matched to an O-X pair.*

The proof of Proposition 5.2 is deferred to the Appendix. Roth Sonmez and Unver (2007) show a similar result to Proposition 5.2 and a similar result can also be derived from Unver (2009). Both these papers however assumed that *there are no tissue type incompatibilities between patients and other patients' donors* in order to approximate a large market. Our result provides a mathematical foundation to essentially justify their assumption. In addition, both papers show that exchanges of size at most $k = 4$ are needed to find an efficient allocation (Unver (2009) also analyzes a dynamic world). The difference from our result (we need at most 3-way exchanges) follows from the fact that they assumed that there are more A-B pairs than B-A pairs (Unver assumes that the probability for a pair to be of type A-B is greater than the probability that it will be of type B-A). In fact simulations by Roth Sonmez and Unver (2007) find very few four way exchanges are needed. It is important to note that although $\mu_{\text{A-B}} = \mu_{\text{B-A}}$ the probability that the number of each of such pairs is different is positive, the difference between the number of these pairs, as Lemma 5.1 implies, will almost always be sufficiently small to make 4-way exchanges unnecessary.[21] Independently of our work, Toulis and Parkes (2010) study 2-way exchanges using random graphs, and provide a very similar efficient allocation.

5.1 Sketch of proof for Proposition 5.2

We will use a simple extension of the Erdos-Renyi Theorem (Lemma 9.5 in the appendix) to l-partite directed graphs ($l \geq 2$) which asserts that if at most one of the l sets (parts of the graph) does not grow to infinity then almost every such large graph consists of a *perfect* allocation, that is an allocation which matches all the pairs in the smallest part of the graph.

[21] Our result would hold also in a model at which $\mu_{\text{A-B}} \neq \mu_{\text{B-A}}$ but the difference between these two probabilities is sufficiently small.

We assume that the number of A-B pairs is at least the number of B-A pairs (for the converse a symmetric argument holds). An application of the Erdos-Renyi Theorem establishes that all selfdemanded pairs can be matched using 2-way or 3-way exchanges to each other with high probability. Similarly all B-A pairs can be matched to A-B pairs using 2-way exchanges. We choose such a preliminary allocation arbitrarily, say M_1.

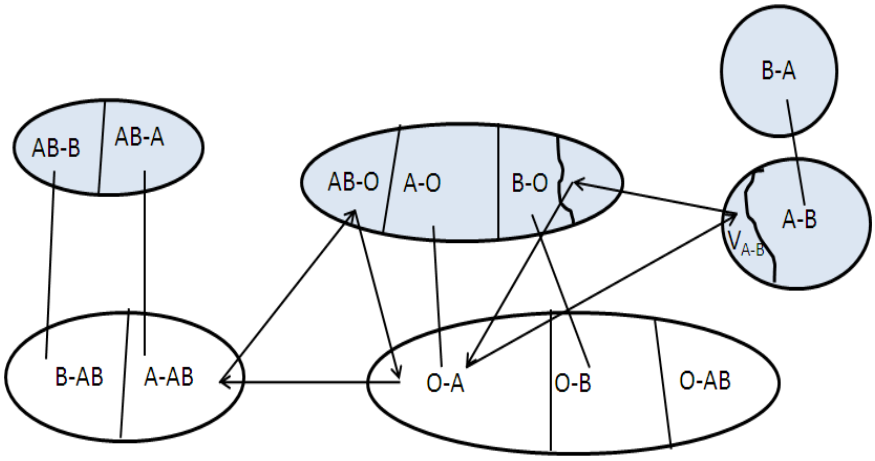

Figure 3: The structure of an efficient allocation in the graph $D(m)$ (excluding all selfdemanded pairs). All B-A pairs are matched to A-B (assuming there are more B-A than A-B), the remainder of the A-B pairs ($V_{\text{A-B}}$) are matched in 3-way exchanges using O-A's and B-O's. AB-O are matched in 3-ways each using two overdemanded pairs, and every other overdemanded pair is matched to a corresponding underdemanded pair.

Let $V_{\text{A-B}}$ be the set of A-B pairs that are not matched so far by M_1 (see Figure 3). By Lemma 5.1 the cardinality of the set $V_{\text{A-B}}$ is smaller than both the size of the set of B-O pairs and the size of the set of O-A pairs. Again by an extension of the Erdos-Renyi theorem this graph almost always contains a perfect allocation, implying that all A-B pairs can be matched (i.e. those not matched to B-A pairs are matched in 3-way exchanges of the form A-B,B-O,O-A). Similarly one can match with high probability all AB-O pairs using 3-way exchanges each containing A-AB pairs and O-A pairs.[22] Using our assumptions on $\bar{\gamma}$ one can show that there are many more O-B pairs and O-A pairs than B-O and A-O pairs respectively that are yet to be matched. Therefore all remaining overdemanded pairs can be matched to underdemanded pairs (again by considering the bipartite graphs induced by those pairs).

Finally the efficiency of our construction roughly follows by observing that (i) all pairs

[22]Here is where we use Assumption B. If assumption B would not hold, then the construction allocation slightly changes by having some AB-O pairs matched in 3-way exchanges using O-B and B-AB pairs, and if necessary also some AB-O pairs matched in 2-way exchanges using O-AB pairs.

but underdemanded are matched, and that (ii) no overdemanded pair can help more than 2 underdemanded pairs to get a transplant, and AB-O is the only pair type that can help 2 underdemanded pairs to get a transplant. Therefore, more than 3-way exchanges are not needed to obtain efficiency.

5.2 Remarks on efficient allocations

By construction every pair whose type is colored in Figure 3 (as well as all selfdemanded pairs) is matched, implying that we obtained a 3-efficient allocation. Roth et al. (2007b) considered 4-way exchanges with pairs AB-O,O-A,A-B and B-AB to obtain efficiency (see Figure 4). However, such an exchange uses an AB-O pair and an A-B pair that is not matched to a B-A pair. But pairs of both these types can all be matched in 3-way exchanges as in Figure 4, implying that using such a 4-way will result in fewer transplants. (That is, the 4-way exchange is made at the expense of two 3-way exchanges, see Figure 4.)

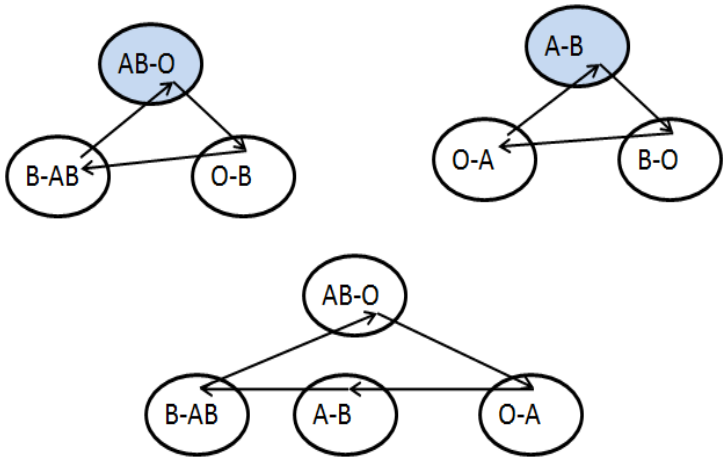

Figure 4: The possible 4-way exchange uses the bottlenecks of the 3-way exchanges - AB-O pairs and A-B pairs.

From Proposition 5.2 and its proof one can derive the efficiency loss between different k's in large graphs.

Corollary 5.3. *Let $W_k(m)$ be the size of a k-efficient allocation in $D(m)$.*

1. *For every $k \geq 2$, $\lim_{m\to\infty} \Pr\left(W_3(m) \geq W_k(m)\right) = 1$.*
2. *For every $\epsilon > 0$, $\lim_{m\to\infty} \Pr\left(W_3(m) - W_2(m) \leq (1+\epsilon)(\mu_{AB-O})m + \epsilon\mu_{A-B}m\right) = 1$.*
3. *In almost every large $D(m)$, in all efficient allocations all pairs whose type is not underdemanded are matched.*

The second part of Corollary 5.3 follows by constructing a 2-way efficient allocation in a similar way as in Proposition 5.2.[23]

One possibly undesirable feature of the efficient allocation is that underdemanded pairs of type O-AB will all be left unmatched. While it is inevitable that many underdemanded pairs will be left unmatched, there is sometimes discomfort in medical settings having *a priori* identifiable pairs seemingly singled out. A natural outcome would be that hospitals would seek to match such pairs internally, a point to which we will return later, when we observe that precisely these internal matches account for most of the efficiency cost of individual rationality.

Until this point nothing has been said about individual rationality in the Bayesian setting. In the next section we study the efficiency cost of requiring an allocation to be individually rational in large exchange pools.

6 Individual Rationality is Not Very Costly in Large Random Compatibility Graphs

In Section 3 we saw that individually rational allocations can harm efficiency, and provided worst case tight bounds. In this section we derive a very much smaller upper bound on this loss for large random compatibility graphs.

One way to bound the efficiency loss is by attempting to construct an efficient allocation as in Proposition 5.2, while making sure that the pairs each hospital can internally match are part of the efficient allocation. Unfortunately such an allocation is not always feasible.

Consider for example the following two *unbalanced* 3-way exchanges (B-O,O-A,A-B) and (A-O,O-B,B-A). Too many 3-way internal exchanges of the second type, for example, as well as other internal exchanges that include O-B pairs but not B-O pairs (see e.g. Figure 5), could possibly lead to a situation in which, to fulfill individual rationality requirements set by internal exchanges, more O-B pairs would potentially need to be matched than the total number of B-O pairs. This can harm efficiency since as Theorem 5.2 suggests more transplants are obtained by choosing the two 2-way exchanges rather than the 3-way exchange in Figure 5.

Individual rationality, however, does not require the clearinghouse to match a specific maximum set of pairs that each hospital can internally match, but only to guarantee to match at least the *number* of pairs each hospital can internally match. For example if a hospital has an internal unbalanced exchange A-O,O-B,B-A and an internally unmatched

[23] In particular AB-O pairs will be matched to O-AB pairs using 2-way allocations rather than being matched in a 3-way as described in Proposition 5.2.

O-A pair, then to satisfy individual rationality it is sufficient to match the A-O,B-A and O-A pairs.

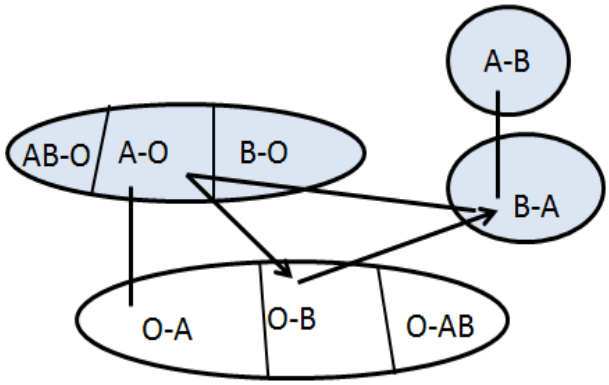

Figure 5: A 3-way (internal) exchange that matches an O-B pair for which there may not be a corresponding B-O pair.

As the above discussion suggests, individually rational allocations may contain (many) more underdemanded pairs of a specific type than its reciprocally overdemanded type. However if hospitals are not "too big" this is not likely to happen, since unbalanced three way exchanges are unlikely to occur among a small number of patient donor pairs. We say that a hospital of size c is **regular** if by randomly choosing an internal allocation that maximizes the number of underdemanded pairs, for any underdemanded type X-Y the expected number of matched X-Y pairs is less than the expected number of overdemanded Y-X pairs in its pool.

We formalize this. For any type $t \in \mathcal{P}$ and set of pairs S we denote by $\tau(S,t)$ the set of pairs with type t in S and for a set of types $T \subseteq \mathcal{P}$ let $\tau(S,T) = \cup_{t\in T}\tau(S,t)$ and let $\mu_T = \sum_{t\in T}\mu_t$. For any set of pairs V let $M^V_{\mathcal{P}^\mathcal{U}}$ be a (random) allocation in V that maximizes the number of matched underdemanded pairs in V.

Definition 6.1. *We say that $c > 0$ is a* regular *size if for every underdemanded type X-Y$\in \mathcal{P}^\mathcal{U}$*

$$E_V[\int |\tau(M^V_{\mathcal{P}^\mathcal{U}}(V), X\text{-}Y)|dF(V)|\#V = c] < \mu_{Y\text{-}X}c, \tag{1}$$

where $F(V)$ is any distribution over all allocations that maximize the number of matched underdemanded pairs in a given set of pairs V.

Using simulations with distributions from clinical data (see Ashlagi et al. (2010c)) we find that hospitals of size up to at least $c = 100$ are regular. This allows us to state our first main result.

Theorem 6.2. *Suppose every hospital size is regular and bounded by some $\bar{c} > 0$ and let $\epsilon > 0$. In almost every large graph $D(H_n)$ there exists an individually rational allocation using exchanges of size at most 3, which is at most $\mu_{AB\text{-}O}m + \epsilon\mu_{A\text{-}B}m$ smaller than the efficient allocation, where m is the number of pairs in the graph.*

As suggested in the theorem statement (and as we will show in the proof) most of the efficiency loss comes from matching of (otherwise unmatched) underdemanded O-AB pairs, in 2-way exchanges to AB-O pairs. This means that the **efficiency loss is only about 1%**, which is the (simulated) frequency of the AB-O pairs. Note also that, as remarked earlier, it is hard to regret this small decrease in the total number of matched pairs, since no O-AB pairs would have been matched had the goal been to maximizes the number of transplants.[24]

Theorem 6.2 is a limit theorem, but Table 1 showed simulation results that demonstrate that the cost of individual rationality is very low even for sizes of exchange pools observed in present-day clinical settings.[25]

The allocation constructed in the proof of Theorem 6.2 can fail to be individually rational with low probability, although it is always close to being efficient. However, by Theorem 6.2 and Proposition 3.2 one can also construct in *every* graph an allocation that is individually rational such that in almost every graph it will be within the indicated efficiency bound [26]

Corollary 6.3. *Suppose every hospital size is regular and bounded by some $\bar{c} > 0$, and let $\epsilon > 0$. The size of a maximum IR allocation is at most $\mu_{AB\text{-}O}m + \epsilon\mu_{A\text{-}B}m$ smaller then the size of an efficient allocation in almost every large graph $D(H_n)$, where m is the number of pairs in the graph.*

6.1 Sketch of proof for Theorem 6.2

The main tool in proving the theorem is an individual rationality lemma that focuses only on underdemanded pairs, and shows that with high probability there exists a *satisfiable* set of underdemanded pairs which can be matched, where a satisfiable set is a set in which (i) for each hospital, the set contains at least the number of underdemanded pairs the hospital can internally match, and (ii) for each underdemanded type X-Y the number of pairs in the set is the same number as the total number of overdemanded Y-X pairs in the entire pool.

[24] We conjecture that the requirement that every hospital size be regular can be relaxed (to a weaker definition of regular) or eliminated entirely.

[25] Independently of our work, Toulis and Parkes (2010) show that if only 2-way exchanges are possible (also internally) then, also using random graphs, there is no efficiency loss at all. Their result follows from Theorems 5.2, 6.2 and their proofs.

[26] In this alternative consruction, one first finds a maximum allocation within each hospital, and then finds a maximum allocation in the entire graph subject to matching all pairs that are matched in the first phase.

We formalize this. For every $h \in H_n$ let V_h be the set of pairs of hospital h. For a hospital $h \in H_n$ and a set of pairs $S \subseteq V_{H_n}$ we denote by $\alpha(S,h) = V_h \cap S$ the set of pairs in S belonging to h. Note that $\tau(M^{V_h}_{\mathcal{P}^{\mathcal{U}}}(V_h), \mathcal{P}^{\mathcal{U}})$ is a maximum set of underdemanded pairs h can internally match. We let $U_{H_n} = \tau(V_{H_n}, \mathcal{P}^{\mathcal{U}})$ and $O_{H_n} = \tau(V_{H_n}, \mathcal{P}^{\mathcal{O}})$ be the set of all underdemanded and overdemanded pairs in H_n respectively.

Definition 6.4. *A set of underdemanded pairs $S \subseteq \tau(V_{H_n}, \mathcal{P}^{\mathcal{U}})$ is called a **satisfiable set** if*

1. $|\alpha(S,h)| \geq |\tau(M^{V_h}_{\mathcal{P}^{\mathcal{U}}}(V_h), \mathcal{P}^{\mathcal{U}})|$ *for all* $h \in H_n$.
2. $|\tau(S, X\text{-}Y)| = |\tau(V_{H_n}, Y\text{-}X)|$ *for all* $X\text{-}Y \in \mathcal{P}^{\mathcal{U}}$.

Note that the first part can be thought of as individual rationality with respect to underdemanded pairs.[27]

Lemma 6.5 (underdemanded rationality lemma)**.** *Suppose every hospital size is regular and bounded by some $\bar{c} > 0$. With probability $1 - o(1)$, there exists a satisfiable set S_n in $D(H_n)$ and a perfect allocation in the bipartite subgraph induced by S_n and $\tau(V_{H_n}, \mathcal{P}^{\mathcal{O}})$.*

The existence of a satisfiable set follows almost directly from the definition of regularity (choose for each hospital a maximum set of underdemanded pairs it can internally match to satisfy condition 1 of a satisfiable set) and from the fact that there are more underdemanded pairs of type X-Y than Y-X pairs (so one can add from each underdemanded type X-Y enough pairs to satisfy condition 2 of a satisfiable set).

However we also need to show that a perfect allocation as described in Lemma 6.5 exists; the problem with the above naive construction is that it adds information about non-existence of internal edges which we wish to avoid in order to be able to use properties of large random graphs with independent edge probabilities. Thus to see that Lemma 6.5 holds, we will construct a satisfiable set S_n with some additional properties; We partition the hospitals into two sets H^1_n and H^2_n each with $\frac{n}{2}$ hospitals, and find a satisfiable set S_n such that (i) the number of underdemanded pairs of each type X-Y in S_n belonging to H^1_n (H^2_n) equals the number of overdemanded pairs Y-X belonging to H^2_n (H^1_n). Then, using Erdos-Renyi type of results, we show that one can match all overdemanded pairs of type Y-X belonging to H^1_n (H^2_n) to X-Y underdemanded pairs in S_n belonging to H^2_n (H^1_n).

We continue with the proof sketch of Theorem 6.2. In the efficient allocation in a random compatibility graph the only pairs that are not matched have underdemanded types. First

[27] Even if a hospital can internally match more pairs using fewer underdemanded pairs, it is reasonable to consider this condition since pairs of other types will be "easy" to match as suggested by Proposition 5.2.

we find an allocation as described in Lemma 6.5, in particular an allocation that matches each overdemanded pair to an underdemanded pairs using 2-way exchanges. Furthermore the selfdemanded pairs can be perfectly matched (using only other selfdemanded pairs) using 2-way or 3-way exchanges.

Finally since with a positive probability every hospital cannot match all its A-B pairs and all its B-A pairs, there will be a linear number of A-B and B-A pairs that cannot be internally matched. Therefore one can find a perfect allocation in the graph induced by these pairs that matches all the the A-B and B-A pairs that the hospitals can internally match.

The efficient allocation would have, in addition, three way matches with AB-O pairs and with the excess A-B or B-A pairs.

7 Kidney Exchange Mechanisms

We have seen that a mechanism that is individually rational for hospitals need not be costly in terms of lost transplants, and individual rationality can be seen as a necessary condition for full participation in a world in which a hospital can withdraw participation after seeing the allocation proposed by the centralized mechanism. But a mechanism that makes it individually rational for hospitals to participate may still not be sufficient to elicit full participation if it does not also make it a dominant strategy, or a Bayesian equilibrium, for hospitals to reveal all their patient-donor pairs. We next begin the exploration of the incentive properties of exchange mechanisms, starting (as in the case of individual rationality) with some negative worst-case results.

A kidney exchange **mechanism**, φ, maps a profile of incompatible pairs $V = (V_1, V_2, \ldots, V_n)$ to an allocation, denoted by $\varphi((V_h)_{h \in H_n})$. A mechanism φ is IR if for every profile V, $\varphi(V)$ is IR. Efficient and maximal mechanisms are defined similarly.

Every kidney exchange mechanism φ induces a game of incomplete information $\Gamma(\varphi)$ in which the players are the hospitals. The type of each hospital h is its set of incompatible pairs. The realized type will be denoted by V_h and at this point we assume no prior over the set of types. At strategy σ_h hospital h reports a subset of its incompatible pairs $\sigma_h(V_h)$. For any strategy profile σ let $\sigma(V) = (\sigma_1(V_1), \ldots, \sigma_n(V_n))$ be the profile of subsets of pairs each hospital submits under σ given V. Therefore, for any profile $V = (V_1, \ldots, V_n)$, at strategy profile σ mechanism φ chooses the allocation $\varphi(\sigma(V))$.

A kidney exchange mechanism does not necessarily match all pairs in $V_{H_n} = \cup_{h \in H_n} V_h$, either because it didn't match all reported pairs or because hospitals did not report all pairs. Therefore we assume that each hospital also chooses an allocation in the set of its pairs that are not matched by the mechanism. Formally, let φ be a kidney exchange mechanism and let σ be a strategy profile and V_h be the type of each hospital. After the mechanism chooses

$\varphi(\sigma(V))$, h finds an allocation in $V_h \setminus \varphi(\sigma(V))(V_h)$, where $\varphi(\sigma(V))(V_h)$ is the set of all pairs in V_h that are matched by the allocation $\varphi(\sigma(V))$. In particular every hospital $h \in H_n$ has an allocation function φ_h that maps any set of pairs X_h to an allocation $\varphi_h(X_h)$.

Since each hospital wishes to maximizes the number of its own matched pairs, the utility of hospital h, u_h, at profile V and strategy profile σ, is defined by the number of pairs in V_h who are matched by the centralized match, plus the number of its remaining pairs that h can match using internal exchanges:

$$u_h(\sigma_h(V_h), \sigma_{-h}(V_{-h})) = |\varphi(\sigma(V))(V_h)| + |\varphi_h(V_h \setminus \varphi(\sigma(V))(V_h))(V_h)|. \quad (2)$$

In the next section we study incentives of hospitals in the games induced by kidney exchange mechanisms.

8 Incentives

Loosely speaking, most of the kidney exchange mechanisms presently employed choose an efficient allocation in the (reported) exchange pool.[28] As already emphasized, maximizing the number (or the weighted number) of transplants in the pool of patient-donor pairs reported by hospitals is not the same as maximizing the number of transplants in the whole pool, unless the whole pool is reported. We next consider the tensions between achieving efficiency, and making reporting of the whole pool a dominant strategy for each hospital.

8.1 Strategyproofness–negative results for compatibility graphs

Section 3 showed that for any largest feasible exchange size $k > 2$, no individually rational mechanism can be efficient, and obtained discouraging worst case bounds (although efficiency can be achieved for $k = 2$). Here we show that for $k \geq 2$, no mechanism that always produces a k-maximal allocation (even if not efficient) can be individually rational and strategyproof, again with discouraging worst case bounds.

A mechanism φ is strategyproof if it makes it a dominant strategy for every hospital to report all of its incompatible pairs in the game $\Gamma(\varphi)$; Formally, φ is **strategyproof** if for every hospital h, every V_h, every strategy σ'_h, and every V_{-h}

$$u_h(\varphi(V_h, V_{-h})) \geq u_h(\varphi(\sigma'_h(V_h), V_{-h})). \quad (3)$$

In an unpublished note Roth et al. (2007a) showed that (even for a maximum exchange size $k = 2$):

[28] The mechanisms often maximize a *weighted* sum of transplants rather than a simple sum, to implement priorities, such as for children and for how difficult it is to match a patient (due to high PRA levels).

Proposition 8.1 (Roth et al. (2007a))**.** *No IR mechanism is both maximal and strategyproof.*

Proof. Consider a setting with two hospitals $H_2 = \{a, b\}$ such that $V_a = \{a_1, a_2, a_3, a_4\}$ and $V_b = \{b_1, b_2, b_3\}$. Further assume the compatibility graph induced by V_{H_2} is given in Figure 6.

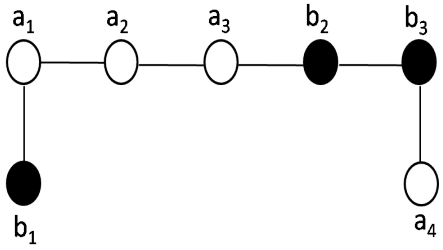

Figure 6

Note that every maximal allocation leaves exactly one node unmatched. Suppose φ is both maximal and IR. We show that if a and b submit V_a and V_b respectively, at least one hospital strictly benefits from withholding a subset of its nodes. Let $v \in V_{H_2}$ be unmatched in $\varphi(V_a, V_b)$. If $v \in V_a$ then $u_a(\varphi(V_a, V_b)) = 3$. However, by withholding a_1 and a_2, a's utility is 4 since the maximal allocation in $V \setminus \{a_1, a_2\}$ matches both a_3 and a_4, and a can match both a_1 and a_2 via an internal exchange. If $v \in V_b$ then by a symmetric argument hospital b would benefit by withholding b_2 and b_3. ∎

Strategyproof mechanisms do exist, e.g. a mechanism that chooses allocations that maximize the number of matched nodes using only internal exchanges. Unfortunately, no such mechanism is close to efficient (again, even for exchange size k=2):

Proposition 8.2. *For $k \geq 2$, no IR strategyproof mechanism can always guarantee more than $\frac{1}{2}$ of the efficient allocation.*

Proof. Consider the same setting as in the proof of Proposition 8.1 (see Figure 6) and suppose φ is an IR strategyproof mechanism which always guarantees more than 1/2 of the efficient allocation. Note that either $u_a(\varphi(V_a, V_b)) \leq 3$ or $u_b(\varphi(V_a, V_b)) \leq 2$. Suppose $u_a(\varphi(V_a, V_b)) \leq 3$. As in the proof of Proposition 8.1, in order for it not to be beneficial for a to withhold a_1 and a_2, the mechanism cannot match all pairs in $\{a_3, a_4\} \cup V_b$. Thus φ can choose at most a single exchange of size 2 in $\{a_3, a_4\} \cup V_b$, which is only half of the maximum (efficient) number, and not more, as required by assumption. The case in which $u_b(\varphi(V_a, V_b)) \leq 2$ is similar. ∎

By allowing randomization between allocations (in particular allowing inefficient allocations to be chosen with positive probability) one can hope to improve efficiency in expectation. Strategyproofness in this case means that, for any reports of other hospitals,

no hospital h is better off in expectation reporting anything other than its type V_h. However, even random mechanisms do not reconcile individual rationality, strategyproofness and efficiency:

Proposition 8.3. *For $k \geq 2$, no IR strategyproof (in expectation) randomized mechanism can always guarantee more than $\frac{7}{8}$ of the efficient allocation.*

Proof. Consider the same setting as in the proof of Proposition 8.1 (see Figure 6) and assume there exists a randomized IR strategyproof mechanism φ that guarantees more than 7/8 of the efficient allocation in every possible V. Any allocation leaves at least one node unmatched. Therefore either $E[u_a(\varphi(V_a, V_b))] \leq 3.5$ or $E[u_b(\varphi(V_a, V_b))] \leq 2.5$. Suppose $E[u_a(\varphi(V_a, V_b))] \leq 3.5$. We argue that under the mechanism φ, hospital a benefits from withholding a_1 and a_2. Since φ guarantees more than 7/8 of the efficient allocation in $\{a_3, a_4, b_1, b_2, b_3\}$, φ will choose the allocation containing exchanges a_3, b_2 and b_3, a_4 with probability more than 3/4. Therefore a's expected utility from reserving 2 transplants to do internally will be $2+c$ for some $c > 1.5$. A similar argument holds if $E[u_b(\varphi(V_a, V_b))] \leq 2.5$ ∎

Ashlagi et al. (2010b) study dominant strategy mechanisms for $k = 2$ and provide a strategyproof (in expectation) randomized mechanism which guarantees 0.5 of the 2-efficient allocation.[29] But it remains an open question whether the bounds established in this section can be achieved.

Strategyproofness is independent of any probability distribution of the underlying compatibility graphs. However, in the case of compatibility of kidneys, we know a lot about the (approximate) distribution of compatibility graphs that might be useful for finding mechanisms that can achieve (almost) efficient allocations as Bayesian equilibria.[30] We proceed by studying the Bayesian setting in a large random kidney exchange program, in the spirit of recent advances in the study of two sided matching in large markets (cf. Immorlica and Mahdian (2005) , Kojima and Pathak (2009), Kojima et al. (2010), and Ashlagi et al. (2010a)).

[29] The model in Ashlagi et al. (2010b) does not allow hospitals to choose an internal allocation after the mechanism has chosen an allocation. However their algorithm works in our model. Specifically their mechanism randomly partitions hospitals into two sets and chooses randomly an allocation with maximum number of matched nodes among allocations that satisfy (i) there are no edges between the nodes of two hospitals within each set, and (ii) are 2-efficient within each hospital.

[30] An efficiency approximation gap between the Bayesian approach and prior free approach has been shown for example by Babaioff et al. (2010) in an online supply problem.

8.2 The Bayesian setting

To study hospitals' incentives in a given mechanism we consider a Bayesian game in which hospitals strategically report a subset of their set of incompatible pairs, and the mechanism chooses an allocation. Thus a **kidney exchange game** is now a Bayesian game $\Gamma(\varphi) = (H, (T_h)_{h\in H}, (u_h)_{h\in H})$ where H is the set of hospitals, u_h is the utility function for hospital h, and T_h is the set of possible private types for each hospital, drawn independently from a known distribution. The type for each hospital is the subgraph induced by its pairs in the random compatibility graph, i.e. after the graph is generated, each hospital observes its own subgraph.

The expected utility for hospital h at strategy profile σ given V_h is

$$E_{V_{-h}}[u_h(\varphi(\sigma_h(V_h), \sigma_{-h}(V_{-h}))]. \tag{4}$$

Let σ be a strategy profile and let $\epsilon > 0$. Strategy σ_h is an ϵ-best response against σ_{-h} if for every σ'_h and every V_h

$$E_{V_{-h}}[u(\varphi(\sigma_h(V_h), \sigma_{-h}(V_{-h}))] \geq E_{V_{-h}}[u(\varphi(\sigma'_h(V_h), \sigma_{-h}(V_{-h}))] - \epsilon. \tag{5}$$

σ is an ϵ-Bayes Nash equilibrium if for every hospital h, σ_h is an ϵ best response against σ_{-h}. For $\epsilon = 0$, σ is the standard Bayes Nash equilibrium.

A particular strategy which will interest us is the **truth-telling** strategy: a hospital always reports its entire set of incompatible pairs. To analyze mechanisms for large random exchange pools, it will be useful to consider a sequence of random kidney exchange games $(\Gamma^1(\varphi), \Gamma^2(\varphi), \ldots)$, where $\Gamma^n(\varphi) = (H_n, (T_h)_{h\in H_n}, (u_h)_{h\in H_n})$ denotes a random kidney exchange game with $|H_n| = n$ hospitals.

8.2.1 The status quo

A stylized version of current kidney exchange mechanisms is the following:
Maximum Transplants mechanism (MT): *for any set of incompatible pairs V choose uniformly at random an efficient allocation in V.*

In Section 1.1. we observed that mechanisms that choose weighted maximum allocations can violate individually rationality. Here we observe that withholding pairs in the MT mechanism can provide a non-negligible benefit to a hospital even in a large random graph.

Proposition 8.4. *In the sequence of games $\Gamma^1(MT), \Gamma^2(MT), \ldots, \Gamma^n(MT), \ldots...$ there exist no $\epsilon(n) = o(1)$ such that reporting truthfully is an $\epsilon(n)$-Bayes Nash equilibrium in $\Gamma^n(MT)$.*

Proof. Suppose that all hospitals truthfully report all their pairs. It is sufficient to provide an example of a compatibility graph for some hospital a, such that a is better off not reporting truthfully. Let a be an arbitrary hospital and let $V_a = \{a_1, a_2\}$ where a_1 and a_2 are an O-B pair and a B-O pair respectively and a_1, a_2 is an internal exchange. For sufficiently large n, by Proposition 5.2 in any efficient allocation the set of reported pairs by all hospitals satisfies the following: in almost every efficient allocation in $D(H_n)$ a constant fraction of the O-B pairs will not be matched. Therefore since by definition of MT the O-B pairs that are not matched are chosen randomly, the probability that any O-B pair will not be matched is at least some constant probability $q > 0$ (not depending on n). Therefore if hospital a reports both pairs a_1 and a_2, the expected utility for a is $2 - q$ and by not reporting both pairs h obtains a utility of 2. ∎

Essentially the existence of any reciprocally overdemanded and underdemanded types in the system drives Proposition 8.4.[31]

We simulated the MT mechanism and examined two types of behavior for hospitals: *truth-telling*, in which a hospital reports all of its incompatible pairs to the mechanism, and a naive strategy called *withhold internal matches*, in which a hospital withholds a maximum set of pairs it can match internally. As depicted in Figure 7, withholding provides more transplants on average than truth-telling for an arbitrary hospital given that all other hospitals are truth-telling. The benefit from withholding becomes even higher when all other hospitals also withhold internal matches (see Figure 7).

Following these findings we compared the efficiency achieved when hospitals use the withhold internal matches strategy, to the efficiency achieved when hospitals report truthfully to the MT (non IR) mechanism. The efficiency loss is about 10% in both $k = 3$ and $k = 2$ (see Table 2).

8.2.2 Individual rationality is not sufficient

Following the results of the previous sections an important open question is how to design a kidney exchange mechanism that minimizes the efficiency loss in equilibrium.

For any set of pair types $T\mathcal{P}^{\mathcal{U}}$ let $\mathcal{M}_T^V$ be the set of allocations in V that match the maximum number of pairs in V whose type belong to T. Following Theorem 6.2 one natural candidate for a mechanism is to randomly choose a maximum IR allocation.

[31] The MT mechanism might further deepen this issue: consider hospitals that withhold internal unbalanced 3-way exchanges. Since the expected number of each of the two unbalanced exchanges is different, either more A-B or more B-A pairs will be withheld by hospitals. If this difference is "large", one of these pair types in fact will play the role of a new overdemanded type. For efficiency, one wishes to overcome imbalances rather than create new ones.

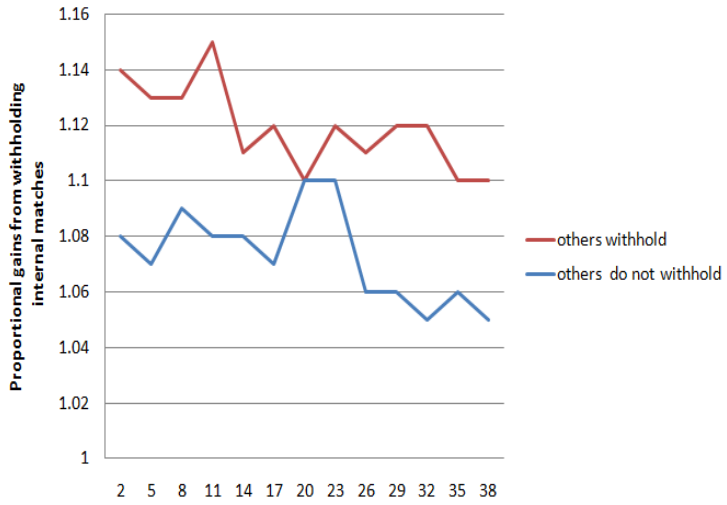

Figure 7: Withholding internal matches vs. reporting truthfully in MT mechanism (k=3).

		k=2		k=3		
No. of hospitals	No. of Pairs	Withholding Strategy	IR	Withholding Strategy	IR	Efficient
12	131	55.84	60.6	70.15	81.43	81.83
14	154	68.64	74.72	85.44	97.82	98.07
16	173	77.44	84.2	96.57	109.01	109.41
18	191	87.84	95.62	109.76	121.81	122.1
20	227	107.74	116.68	132.32	144.09	144.35

Table 2: Number of transplants achieved under two different strategies: (i) each hospital withholds an efficient internal allocation and (ii) each hospital reports reports truthfully.

RandomIR mechanism:

Input: *a profile of incompatible pairs* $(B_1, B_2, \ldots, B_n)$.

Step 1: *randomly choose a maximum allocation* $M_h \in \mathcal{M}_{\mathcal{P}}^{B_h}$ *for every hospital* $h \in H_n$.

Step 2: *choose a maximum allocation in* $B_{H_n} = \cup_{h \in H_n} B_h$ *that matches all pairs in* $\cup_{h \in H_n} M_h(B_h)$.

The RandomIR mechanism is by construction individually rational, and by a similar result to Corollary 6.3 the efficiency loss is small.

However, RandomIR will quickly run into incentive problems for the same reason as the MT mechanism: Consider a hospital a which has the compatibility graph on the left side of Figure 8, i.e. the efficient internal allocation for a uses the 3-way allocation A-O,A-A,A-A. By similar arguments as in the proof of Proposition 8.4 hospital a will be better off withholding the A-O,O-A internal match, since its O-A pair is not as likely to be matched by the central exchange as are both A-A pairs.

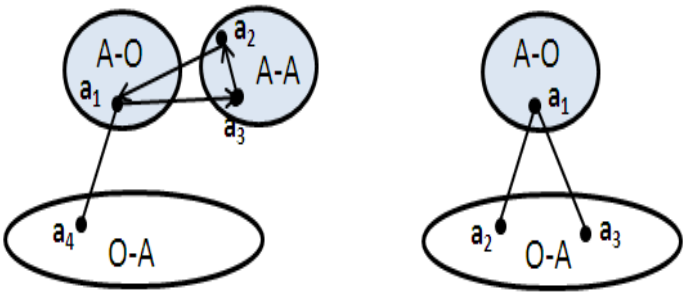

Figure 8: On the left side, hospital a has a unique efficient allocation of size 3, but can also internally match the A-O and O-A pairs. On the right side hospital a has one overdemanded A-O pair and two O-A pairs and it can internally match either one.

As underdemanded pairs are the ones that compete for being matched, a natural attempt to solve this problem is to change Step 1 in the RandomIR mechanism so that for each hospital h it will choose from ${\mathcal{M}_{\mathcal{P}^{\mathcal{U}}}}^{B_h}$ rather than from $\mathcal{M}_{\mathcal{P}}^{B_h}$:
(New) Step 1: randomly choose a maximum allocation $M_h \in {\mathcal{M}_{\mathcal{P}^{\mathcal{U}}}}^{B_h}$ *for every hospital* $h \in H_n$.

Unfortunately truth-telling is still not a Bayes-Nash equilibrium in the game induced by the (new) RandomIR mechanism. To see this suppose all hospitals but a report truthfully and suppose a has the compatibility graph on the right side of Figure 8. We argue that hospital a is better off not revealing the overdemanded pair a_1. If hospital a withholds a_1 there is some probability p that each of a_2 and a_3 will be matched. Therefore if both pairs a_2 and a_3 are matched or if neither is matched by the mechanism, hospital a will end up with two pairs matched. If, however, exactly one of these pairs, say a_2, is matched by the mechanism, a will end with 3 matches since it can still match a_1 to a_3. Hence, a's utility is

$$2p^2 + 2(1-p)^2 + 3 * 2(1-p)p. \tag{6}$$

By reporting truthfully, the mechanism guarantees to match a_1 and one of a_2 or a_3, while the remaining pair will be matched with probability $p + o(1)$ (since the market is large the probability remains "very close" to p). Therefore a's expected utility is $2 + p + o(1)$. A simple calculation shows that for any $0 < p < \frac{1}{2}$, a is better off withholding a_1. (Loosely speaking since $\bar{\gamma} < \frac{1}{2}$, there are at least twice as many O-A pairs as A-O pairs, implying that the probability that any particular O-A pair that is not guaranteed to be matched will be chosen to be matched by the mechanism is likely to be $p < \frac{1}{2}$.)

Notice that in this case hospital a has an incentive to withhold a single overdemanded pair rather than an internal exchange. To prevent hospitals from wishing to withhold overdemanded pairs as in this example, a mechanism will likely have to give priority to underdemanded pairs from hospitals that contribute overdemanded pairs. We discuss this next.

8.2.3 Towards a new mechanism

So far we have seen natural mechanisms that are either manipulable by withholding an internal exchange that contains an overdemanded pair or by withholding an overdemanded pair by itself. Either way, the main obstacle in designing a Bayesian incentive compatible mechanism is preventing hospitals from withholding their overdemanded pairs.

In this section we present a mechanism for kidney exchange which makes the truth-telling strategy profile an approximate Bayes-Nash equilibrium, assuming that hospitals satisfy a stronger regularity condition. This stronger regularity condition, which now deals with each underdemanded type and its reciprocal overdemanded type separately, will allow us to separate the reporting problem for each type of overdemanded pair. This will allow a mechanism in which there is no incentive to withhold an overdemanded pair of some type in order to influence the match probability of an underdemanded pair that is not of its reciprocal type.

Recall that $\mathcal{M}_T^V$ is the set of allocations in V that maximize the number of matched pairs in V whose type belongs to T. If $T = \{t\}$ is a singleton we will just write $\mathcal{M}_t^V$.

Definition 8.5. *We say that $c > 0$ is a* strongly regular *size if for every underdemanded type X-Y$\in \mathcal{P}^{\mathcal{U}}$*

$$E_V[\#\tau(M_{X\text{-}Y}^V(V), X\text{-}Y)|\#V = c] < \mu_{Y\text{-}X}c, \tag{7}$$

where $M_{X\text{-}Y}^V$ is an arbitrary allocation in $\mathcal{M}_{X\text{-}Y}^V$.

Using simulations we find that hospitals of size up to at least $c = 30$ are strongly regular. Throughout this section we assume that every hospital's size is strongly regular and bounded. The mechanism we introduce provides an allocation that uses only 2-way exchanges with similar properties to the one constructed in the proof of Theorem 6.2: (i) overdemanded X-Y pairs are matched in 2-way exchange to Y-X pairs, (ii) A-B and B-A pairs will be matched to each other in 2-way exchanges, (iii) and selfdemanded pairs will be matched to each other.

The key to the mechanism lies in how property (i) is implemented, i.e. in how the mechanism will choose for each underdemanded type which pairs will be matched. In particular we will use a lottery, called the *underdemanded lottery*, to determine for each underdemanded type X-Y$\in \mathcal{P}^{\mathcal{U}}$ a *set* of X-Y pairs, denoted by S_h(X-Y), that will be matched for each hospital h (ideally we want to match all overdemanded pairs, so the total number of X-Y pairs that will be matched equals the total number of Y-X pairs in the pool).

We describe the underdemanded lottery for a given underdemanded type X-Y$\in \mathcal{P}^{\mathcal{U}}$. For each h, S_h(X-Y) will be initialized to be a set of X-Y pairs with maximum cardinality that h can match internally, and the lottery will output for each hospital a set of pairs that are chosen to be matched.

Underdemanded lottery

Input: *a set of hospitals H_n, a profile of subsets of pairs $(B_1, \ldots, B_n)$, an underdemanded type X-Y, and an integer $0 < \theta < |\tau(B_{H_n}, X\text{-}Y)|$ which is interpreted as the number of X-Y pairs that we want to choose in total.*[32]

Initialization: *For each hospital h let $Q_h(X\text{-}Y) = |\tau(B_h, X\text{-}Y)|$ and let $S_h(X\text{-}Y)$ be an arbitrary maximum set of X-Y pairs h can internally match in B_h.*[33]

Main Step: *Let J be a bucket containing $Q_h(X\text{-}Y)$ balls for each hospital h. As long as $\sum_{h \in H_n} |S_h(X\text{-}Y)| < \theta$:*

1. *Choose uniformly at random a ball from J without replacement. If the ball belongs to hospital h, then add an arbitrary X-Y pair to $S_h(X\text{-}Y)$ from $B_h \setminus S_h(X\text{-}Y)$ if such exists.*

The important change from RandomIR is the Main Step in the underdemanded lottery; consider the graph of hospital a in the right hand side of Figure 8. If a_2 was chosen to be matched in the (new) Step 1 of RandomIR, then in Step 2 only remaining pairs are considered to be chosen from. In the underdemanded lottery, S_a(X-Y) is initialized to either $\{a_2\}$ or $\{a_3\}$, assume that S_a(X-Y) $= \{a_2\}$ without loss of generality. However, *two* balls are initially placed in the bucket J for hospital a, and if either one of them is drawn, a_3 is added to S_a(X-Y). That is, the fact that the hospital can internally match one of its underdemanded pairs now increases the probability that another of its underdemanded pairs of the same type will be matched.

We are now ready to state the mechanism. For ease of exposition we assume throughout this section that n is even.

The Bonus Mechanism

Input: *a set of hospitals $H_n = \{1, \ldots, n\}$ and a profile of incompatible pairs $(B_1, B_2, \ldots, B_n)$, each of a strongly regular size.*

Step 1 [Match selfdemanded pairs]: *find a maximum allocation, M_S in the graph induced by all selfdemanded pairs B_{H_n}.*

Step 2 [Match A-B and B-A pairs]: *for each hospital h choose randomly an allocation $M_h \in \mathcal{M}_{\mathcal{P}^{\mathcal{R}}}{}^{B_h}$.*[34] *Find a maximum allocation,M_R in the graph induced by A-B and B-A pairs among those that maximize the number of matched pairs in $\cup_{h \in H_n} \tau(M_h(B_h), \mathcal{P}^{\mathcal{R}})$.*

Step 3 [Match overdemanded and underdemanded pairs]: *Partition the set of hospitals into two sets $H_n^1 = \{1, \ldots, \frac{n}{2}\}$ and $H_n^2 = \{\frac{n}{2} + 1, \ldots, n\}$. For each underdemanded type X-Y$\in \mathcal{P}^{\mathcal{U}}$*

[32]The parameter θ is not set here to be the number of Y-X pairs in B_{H_n} since the mechanism will apply the lottery twice, each time with a different set of $\frac{n}{2}$ hospitals and θ will be the number of Y-X pairs in the set of other $\frac{n}{2}$ hospitals. This will be further discussed below.

[33]Formally S_h(X-Y) $= \tau(M^{B_h}_{\text{X-Y}}(B_h), \text{X-Y})$ for some $M^{B_h}_{\text{X-Y}} \in \mathcal{M}^{B_h}_{\text{X-Y}}$.

[34]Recall that $\mathcal{P}^R = \{\text{A-B}, \text{B-A}\}$.

and for each $j = 1, 2$.[35]

(3a) *Set* $\theta_j(Y\text{-}X) = |\tau(B_{H_{n^{3-j}}}, Y\text{-}X)|$ *to be the number of Y-X pairs in the set* $B_{H_{n^{3-j}}}$. *Then, using the underdemanded lottery procedure with the inputs* $(B_h)_{h \in H_{n^j}}$, $\theta_j(Y\text{-}X)$ *and X-Y, construct a subset* $S_h(X\text{-}Y)$ *one for each hospital in* $h \in H_{n^j}$.

(3b) *Find a maximum allocation* $M^j_{X\text{-}Y}$ *in the subgraph induced by the sets of pairs* $\cup_{h \in H^j_n} S_h(X\text{-}Y)$ *and* $\tau(B_{H^{3-j}_n}, Y\text{-}X)$.[36]

Step 4 [Output]: *Let* $M_U = \cup_{j=1,2} \cup_{X\text{-}Y \in \mathcal{P}^{\mathcal{U}}} M^j_{X\text{-}Y}$. *Output* $M_S \cup M_R \cup M_U$.

We can now state our second main result.

Theorem 8.6. *Let* H_n *be a set of hospitals. If every hospital size is strongly regular, the truth-telling strategy profile is an* $\epsilon(n)$*-Bayes-Nash equilibirum in the game induced by the Bonus mechanism, where* $\epsilon(n) = o(1)$. *Furthermore for any* $\epsilon > 0$, *the efficiency loss under the truth-telling strategy profile in almost every* $D(H_n)$ *is at most* $\mu_{AB\text{-}O} m + \epsilon \mu_{A-B} m$, *where* m *is the number of pairs in the pool.*

We conjecture that, here too, the strong regularity assumption can be relaxed and even entirely eliminated:

Conjecture: *There exists a mechanism* φ *such that the truth-telling strategy profile is an* $\epsilon(n)$*-Bayes-Nash equilibrium for* $\epsilon(n) = o(1)$ *in the game induced by* φ. *Furthermore under the truth-telling strategy the efficiency loss is at most* $\mu_{AB\text{-}O} m + \epsilon \mu_{A-B} m$ *for any* $\epsilon > 0$, *where* m *is the number of pairs in the pool.*

9 Conclusions and open questions

Kidney exchange in the United States has grown from being carried out rarely in only a few hospitals, to being carried out regularly in a variety of kidney exchange networks of hospitals, and it is presently being explored at the national level.

The National Kidney Paired Donation Pilot Program was approved by the OPTN/UNOS Board of Directors in June 2008, and ran its first two match runs in October and December

[35] In order to choose the sets of underdemanded pairs of each type that will be matched we partition the set of hospitals into two sets, each with $\frac{n}{2}$ hospitals; For each set in the partition we will match the overdemanded pairs of the hospitals in one set to the chosen underdemanded pairs of the hospitals in the other set in order to avoid lack of independence (see also Section 6.1) and the proof of Theorem 6.3 for further discussion.

[36] The size of $|\cup_{h \in H^j_n} S_h(\text{X-Y})|$ will equal $|\tau(B_{H^{3-j}_n}, \text{Y-X})|$ with high probability and therefore the maximum allocation in this subgraph will match with high probability all pairs in $\cup_{h \in H^j_n} S_h(\text{X-Y})$.

2010, with 43 patient-donor pairs in October and 62 in December, registered by kidney exchange consortia representing 77 transplant programs. For the purposes of the present paper it is notable that only a small fraction of the patient-donor pairs registered in the participating hospitals were enrolled in the national pilot program.[37] So the problem of full participation by hospitals is both real and timely. It has also begun to be observed in the active kidney exchange networks that are fully operational.

The present paper observes that one contributory cause of the lack of full participation is that the matching algorithms currently employed in practice do not make it individually rational for hospitals to always contribute all their patient-donor pairs. We show that, in worst cases, this could be very costly, but we prove that in large markets it is possible to redesign the matching mechanisms to guarantee individually rational allocations to hospitals, at very modest cost in terms of "lost" transplants. Note that these "lost" transplants are not really lost if instead hospitals would have withheld patient-donor pairs; on the contrary, we show that individually rational allocations produce a big gain in transplants compared to having hospitals withhold pairs.

To obtain analytical results about large markets we approximate them as large random graphs whose properties we can study with limit theorems based on the classical results of Erdos and Renyi. But we also show by simulation with clinically relevant distributions of patients and donors that these main results apply on the scale of exchange we are presently seeing.

However the highly interconnected compatibility graphs that we see in the limit theorems are far from perfect approximations of the much sparser compatibility graphs we see in practice, and this is especially true for the very most highly sensitized patients. This raises a number of open questions that are likely to arise in practice.

The first of these questions is how to model the situation facing highly sensitized patients, who will be only sparsely connected in the compatibility graph, because they may be compatible with a very small number of donors, even in a large graph of finite size. This is closely related to the second question, which is how to effectively integrate nondirected donors and chains with the cyclic exchanges that have been used initially in the national pilot program and that are the subject of the present paper. In addition to cycles of length k, there has been growing use of various kinds of chains in kidney exchange, and it remains an open question how the relative importance of chains and cyclic exchanges will change as

[37]We hasten to note that there are many reasons other than the incentive problems discussed here that contribute to this initial very low participation rate. These include the new bureaucratic procedures for enrolling patients, the novelty and lack of track record of the national program, the desire to start small and see what happens, the exclusion of chains and nondirected donors, etc. See http://optn.transplant.hrsa.gov/resources/KPDPP.asp

the size of the pool (and the number of non-directed donors) grow large. It seems likely that, even in large markets, chains will be especially helpful to the most highly sensitized patients (cf. Ashlagi et al. 2010c).

A third open question is under what conditions individually rational and incentive compatible mechanisms exist that are as efficient as we have shown them to be under regularity conditions on the size of hospitals. We conjecture that these regularity conditions can be relaxed. In any case, such mechanisms could be useful in eliciting full participation in a full scale national exchange, as it appears from simulations that hospitals are in fact of regular size (although the largest hospitals may not be strongly regular). However, our results suggest that the benefits of a national exchange could also be realized if there was sufficient regulatory power to require transplant centers to either participate fully or not at all, since that would reduce the strategy space so that individual rationality would be the primary consideration.

The final open question we raise here is how these strategic concerns would be different in a world in which the players are not only hospitals and a (single) centralized exchange, but in which there are multiple kidney exchange networks, some with strategic concerns of their own. This is, of course, the situation that is currently in place.

In conclusion, as kidney exchange has grown, the strategy sets, the strategic players, and hence the incentive constraints have changed. The new incentive issues, concerning full participation by hospitals, arise out of the growth of kidney exchange, and are potential obstacles to further growth. However the results of this paper strongly suggest that these new barriers can also be overcome.

References

D. J. Abraham, A. Blum, and T. Sandholm. Clearing Algorithms for Barter Exchange Markets: Enabling Nationwide Kidney Exchanges. In *Proceedings of the 8th ACM Conference on Electronic commerce (EC)*, pages 295–304, 2007.

I. Ashlagi, M. Braverman, and A. Hassidim. Matching with Couples Revisited. Working paper, 2010a.

I. Ashlagi, F. Fischer, I. Kash, and A.D. Procaccia. Mix and Match. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC), 2010*, pages 295–304, 2010b.

I. Ashlagi, D. S. Gilchrist, A. E. Roth, and M. A. Rees. Nonsimultaneous Chains and Dominos in Kidney Paired Donation - Revisited. Unpublished, 2010c.

M. Babaioff, L. Blumrosen, and A. Roth. Auctions with Online Supply. In *Proceedings of the 11th ACM Conference on Electronic commerce (EC)*, 2010.

P. Biro, D.F. Manlove, and R. Rizzi. Maximum weight cycle packing in directed graphs, with application to kidney exchange programs. *Discrete Mathematics, Algorithms and Applications*, 1(4):499–517, 2009.

J. Edmonds. Paths, Trees, and Flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.

J. Edmonds, D.W. Gjertson, and J. M. Cecka. Matroids and the greedy algorithm. *Programming*, 1:127–136, 1971.

N. Immorlica and M. Mahdian. Marriage, Honesty, and Stability. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 53–62, 2005.

S. Janson, T. Luczak, and R. Rucinski. *Random Graphs*. A John Wiley & Sons, 2000.

W. S. Jevons. *Money and the Mechanism of Exchange*. New York: D. Appleton and Company., 1876. http://www.econlib.org/library/ YPDBooks/Jevons/jvnMME.html.

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

F. Kojima, P. Pathak, and A. E. Roth. Matching with Couples: Stability and Incentives in Large Markets. Working paper, 2010.

S. Lieder and A. E. Roth. Kidneys for sale: Who disapproves, and why? *American Journal of Transplantation*, 10:1221–1227, 2010.

M. A. Rees, J. E. Kopke, R. P. Pelletier, D. L. Segev, M. E. Rutter, A. J. Fabrega, J. Rogers, O. G. Pankewycz, J. Hiller, A. E. Roth, T. Sandholm, M. U. Ünver, and R. A. Montgomery. A non-simultaneous extended altruistic donor chain. *New England Journal of Medecine*, 360:1096–1101, 2009.

A. E. Roth. Repugnance as a constraint on market. *Journal of Economic Perspectives*, 21(3): 37–58, 2007.

A. E. Roth. What have we learned from market design? *Hahn Lecture, Economic Journal*, 118:285–310, 2008.

A. E. Roth, T. Sönmez, and M. U. Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119:457–488, 2004.

A. E. Roth, T. Sönmez, and M. U. Ünver. A kidney exchange clearinghouse in New England. *American Economic Review Papers and Proceedings*, 95(2):376–380, 2005a.

A. E. Roth, T. Sönmez, and M. U. Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125:151–188, 2005b.

A. E. Roth, T. Sönmez, and M. U. Ünver. Notes on Forming large markets from small ones: Participation Incentives in Multi-Center Kidney Exchange. Unpublished, 2007a.

A. E. Roth, T. Sönmez, and M. U. Ünver. Efficient kidney exchange: coincidence of wants in markets with compatibility-based preferences. *American Economic Review*, 97:828–851, 2007b.

S. L. Saidman, A. E. Roth, T. Sönmez, M. U. Ünver, and F. L. Delmonico. Increasing the Opportunity of Live Kidney Donation by Matching for Two and Three Way Exchanges. *Transplantation*, 81:773–782, 2006.

P. I. Terasaki, D.W. Gjertson, and J. M. Cecka. Paired kidney exchange is not a solution to ABO incompatibility. *Transplantation*, 65:291, 1998.

P. Toulis and David Parkes. A Random Graph Model of Kidney Exchanges : Optimality and Incentives. Working paper, 2010.

M. U. Ünver. Dynamic Kidney Exchange. *Review of Economic Studies*, 77(1):372–414, 2010.

S. Zenios, E. S. Woodle, and L. F. Ross. Primum non nocere: Avoiding harm to vulnerable candidates in an indirect kidney exchange. *Transplantation*, 72:648–654, 2001.

Appendix A

9.1 Preliminaries

9.1.1 Useful Bounds

Lemma 9.1 (Chernoff and Heoffding bounds (see e.g. Alon and Spencer (2008))**.** *Let $X_1, X_2, \ldots, X_n$ be independent bernoulli random trials with $\Pr(X_i = 1) = p$ for every $i = 1, \ldots, n$ and let $X = \sum_{i=1}^{n} X_i$.*

(i) For any $\delta \in (0, 1]$

$$\Pr\left(X < (1-\delta)pn\right) < e^{\frac{-np\delta^2}{2}}. \tag{8}$$

(ii) For any $\delta < 2e - 1$

$$\Pr\left(X > (1+\delta)pn\right) < e^{\frac{-np\delta^2}{4}}. \tag{9}$$

(iii) For any $\delta > 0$

$$\Pr\left(|X - E[X]| \geq \delta\right) < 2e^{\frac{-2\delta^2}{n}}. \tag{10}$$

9.1.2 Proof of Lemma 5.1

For each pair type X-Y let $Z_{\text{X-Y}}(m)$ be the random variable that indicates the number of X-Y pairs in $D(m)$.

Claim 9.2. *Let* $0 < \delta < 1$ *and* $D(m)$ *be a random compatibility graph and consider the following event:*

$$B_\delta(m) = \{\forall X\text{-}Y \in \mathcal{P},\ (1-\delta)\mu_{X\text{-}Y}m < Z_{X\text{-}Y}(m) < (1+\delta)\mu_{X\text{-}Y}m\}. \tag{11}$$

Then $\Pr\left[B_\delta(m)\right] = 1 - o(m^{-1})$.

Proof. Let $D(m)$ be a random compatibility graph and let $\delta > 0$. By the first two parts of Lemma 9.1, for every type X-Y

$$\Pr\left[Z_{\text{X-Y}}(m) \notin ((1-\delta)m\mu_{\text{X-Y}}, (1+\delta)m\mu_{\text{X-Y}})\right] < e^{\frac{-m\mu_{\text{X-Y}}\delta^2}{4}} + e^{\frac{-m\mu_{\text{X-Y}}\delta^2}{2}} = o(m^{-1}).$$

Therefore

$$\Pr\left[B_\delta(m)\right] = 1 - \Pr\left[\text{for some X-Y} \in \mathcal{P} : Z_{\text{X-Y}}(m) \notin ((1-\delta)m\mu_{\text{X-Y}}, (1+\delta)m\mu_{\text{X-Y}})\right] \geq$$
$$1 - \sum_{\text{X-Y}\in\mathcal{P}} \Pr\left[Z_{\text{X-Y}}(m) \notin ((1-\delta)m\mu_{\text{X-Y}}, (1+\delta)m\mu_{\text{X-Y}})\right] = 1 - o(m^{-1}),$$

where the last inequality follows since there are a finite number of pair types. ∎

Claim 9.3. *Let* $0 < \delta < \frac{1}{2}$ *and let* $D(m)$ *be a random compatibility graph and consider the following event:*

$$S_\delta(m) = \{|Z_{A\text{-}B}(m) - Z_{B\text{-}A}(m)| < m^{\frac{1}{2}+\delta}\}. \tag{12}$$

Then $\Pr\left[S_\delta(m)\right] = 1 - o(m^{-1})$.

Proof. By Hoeffding's bound (part three of Lemma 9.1),

$$\Pr\left(Z_{\text{A-B}}(m) \geq E[Z_{\text{A-B}}(m)] + m^{\frac{1}{2}+\delta}\right) \leq e^{-\frac{m^{2\delta}}{2}}.$$

Applying the same argument for B-A pairs we obtain the result. ∎

Proof of Lemma 5.1: Let $S_\delta(m)$ and $B_\delta(m)$ be as in Claims 9.3 and 9.2. By these claims we obtain that the probability that either $S_\delta(m)$ or $B_\delta(m)$ do not hold is $o(m^{-1})$. □

9.1.3 Bounded directed random graphs - definitions and Erdos-Renyi extensions

In a random compatibility graph the number of pairs of each type is not fixed. We will need Erdos-Renyi type results for random graphs in which the number of nodes as well as the number of edges are random.

We start by defining a vector that will represent bounds on the number of nodes of each pair type in a given subset of the compatibility graphs. For example, to represent the subgraph induced by all A-O pairs and all O-A pairs, by Lemma 5.1 and the event $B_\delta(m)$ we can use the vector $((1-\delta)\mu_{\text{A-O}}, (1+\delta)\mu_{\text{A-O}}, (1-\delta)\mu_{\text{O-A}}, (1+\delta)\mu_{\text{O-A}})$ for some $\delta < 1$; in particular this vector is a tuple of coefficients for bounding from below and above the number of A-O pairs and the number of O-A pairs in this subgraph. subgraph.

For any $r > 0$ a *quasi-ordered* vector is a vector $\bar{\alpha}_r = (\alpha_{0,1}, \alpha_{0,2}, \alpha_{1,1}, \alpha_{1,2}, \ldots, \alpha_{r-1,1}, \alpha_{r-1,2})$ where $\alpha_{j,1} \leq \alpha_{j,2}$ for all $0 \leq j < r$, and $\alpha_{0,1} \leq \alpha_{1,1} \leq \ldots \leq, \alpha_{r-1,1}$.[38]

The vector $\bar{\alpha}_r$ is called *feasible* if at most one pair type could have zero number of nodes, that is $\alpha_{0,2} > 0$ and for every $j \geq 1$, $\alpha_{j,1} > 0$. Let $\bar{\alpha}_r$ be a feasible vector. We say that a tuple of r sets of nodes $(W_0, \ldots, W_{r-1})$ are $(\bar{\alpha}_r, m)$-feasible if for each $0 \leq j < r$ the interval $[\alpha_{j,1}m, \alpha_{j,2}m]$ contains at least one integer and if the sizes of these sets are drawn from some distribution over all possible r-tuples of integers that belong to $[\alpha_{0,1}m, \alpha_{0,2}m] \times \cdots \times [\alpha_{r-1,1}m, \alpha_{r-1,2}m]$. Note that for every sufficient large m, the interval $[\alpha_{j,1}m, \alpha_{j,2}m]$ contains at least one integer if an only if $\alpha_{j,1} < \alpha_{j,2}$ or $\alpha_{j,1} = \alpha_{j,2}$ is an integer.

Definition 9.4 (Bounded Directed Random Graphs)**.** *A graph is called a* ***bounded directed random graph****, denoted by $D(\bar{\alpha}_1, m, p)$, if it is generated as follows. A $(\bar{\alpha}_1, m)$-feasible set of nodes is generated and between each two nodes v, w a directed edge is generated from v to w with probability at least p.*[39]

A graph is called a ***r-bounded directed random graph****, denoted by $D(\bar{\alpha}_r, m, p)$, if it is generated as follows: first $r \geq 2$ distinct sets of nodes $W_0, W_1, \ldots, W_{r-1}$ which are $(\bar{\alpha}_r, m)$-feasible are generated. Then for each $i = 0, 1, \ldots, r-1$, and for each two nodes $v \in W_i, w \in W_{i+1}$ (i is taken modulo r) a directed edge is generated from v to w with probability at least p.*

The definition of a bipartite graph can naturally be extended to a r-partite graph which contains r sets of nodes each of size exactly m and edges are generated as in Definition 9.4. Whenever there is no confusion we will refer also to a r-bounded directed random graph by a r-partite graph. Note that in any r-partite graph only exchanges of size $k = qr$ for positive integers q are feasible. (When $r = 1$ we think about subgraphs induced by selfdemanded pairs

[38] The vector is called quasi-ordered since only the lower bounds are ordered.

[39] Note that for $\alpha_{0,1} = \alpha_{0,2} = 1$ the number of nodes is m.

of some given type. When $r = 2$ we think about subgraphs with potential 2-way exchanges such as O-A,A-O, and when $r = 3$ we think about subgraphs with potential 3-way exchanges such as AB-O,O-A,A-AB).

Lemma 9.5. *Let* $0 < p < 1$.

1. *For any feasible vector* $\bar{\alpha_1}$*, almost every large* $D(\bar{\alpha}_1, m, p)$ *has a nearly perfect allocation using exchanges of size 2 (i.e. an allocation that matches all nodes but at most one), and a perfect allocation for any* $k \geq 3$ *(i.e. an allocation that matches all nodes).*

2. *Let* $\bar{\alpha}_r$ *be a feasible vector with* $r > 1$*. Almost every large* $D(\bar{\alpha}_r, m, p)$ *contains a perfect allocation, i.e. an allocation that matches all nodes in some set* W_i*. Consequently, if* $j' \leq r - 1$ *is the least index for which* $\alpha_{j',2} < \alpha_{j'+1,1}$*, then every perfect allocation matches all nodes in some set* W_i *for some* $i \leq j'$*.*

Proof. Observe that is sufficient to prove the lemma for exact p since by increasing p for some edges can only increase the probability for the existence of a (nearly) perfect allocation. Throughout the proof we denote by 1_r the positive vector with 2r 1's $(1, 1, \ldots, 1)$.

We begin with the first part. Denote by Q the nearly perfect allocation property. Fix some feasible vector $\bar{\alpha}_1$. The proof for both $k = 2$ and $k \geq 3$ will follow from applying the Erdos-Renyi Theorem to non-directed random graphs.

First consider $k = 2$. Let p_m be the probability that a nearly perfect allocation exists in the non-directed random graph $G(m, p^2)$ (recall that this graph has exactly m nodes and each edge is generated with probability p^2). That is

$$p_m = \Pr\left[G(m, p^2) \models Q\right].$$

Consider the graph $D(1_1, m, p)$. Since a cycle of length 2 has probability p^2 and because $k = 2$

$$\Pr\left[D(1_1, m, p) \models Q\right] = p_m.$$

Let $m(\bar{\alpha_1})$ be such that $[\alpha_{0,1}m, \alpha_{1,1}m]$ contains an integer for every $m \geq m(\bar{\alpha_1})$. We define a sequence $(x_m)_{m \geq m(\bar{\alpha_1})}$ by choosing arbitrarily the integer

$$x_m \in \arg \min_{x \in N \cap [\alpha_{0,1}m, \alpha_{1,1}m]} \Pr\left[D(1_1, x, p) \models Q\right]. \tag{13}$$

Note that the minimum is attained at some value since it is taken over a finite set that includes an integer. Therefore

$$\Pr\left[D(\bar{\alpha}_1, m, p) \models Q\right] \geq Pr\left[D(1_1, x_m, p) \models Q\right] = p_{x_m}.$$

By the Erdos-Renyi Theorem since p is a constant, $p_{x_m} \to 1$ as $m \to \infty$ completing the proof for $k = 2$.

We proceed with $k \geq 3$. If m is even, a perfect allocation exists using only 2-way exchanges with probability $1 - o(1)$. If m is odd we pick arbitrarily $m - 1$ nodes. In the graph induced by these nodes we find a perfect allocation, say M, using 2-way exchanges (again, this can be found with probability $1-o(1)$). Given that such M exists, it is sufficient to find a couple of nodes v, w that are matched to each other in M so that the single unmatched node can form a 3-way exchange with v and w. Such two nodes v and w cannot be found with probability at most $(1-p^2)^{\frac{m}{2}}$, completing the first part.

The second part will follow by a reduction to a bipartite random graph followed by application of the Erdos-Renyi Theorem. Fix a feasible vector $\bar{\alpha}_r$ where $r > 1$ and let Q be the perfect allocation property. First consider $\alpha_{0,1} > 0$. Note that it is enough to prove the result for $k = r$, i.e. there exists a perfect allocation using exchanges of at most (hence exact) size r.

Consider the r-partite graph $D(1_r, m, p)$ with the sets $V_0, V_1, \ldots, V_{r-1}$ each of size exactly m, as in Definition 9.4. For each $i = 0, \ldots, r-1$ and $j = 1, \ldots, m$ let $v_{i,j}$ be the j-th node in the set V_i. We construct a bipartite graph $G(m, m, p^r)$ (as in the Erdos-Renyi Theorem) with sets of nodes V and W as follows. Let $V = V_0$ and for every $j = 1, \ldots m$, let the tuple $(v_{1,j}, v_{2,j}, \ldots, v_{r-1,j})$ be a single node in W (see Figure 9). Let

$$q_m = \Pr\left[G(m, m, p^r) \models Q\right].$$

Fix some $1 \leq j \leq m$ and some $v \in V_0$. Observe that the probability that $D(1_r, m, p)$ contains the cycle $v, v_{1,j}, v_{2,j}, \ldots, v_{r-1,j}$ is p^r. The probability that there exists an edge between $(v_{1,j}, v_{2,j}, \ldots, v_{r-1,j})$ and v is also p^r (see Figure 9). Therefore

$$\Pr\left[D(1_r, m, p) \models Q\right] \geq q_m.$$

Let $m(\bar{\alpha_r})$ be the least integer such that for every $m \geq m(\bar{\alpha_r})$ each interval $[\alpha_{j,1}m, \alpha_{j,2}m]$ contains an integer. We define a sequence $(x_m)_{m \geq m(\bar{\alpha_r})} \in N^r$ of r-tuples by choosing arbitrarily

$$x_m \in \arg \min_{x \in N^r \cap [\alpha_{0,1}m, \alpha_{0,2}m] \times \cdots \times [\alpha_{r-1,1}m, \alpha_{r-1,2}m]} \Pr\left[D(x, 1, p) \models Q\right]. \tag{14}$$

For every $m \geq m(\bar{\alpha}_r)$ let $\tilde{x}_m = \min_{i=0}^{r-1}((x_m)_i)$. Since only increasing the sizes of sets which are not the smallest ont increases the probability of a perfect allocation

$$\Pr\left[D(\tilde{x}_m, 1, p) \models Q\right] \geq \Pr\left[D(x_m, 1, p)\right].$$

Therefore

$$\Pr\left[D(\bar{\alpha}_r, m, p) \models Q\right] \geq \Pr\left[D(x_m, 1, p) \models Q\right] \geq q_{x_m}.$$

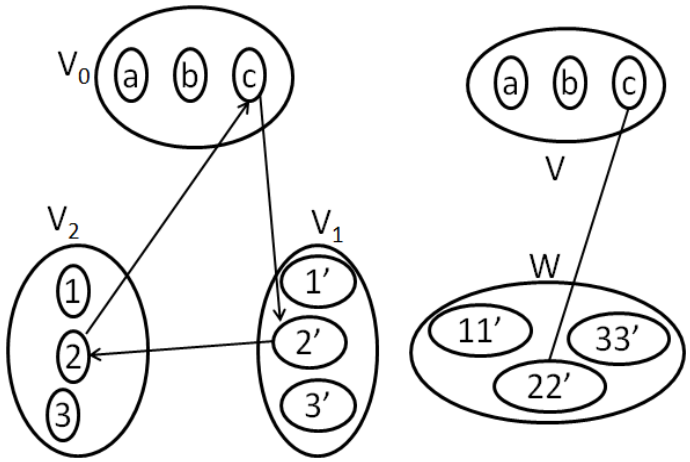

Figure 9: The graph on the left is a directed 3-partite graph and the reduction to a bipartite graph works as follows. We arbitrarily join nodes j and j' into a single node (for each $j = 1, 2, 3$) in the reduced graph. An edge in the reduced graph corresponds to a cycle in the 3-partite graph but not vice versa, for example the cycle a,1',2 has no "representative" edge in the reduced graph.

As in the first part, $q_{x_m} \to 1$ as $m \to \infty$. Therefore almost every $D(\bar{\alpha}_r, m, p)$ contains a perfect allocation. Denote by $W_0, W_1, \ldots, W_{r-1}$ the sets of pairs in each part in $D(\bar{\alpha}_r, m, p)$.

It remains to prove the result for $\alpha_{0,1} = 0$. Consider the sequence of realized sets $W_0^m, W_1^m, \ldots, W_{r-1}^m$. for $m \geq m(\bar{\alpha}_r)$. We partition this sequence into two subsequences, one (x_{m_j}) in which $W_0^m = \min(|W_0^m|, |W_1^m|, \ldots, |W_{r-1}^m|)$ and the other (y_{m_j}) in which which $W_0^m > \min(|W_0^m|, |W_1^m|, \ldots, |W_{r-1}^m|)$. For the latter subsequence the proof is similar to the case $\alpha_{0,1} > 0$. We need to show that the probability for the existence of a perfect allocation converges to one in subsequence (x_{m_j}). Note that if W_0 would have been the same size as the second smallest set in each element in the sequence again the same proof would follow. Therefore, the proof follows by observing that if a r-partite graph contains a perfect allocation, it also contains one after removing some nodes from W_0. ∎

9.2 Proof of Proposition 5.2

The proof is by construction. Let $D(m)$ be a random compatibility graph. We need to show that an allocation with the properties described in the proposition exists in $D(m)$ with probability $1 - o(1)$. Let δ be a constant such that $0 < \delta < \min\{\frac{1-2.5\tilde{\gamma}}{1+2.5\tilde{\gamma}}, 0.01, \frac{\tilde{\gamma}}{100}\}$.

Let $B_\delta(m)$ and $S_\delta(m)$ be the events defined in (11) and (12) respectively. Since $\Pr[B_\delta(m)] = 1 - o(m^{-1})$ we will assume throughout the proof that the events $B_\delta(m)$ and $S_\delta(m)$ occur (we will the probability that either one of these events does not occur towards non-existence of a desired allocation). Let V be the set of realized pairs in $D(m)$. While we assume that the

type of pair is realized we assume that the edges are yet to be realized.

Claim 9.6. *1. With probability $1-o(1)$ there exists a perfect allocation using only 2-way or 3-way exchanges in the subgraph induced by only selfdemanded pairs.*

2. With probability $1-o(1)$ there exists a perfect allocation in the subgraph induced by only A-B and B-A pairs. In particular either all A-B pairs or all B-A pairs are matched under such an allocation.

Proof. Since $B_\delta(m)$ occurs, for every selfdemanded type X-X the subgraph induced by only X-X pairs is a bounded directed graph, $D(((1-\delta)\mu_{\text{X-X}}, (1+\delta)\mu_{\text{X-X}}), m, \gamma_H)$. Therefore the first part follows by the first part of Lemma 9.5.

Similarly the graph induced by only A-B and B-A pairs is a 2-bounded directed graph, $D(((1-\delta)\mu_{\text{A-B}}, (1+\delta)\mu_{\text{A-B}}, (1-\delta)\mu_{\text{B-A}}, (1+\delta)\mu_{\text{B-A}}), m, \gamma_H)$. Hence the second part follows by the second part of Lemma 9.5. ∎

Let M_1 be an allocation in V that satisfies both parts of Claim 9.6. We will assume that such M_1 exists, and count the low probability it doesn't towards failure of the desired allocation to exist. Further assume that M_1 matches all B-A pairs, and in particular $Z_{\text{A-B}}(m) \geq Z_{\text{B-A}}(m)$ (the proof proceeds similarly if all B-A pairs are matched).

Let V' be the set of pairs that are not matched by M_1 in V. In particular V' contains all overdemanded pairs, underdemanded pairs and the A-B pairs that are not matched by M_1. The next Claim shows that all A-B pairs that are not matched by M_1 can be matched as in the hypothesis. Recall that for a set of pairs S and type t, $\tau(S,t)$ denotes the set of pairs in S whose type is t.

Claim 9.7. *With probability $1-o(1)$ there exists a perfect allocation in the subgraph induced by the sets of pairs $\tau(V', A\text{-}B)$, $\tau(V', B\text{-}O)$ and $\tau(V', O\text{-}A)$, which matches all pairs in $\tau(V', A\text{-}B)$.*

Proof. Let $\bar{\alpha}_3 = (0, 2\delta\mu_{A-B}, (1-\delta)\mu_{\text{B-O}}, (1+\delta)\mu_{\text{B-O}}, (1-\delta)\mu_{\text{O-A}}, (1+\delta)\mu_{\text{O-A}})$. Since both $B_\delta(m)$ and $S_\delta(m)$ occur the subgraph induced by the pairs in the statement is a 3-bounded directed random graph $D(\bar{\alpha}_3, m, \gamma_H)$, and the result follows by the second part of Lemma 9.5. ∎

Let M_2 be a perfect allocation as in Claim 9.7 (again assuming it exists). By Lemma 5.1 the size of of this allocation is $o(m)$.

As the hypothesis suggests we wish to match every AB-O pair in a 3-way exchange using one O-A pair and one A-AB pair (see Figure 3). Furthermore we need to match every other overdemanded pair X-Y in a 2-way to a Y-X pair. Although we have already used some

O-A pairs in M_2, the following claim shows that there are sufficiently many O-A pairs that are not matched by M_2 that can be used in order to match all A-O and AB-O pairs as we have just described. Similarly there are sufficiently many A-AB pairs to match all AB-A and AB-O pairs.

Claim 9.8. *1.* $Z_{O\text{-}A}(m) \geq (1+\delta)m(\mu_{A\text{-}O} + \mu_{AB\text{-}O}) + \lambda m$ *for some* $\lambda > 0$.

2. $Z_{A\text{-}AB}(m) \geq (1+\delta)m(\mu_{AB\text{-}A} + \mu_{AB\text{-}O})$.

Proof. Recall that $\frac{1}{\rho}$ is the probability that an arbitrary patient and an arbitrary donor are incompatible. since $B_\delta(m)$ occurs

$$Z_{\text{O-A}}(m) \geq \mu_{\text{O-A}}(1-\delta)m = \rho\mu_O\mu_A(1-\delta)m > \rho\mu_O\bar{\gamma}(\mu_A + \mu_{AB})(1+\delta)m,$$

where the last inequality follows since $\mu_{AB} < \mu_A$, and $\delta < \frac{1-2.5\bar{\gamma}}{1+2.5\bar{\gamma}} < \frac{1-2\bar{\gamma}}{1+2\bar{\gamma}}$, completing the first part. to see that the second part follows note that

$$Z_{\text{A-AB}}(m) \geq \mu_{\text{A-AB}}(1-\delta)m = \rho\mu_A\mu_{AB}(1-\delta)m > \rho\mu_{AB}\bar{\gamma}(\mu_O + \mu_A)(1+\delta)m,$$

where the last inequality follows because $\mu_O + \mu_A < 2.5\mu_A$ (see Assumption B and Footnote 21) and $\delta < \frac{1-2.5\bar{\gamma}}{(1+2.5\bar{\gamma})}$. ∎

Let $M' = M_1 \cup M_2$ and let V'' be the set of all pairs that are not matched by M'. Consider the subgraph induced by the sets of pairs $\tau(V', \text{AB-O})$, $\tau(V', \text{O-A})$ and $\tau(V', \text{A-AB})$. Observe that this graph is a 3-bounded directed random graph; indeed by Claim 9.8 there exist constants c_1 and c_2 such that the number of pairs in in $\tau(V', \text{A-AB})$ and $\tau(V', \text{AB-O})$ is at least $c_1 m$ and $c_2 m$. Therefore by Lemma 9.5 with high probability there exists a perfect allocation that will match all AB-O pairs will be matched.

To complete the construction it remains to show that for every overdemanded type X-Y except AB-O the graph induced by all X-Y and Y-X pairs that are not yet matched contains a perfect allocation exchanges of size 2. This follows from similar arguments as above.

It remains to show that one cannot obtain more transplants by allowing exchanges of size more than 3. Let e be an exchange of any size and let $\tau(e, \text{X-Y})$ be the set of pairs in e whose type is X-Y. It is enough to show that

$$\sum_{t \in \mathcal{P}^{\mathcal{U}}} \tau(e,t)| \leq 2|\tau(e, \text{AB-O})| + \sum_{t \in \mathcal{P}^{\mathcal{O}} \setminus \{\text{AB-O}\}} |\tau(e,t)| \tag{15}$$

We say that a pair v *helps* pair y if the there is either a directed edge from v to w or there is a directed path $v, z_1, z_2, \ldots, z_r, w$ where each z_i, $i \geq r$ is a selfdemanded pair. Observe that every underdemanded O-X pair must be helped by some overdemanded Y-O pair. Similarly any underdemanded pair X-AB must help an overdemanded pair AB-Y pair. Finally since an O-X underdemanded pair can help an underdemanded pair Y-AB but not vice versa, we obtain the bound. □

9.3 Individual Rationality and the Proof of Theorem 6.2:

Before we prove Theorem 6.2 it will be useful to write Claims 9.2 and 9.3 with respect to $D(H_n)$ rather than $D(m)$. We will need to rewrite the events (11) and (12) accordingly.

Lemma 9.9. *Let $0 < \delta < \frac{1}{2}$ and let $H_n = \{1, \ldots, n\}$. Moreover let χ_{H_n} be a random variable which denotes the size of all hospitals, that is $\chi_{H_n} = \sum_{h \in H_n} |V_h|$. Consider the events*

$$W_\delta(H_n) = \{\forall X\text{-}Y \in \mathcal{P},\ (1-\delta)\mu_{\ X\text{-}Y}\chi_{H_n} < |\tau(V_{H_n}, X\text{-}Y)| < (1+\delta)\mu_{X\text{-}Y}\chi_{H_n}\}, \tag{16}$$

and

$$S_\delta(H_n) = \{||\tau(V_{H_n}, A\text{-}B)| - |\tau(V_{H_n}, B\text{-}A)|| = o(n)\}. \tag{17}$$

If every hospital $h \in H_n$ is of a positive and bounded size then

$$\Pr\left[W_\delta(H_n), S_\delta(H_n)\right] = 1 - o(1). \tag{18}$$

9.3.1 Proof of Theorem 6.2:

Let $D(H_n)$ be a random compatibility graph with the set of hospitals H_n. We prove the result for the case in which each hospital has the same regular size $c \leq \bar{c}$. The proof for the general case is similar (using the fact that the size of each hospital is bounded).

Let RHS(1) and LHS(1) be the the right hand side and left hand side of inequality (1) respectively (see Definition 6.4). Fix $\delta > 0$ such that $\delta < \min(RHS(1) - LHS(1), 0.01, \frac{\bar{\gamma}}{100})$.

We assume that both events $W_\delta(H_n)$ and $S_\delta(H_n)$ as defined in (16) and (17) respectively occur and count the low probability it doesn't towards failure for the existence of an allocation with the properties described in the theorem. Lemma 6.5 is a key step in the proof.

Proof of Lemma 6.5:

One way to construct a satisfiable set S_n would be to first (i) choose randomly for each hospital a maximum set of underdemanded pairs it can internally match (by regularity and law of large numbers this will satisfy the first property of Definition 6.4), and (ii) add arbitrary pairs of each underdemanded type so that the second property of Definition 6.4 is satisfied.

Suppose S_n is constructed as above. We want to show that with high probability for each underdemanded type X-Y$\in \mathcal{P}^{\mathcal{U}}$ a perfect allocation exists in the subgraph induced by $\tau(S_n, \text{X-Y})$ and the overdemanded pairs in $\tau(V_{H_n}, \text{Y-X})$. Unfortunately Lemma 9.5 cannot be directly applied since these graphs are not 2-bounded directed random graph due to lack of independence of each edge in the graph (recall that we already have partial information on internal edges after phase (i) of the process above). Although it is true that with high

probability such a perfect allocation exists we use a slightly more subtle construction for a satisfiable set.

Instead we will partition the set of hospitals into two sets H_n^1 and H_n^2 each with $\frac{n}{2}$ hospitals, and find a satisfiable set S_n such that (i) the number of underdemanded pairs of each type X-Y in S_n belonging to H_n^1 (H_n^2) equals the number of overdemanded pairs Y-X belonging to H_n^2 (H_n^1). Then we will match overdemanded pairs of type Y-X H_n^1 (H_n^2) to X-Y underdemanded pairs in S_n belonging to H_n^2 (H_n^1), using the observation that these subgraphs are 2-uniform directed random graphs.

For every hospital $h \in H_n$, let M_h be a random allocation that maximizes the number of underdemanded pairs in the subgraph induced by its set of pairs V_h. For simplicity we will assume throughout the proof that n is even. We partition the set of hospitals into two sets $H_n^1 = \{1, \ldots, \frac{n}{2}\}$ and $H_n^2 = \{\frac{n}{2}+1, \ldots, n\}$. Define for each $j = 1, 2$

$$S_n^j = \cup_{h \in H_n^j} \tau(M_h(V_h), \mathcal{P}^{\mathcal{U}}). \tag{19}$$

and let $S = S_n^1 \cup S_n^2$. By construction S satisfies the first property in Definition 6.4. Consider the following events for $j = 1, 2$:

$$Q_n^j = \{\forall \text{X-Y} \in \mathcal{P}^{\mathcal{U}},\ |\tau(S_n^j, \text{X-Y})| < (1-\delta)\mu_{\text{Y-X}} \frac{n}{2} c\}.$$

By the regularity assumption and the law of large numbers $Pr[Q_n^j] = 1 - o(1)$ for both $j = 1, 2$, and therefore $Pr[Q_n^1, Q_n^2] = 1 - o(1)$.

Consider the events $W_\delta(H_n^j)$ for each $j = 1, 2$, where $W_\delta(H_n^j)$ is defined as in (16). Since the size of each H_n^j is $\frac{n}{2}$, from Lemma 9.9 and the fact that there are only two sets in the partition with probability $1 - o(1)$ both $W_\delta(H_n^1)$ and $W_\delta(H_n^2)$ occur.

Therefore with probability $1 - o(1)$ for each $j = 1, 2$

$$|\tau(S_n^j, \text{X-Y})| < |\tau(V_{H_n^{3-j}}, \text{Y-X})|. \tag{20}$$

Finally for each $j = 1, 2$ we add to S_n^j arbitrary underdemanded pairs belonging to H_n^j such that (20) becomes an equality for every X-Y$\in \mathcal{P}^{\mathcal{U}}$; Observe that this is feasible by applying Lemma 5.1 for $\frac{n}{2}$ hospitals. By construction $S_n = S_n^1 \cup S_n^2$ is a satisfiable set.

Let X-Y$\in \mathcal{P}^{\mathcal{U}}$ be an arbitrary type and consider the subgraph induced by the sets of pairs $\tau(S_n^1, \text{X-Y})$ and $\tau(V_{H_n^2}, \text{X-Y})$. Note that this is 2-bounded directed random graph (the realization of each edge is independent of the internal allocations M_h for each h since all potential edges in this graph are not internal). Therefore there is perfect matching in this graph with probability $1 - o(1)$. Similarly, a perfect allocation exists with high probability in the graph induced by the sets of pairs $\tau(S_n^2, \text{X-Y})$ and $\tau(V_{H_n^1}, \text{X-Y})$. Finally since there are a finite number of types the proof follows.$\square$

Let M_1 be a perfect allocation as in Lemma 6.5. We assume that such M_1 exists, again assuming that with the failure probability no allocation with the desired properties exists.

So far M_1 matches twice the number of overdemanded pairs in the graph, including for each hospital h the number of underdemanded pairs each h can internally match. As in the proof of Proposition 5.2 there exists a perfect allocation in the subgraph induced by all selfdemanded pairs with probability $1 - o(1)$, say M_2.

Finally we will show that there exists a perfect allocation in the subgraph induced by all A-B and B-A pairs which matches for each hospital at least the same number of A-B and B-A pairs it can internally match.

For each hospital there exist probabilities $\epsilon_{\text{A-B}} > 0$ and $\epsilon_{\text{B-A}} > 0$ not depending on n for not matching all their A-B and B-A pairs respectively. Therefore there exists $\epsilon > 0$ not depending on n such that with probability $1 - o(1)$ the number of A-B pairs that cannot be internally matched is at least ϵn and the expected number of B-A pairs that cannot be internally matched is at least ϵn, i.e linear in n.

However by Lemma 5.1 the difference between the number of A-B and B-A pairs is sublinear with high probability, that is, with probability $1 - o(1)$

$$||\tau(V_{H_n}, \text{A-B})| - |\tau(V_{H_n}, \text{B-A})|| = o(n). \tag{21}$$

Suppose that $|\tau(V_{H_n}, \text{A-B})| > |\tau(V_{H_n}, \text{B-A})|$ (the proof proceeds similarly if the converse inequality holds). By (21), with probability $1-o(1)$ there exists $W \subseteq \tau(V_{H_n}, \text{A-B})$ such that (i) $|W| = |\tau(V_{H_n}, \text{B-A})|$ and (ii) for each hospital h, W contains at least the number of A-B pairs it can internally match.

Using similar arguments as in the proof of Lemma 6.5 there exists with high probability a perfect allocation in the graph induced by the sets of pairs W and $\tau(V_{H_n}, \text{B-A})$, say M_3.

It remains to bound the efficiency loss, which will follow from Proposition 5.2. We consider an efficient allocation M' as in Proposition 5.2 and let $M = M_1 \cup M_2 \cup M_3$. In both M and M' all selfdemanded pairs are matched. M matches each AB-O pair in a 2-way exchange to an O-AB pair rather than carrying out a 3-way exchange as in M'. In both allocations M and M' after excluding all exchanges in which an AB-O pairs is part of, all overdemanded pairs are matched and the same number of underdemanded pairs are matched. Finally by (21) M leaves $o(n)$ A-B or B-A pairs unmatched matched whereas M' matches all A-B and B-A pairs. □

9.4 Proof of Theorem 8.6:

Let H_n be a set of bounded and strongly regular sized hospitals and let H_n^1 and H_n^2 be as in the theorem, i.e. a partition of H_n to two sets of hospitals each of size $\frac{n}{2}$. For simplicity we

will assume that all hospitals have the same size $c > 0$.[40] Fix some hospital $\bar{h} \in H_n$ and fix $V_{\bar{h}}$ to be the set of pairs (type) of hospital $\bar{h}$. Without loss of generality assume that $\bar{h} \in H_n^1$. We assume that all hospitals but $\bar{h}$ report truthfully their set of incompatible pairs.

Denote by φ the Bonus mechanism. We need to show that for any subset of pairs $B_{\bar{h}} \subseteq V_{\bar{h}}$

$$E_{V_{-h}}[u(\varphi(V_{\bar{h}}, V_{-h})] \geq E_{V_{-h}}[u(\varphi(B_{\bar{h}}, V_{-h})] - o(1). \tag{22}$$

Let RHS(7) and LHS(7) be the the right hand side and left hand side of inequality (7) respectively (see Definition 8.5). Fix $\delta > 0$ such that $\delta < \min(RHS(7) - LHS(7), 0.01)$. We assume that the events $W_\delta(H_n^1)$, $W_\delta(H_n^2)$, $W_\delta(H_n)$ and $S_\delta(H_n)$ as defined in (16) and (17) occur and as usual count the low probability they don't towards failure of the existence of an allocation as constructed in the Bonus mechanism.[41]

The following claim will imply that the strategic problem of each hospital roughly comes down to to maximizing its expected number of matched underdemanded pairs.

Claim 9.10. *If $\bar{h}$ reports truthfully $V_{\bar{h}}$, all its non-underdemanded pairs that can be internally matched will be matched by φ with probability $1 - o(1)$.*

Proof. As in the proof of Theorem 5.2, in almost every graph there exists a perfect allocation within the set of all selfdemanded pairs, thus Step 1 of the mechanism φ will find a perfect allocation with probability $1 - o(1)$. (Here and below a little bit of care has to be taken to verify that the results about uniform directed random graphs hold even when the internal subgraph of a single hospital $\bar{h}$ of bounded size c is fixed in advance.[42]) Using the same arguments as in the proof of Theorem 6.2 to match A-B and B-A pairs, we obtain in Step 2 of the Bonus mechanism, with probability $1 - o(1)$ a perfect allocation will be found in the graph induced by A-B and B-A pairs that matches all A-B and B-A that can be internally matched. Finally similarly to Lemma 6.5 all overdemanded pairs will be matched in Step 3 (to underdemanded pairs) with probability $1 - o(1)$. Since there are only 3 steps and they are all independent of one another the result follows. ∎

For any $B_{\bar{h}} \subseteq V_{\bar{h}}$ and any underdemanded type X-Y $\in \mathcal{P}^{\mathcal{U}}$. Denote by $\psi_{\text{X-Y}}(B_{\bar{h}})$ the expected number of X-Y pairs in $V_{\bar{h}}$ that will be matched when $\bar{h}$ reports $B_{\bar{h}}$ (both by the mechanism φ and, in the second stage, by $\bar{h}$).

[40] Again, since all hospitals are of bounded size a similar proof follows (one needs ca neglect sizes appear a finite number of times).

[41] Note that the internal graph of hospital $\bar{h}$ is not a random variable since it is fixed. However, Lemma 9.9 still holds since the size of $\bar{h}$ is bounded and does not affect the number of pairs in the limit. We skip here the formal details.

[42] Note that in the compatibility graph, with probability bounded away from zero some hospital has the same internal graph as $\bar{h}$.

Fix an arbitrary subset $B_{\bar{h}} \subseteq V_{\bar{h}}$ and an arbitrary underdemanded type X-Y$\in \mathcal{P}^{\mathcal{U}}$. To see that (22) holds, by Claim 9.10 it is sufficient to show that

$$\psi_{\text{X-Y}}(B_{\bar{h}}) \leq \psi_{\text{X-Y}}(V_{\bar{h}}) + o(1). \tag{23}$$

The following lemma allow us to assume that all X-Y pairs belonging to $\bar{h}$ that are chosen in the underdemanded lottery will be matched:

Claim 9.11. *All X-Y pairs chosen by the underdemanded lottery will be matched by φ with probability $1 - o(1)$, regardless of whether $B_{\bar{h}}$ or $V_{\bar{h}}$ are reported.*

Proof. Suppose $\bar{h}$ reports $B_{\bar{h}}$ (since $B_{\bar{h}}$ is arbitrary all arguments in the proof hold also if $\bar{h}$ reports $V_{\bar{h}}$). Recall that S_h(X-Y) is the set of X-Y pairs belonging to h that are chosen in the underdemanded lottery, and recall that θ_j(Y-X) $= |\tau(B_{H_{n^{3-j}}}, \text{Y-X})|$ for each $j = 1, 2$ (see Step (3a) in the Bonus mechanism).

By our assumption every hospital is strongly regular (see Definition 6.4). Therefore, by the law of large numbers and since $\bar{h}$ is of bounded size, with probability $1 - o(1)$ for each $j = 1, 2$

$$\sum_{h \in H_n^j} |S_h(\text{X-Y})| < \theta_j(\text{Y-X}).^{43}$$

Therefore with high probability the underdemanded lottery will enter the Main Step of the underdemanded lottery.[44].

Since $W_\delta(H_n^1)$ and $W_\delta(H_n^2)$ occur θ_j(Y-X) $< |\tau(B_{H_{n^{3-j}}}, \text{X-Y})|$ and θ_{3-j}(Y-X) $< |\tau(B_{H_{n^{3-j}}}, \text{X-Y})|$ for each $j = 1, 2$. Hence, for each $j = 1, 2$ the size of $\cup_{h \in H_n^j} S_h$(X-Y) at the end of the underdemanded lottery is the same size as the number of reported Y-X pairs by all hospitals in H_n^{3-j}.

In particular each of the two subgraphs containing X-Y and Y-X pairs considered in step (3b) of the bonus mechanism is a 2-bounded directed random graph (here we used that nodes on each side of a graph cannot belong to the same hospital and therefore we still have independence of each edge). Therefore, by Lemma 9.5 both these subgraphs contain a perfect allocation with probability $1 - o(1)$ and by construction all X-Y pairs in these graph will be matched with probability $1 - o(1)$. ∎

From this point on we will assume that all X-Y pairs chosen by the underdemanded lottery all end up matched (again counting the failure probability towards failing to match all underdemanded pairs of hospital $\bar{h}$ that are chosen in the underdemanded lottery).

[43] We don't know if $|B_{\bar{h}}|$ is a strongly regular size, but since it is only one bounded hospital the inquality holds.

[44] Again, we neglect formalizing that hospital $\bar{h}$'s set is fixed and not a random variable

By the Main Step of the underdemanded lottery, adding imaginary X-Y pairs to $B_{\bar{h}}$ (i.e. not from $V_{\bar{h}} \setminus B_{\bar{h}}$) can only increase $\psi_{\text{X-Y}}(B_{\bar{h}})$. We will add g new X-Y pairs to the set $B_{\bar{h}}$ assuming that each of these new pairs cannot be internally matched by $\bar{h}$, where

$$g = |\tau(V_{\bar{h}}, \text{X-Y})| - |\tau(B_{\bar{h}}, \text{X-Y})|.$$

Note that $g \geq 0$, and with a slight abuse of notation we refer from now on to $B_{\bar{h}}$ as the extended set containing the imaginary pairs. We need to show that (23) holds.

Let q and $\tilde{q}$ be the number of X-Y pairs $\bar{h}$ can match internally in $V_{\bar{h}}$ and $B_{\bar{h}}$ respectively. Observe that $\tilde{q} \leq q \leq |\tau(V_{\bar{h}}, \text{X-Y})|$. We will assume that $q < |\tau(V_{\bar{h}}, \text{X-Y})|$, otherwise (23) is satisfied since all pairs in $\tau(V_{\bar{h}}, \text{X-Y})$ will be matched by φ.

Consider the Main Step in the under demanded lottery; When $\bar{h}$ reports $V_{\bar{h}}$, each ball in J belonging to $\bar{h}$ is drawn with some identical probability $p > 0$. Similarly when $\bar{h}$ reports $B_{\bar{h}}$ each ball in J belonging to $\bar{h}$ is drawn with some identical probability $\tilde{p} > 0$. Since the number of X-Y pairs and Y-X belonging to $\bar{h}$ is bounded and the total number of X-Y and Y-X pairs in the pool approaches infinity

$$\tilde{p} = p + o(1). \tag{24}$$

We set $z = |\tau(V_{\bar{h}}, \text{X-Y})|$ and consider the case that $\bar{h}$ reports $V_{\bar{h}}$. In the initialization step of the underdemanded lottery, $S_{\bar{h}}(\text{X-Y})$ is initialized to contain exactly q X-Y pairs of $\bar{h}$, and in the Main step of the lottery, for each one of $\bar{h}$'s that is drawn, an additional X-Y pair belonging to $\bar{h}$ is added to $S_{\bar{h}}(\text{X-Y})$ as long as there are remaining X-Y pairs in $V_{\bar{h}}$. Therefore since $\bar{h}$ has at most $z - q$ additional X-Y pairs (to the initial q ones)

$$\psi_{\text{X-Y}}(V_{\bar{h}}) = q + \sum_{j=1}^{z-q-1} j \binom{z}{j} p^j (1-p)^{z-j} + (z-q) \sum_{j=z-q}^{z} \binom{z}{j} p^j (1-p)^{z-j}. \tag{25}$$

Consider now the case that $\bar{h}$ reports $B_{\bar{h}}$. Again, the initialized set $S_{\bar{h}}(\text{X-Y})$ contains $\tilde{q}$ X-Y pairs, and for each of $\bar{h}$'s balls that is drawn in the Main Step, an additional X-Y pair is added to $S_{\bar{h}}(\text{X-Y})$ (as long as it has such remaining in $B_{\bar{h}}$). Recall that we assumed that all pairs $S_{\bar{h}}(\text{X-Y})$ at the end of the lottery will be matched by the mechanism φ.

Since $\bar{h}$ hasn't reported all its pairs, it can still use pairs in $V_{\bar{h}} \setminus B_{\bar{h}}$ in exchanges to match X-Y pairs in $\tau(V_{\bar{h}}, \text{X-Y}) \setminus S_{\bar{h}}(\text{X-Y})$. By definition of q and the initialization of $S_{\bar{h}}(\text{X-Y})$, $\bar{h}$ cannot match more than an additional $q - \tilde{q}$ X-Y pairs that the mechanism hasn't matched. Therefore

$$\psi_{\text{X-Y}}(B_{\bar{h}}) \leq \tilde{q} + \sum_{j=1}^{z-\tilde{q}-1} \min(j+q-\tilde{q}, z-\tilde{q}) \binom{z}{j} \tilde{p}^j (1-\tilde{p})^{z-j} + (z-\tilde{q}) \sum_{j=z-\tilde{q}}^{z} \binom{z}{j} \tilde{p}^j (1-\tilde{p})^{z-j}, \tag{26}$$

where the second term on the right hand side follows since if j balls are drawn from J, $\bar{h}$ can match at most an additional $q-\tilde{q}$ X-Y pairs, and altogether not more than $z-\tilde{q}$ additional X-Y pairs to the first $\tilde{q}$ pairs.

Since z, p and $\tilde{q}$ are all bounded, by (24) we can replace $\tilde{p}$ with p in the right hand side of (26) and add $o(1)$. Therefore

$$\psi_{\text{X-Y}}(B_{\bar{h}}) \leq \tilde{q} + \sum_{j=1}^{z-\tilde{q}-1} \binom{z}{j} p^j (1-p)^{z-j} \min(j+q-\tilde{q}, z-\tilde{q}) + (z-\tilde{q}) \sum_{j=z-\tilde{q}}^{z} \binom{z}{j} p^j (1-p)^{z-j} + o(1). \tag{27}$$

Since $z-\tilde{q} \geq z-q$

$$\psi_{\text{X-Y}}(B_{\bar{h}}) \leq \tilde{q} + \sum_{j=1}^{z-q-1} \binom{z}{j} p^j (1-p)^{z-j} (j+q-\tilde{q}) + (z-\tilde{q}) \sum_{j=z-q}^{z} \binom{z}{j} p^j (1-p)^{z-j} + o(1) =$$

$$\tilde{q} + \sum_{j=1}^{z-q-1} j \binom{z}{j} p^j (1-p)^{z-j} + (q-\tilde{q}) \sum_{j=1}^{z-q-1} \binom{z}{j} p^j (1-p)^{z-j} + (z-q+q-\tilde{q}) \sum_{j=z-q}^{z} \binom{z}{j} p^j (1-p)^{z-j} + o(1) \leq$$

$$\psi_{\text{X-Y}}(V_{\bar{h}}) + o(1),$$

where the last inequality follows by (25) and since $(q-\tilde{q}) \sum_{j=1}^{z} \binom{z}{j} p^j (1-p)^{z-j} \leq q-\tilde{q}$. We have shown that inequality (23) is satisfied.

To see that the bound on the efficiency loss holds under the truth-telling strategy profile, note that the allocation constructed by φ has the same size/properties as the one constructed in the proof of Theorem 6.3, implying the result. □