Participation (versus free riding) in large scale, multi-hospital kidney exchange

Itai Ashlagi and Alvin E. Roth

Preliminary and Incomplete Version

October 5, 2010

Abstract

As multi-hospital kidney exchange consortia have been formed and a national exchange is contemplated, the set of "players" has grown from patients and their surgeons to include hospitals (or directors of transplant centers). 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, and enrolling only those who are hardest to match, while conducting more easily arranged exchanges internally. This behavior has already started to be observed.

We extend and quantify some of the impossibility results for small markets that illustrate the tension between efficiency and incentives to participate fully, and explore how these extend to large markets, using the theory of random graphs.

We show that the incentives to free ride by withholding pairs can be substantially ameliorated by the appropriate choice of mechanisms in large markets. It appears that achieving (close to) full participation will require some change in existing kidney exchange mechanisms when multiple hospitals are involved. However the cost of making it individually rational for hospitals to participate is low in large markets, while the cost of failing to guarantee individually rational allocations could be large, in terms of lost transplants, if that causes hospitals to match their own internal pairs.

1 Introduction

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.[1] Kidneys for

[1]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).

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.[2]

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, 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 with 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).[3]

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-

[2]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.

[3]Roth et al. (2005a) describe the formation of the New England Program for Kidney Exchange under the direction of Dr. Frank Delmonico, and these proposals were also instrumental in helping establish the Alliance for Paired Donation (APD) under the direction of Dr Mike Rees. Today, in addition to those two large regional exchanges, kidney exchange is practiced by a growing number of hospitals and formal and informal consortia, and there is active discussion of a national exchange being organized (see Roth (2008)). Computer scientists have become involved, and for an interim period an algorithm of Abraham, Blum, and Sandholm (2007) designed to handle large populations was used in the APD, and algorithms of that sort may form the basis of a national exchange.

hospital exchange consortia have been formed and a national exchange is contemplated, 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 national kidney exchange program.

2 Motivation

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.[4] Since around 2004, multi-hospital exchange programs have been organized, such as the New England Program for Kidney Exchange (NEPKE), which organizes the fourteen transplant centers in New England (cf. Roth et al. (2005a)), and the Alliance for Paired Donation (APD), which counts as members several dozen hospitals around the country (with varying degrees of participation). Hospitals participate in a multi-center exchange by reporting a list of incompatible donor-patient pairs to a central repository, and a matching mechanism chooses which exchanges to carry out. 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 examine how much a *centralized* kidney exchange program can increase the number of transplants, 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.[5]. Table 1 shows that the efficiency gains

[4]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

[5]We briefly explain here the way we conduct 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 from centralization grow as the number of (moderate sized) hospitals increases: Centralized kidney exchange can potentially increase transplants by more than 300% compared to the internal exchanges that could be accomplished when we consider 22 hospitals with an average of 11 pairs each.

No. of Hospitals	Num Of Pairs	Decentralized k=2	Centralized k=2	Decentralized Exchange k=3	Centralized Exchange k=3
2	21	3.46	5.26	4.36	6.89
4	42	6.6	13.58	8.32	18.67
6	67	11.72	25.62	14.73	35.97
8	85	14.4	36.52	18.04	49.75
10	108	17.52	47.74	22.87	64.34
12	131	22.32	60.6	28.16	81.83
14	154	26.44	74.72	33.85	98.07
16	173	28.76	84.2	36.58	109.41
18	191	31.78	95.62	39.75	122.1
20	227	38.7	116.68	49.79	144.35
22	252	44.52	131.5	55.85	161.07

Table 1: Centralized vs decentralized kidney exchange. For each $k = 2, 3$ the average number of transplants is given under two settings: (i) there is no centralized mechanism, i.e. each hospital can only match internally, and (ii) there is a centralized mechanism and all pairs belong to its pool.

However, 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 and offers a new mechanism which provides efficient outcomes. We use both theoretical analysis and simulations to investigate the problem.

---

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. 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.

3 Kidney Exchange - Basic Theory

3.1 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.[6] 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 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 $(v_n, v_1) \in E(V)$. 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.[7]

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

We will be interested in finding "large" or efficient allocations, to 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' such that $|C(V, M')| > |C(V, M)|$. M is called **k-Pareto efficient** if there exists no allocation M' such that $C(V, M') \supsetneq C(V, M)$. A matching will be called **efficient** (or **Pareto efficient)** if it is k-efficient (or k-Pareto efficient) for unbounded k, i.e. for no limit on how many transplants can be included in an exchange.[8] Note that every k-efficient allocation is also k-Pareto efficient. 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)).

[6]Pairs that are compatible would go directly to transplantation and not join the exchange pool.

[7]In APD and NEPKE k was originally set to 2, was increased to 3, and now optimization is conducted over even larger exchanges and chains.

[8]In graph theory efficient and Pareto efficient are referred to as maximum and maximal respectively.

A **Kidney Exchange Program** (or simply a Kidney Exchange) consists of a set of hospitals $H = \{1, \ldots, n\}$ and a set of incompatible pairs V_h for each hospital h. The compatibility graph induced by $\cup_{h \in H} V_h$ 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, as will be apparent, we do not rule out hospitals having preferences over which of their patients are transplanted.

An exchange that matches only pairs from the same hospital is called **internal**. Hospital h can match a set of pairs $B_h \subseteq V_h$ **internally** if there exists an allocation in B_h such that all nodes in B_h are matched.

3.2 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 match internally more pairs than the number of its pairs matched in the allocation. Formally, an allocation M in $V = \cup_{h \in H} V_h$ is not **individually rational** if there exists a hospital h and an allocation M_h in V_h such that $|C(V_h, M)| < |C(V_h, M_h)|$.[9]

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.

In the next section we give some basic results regarding efficiency and individual rationality.

3.3 Basic Results: IR and Efficiency

The first result is a negative one:

[9]Other formulations of individual rationality may be appropriate under some circumstances. Consider the following situation were hospital a has three pairs, a_1, a_2 and a_3 and can match internally only a_1 and a_2. Suppose hospital a informed a_1 and a_2 that their patients can each get a transplant. According to our proposed definition an individually rational allocation might match only a_2 and a_3 for hospital a leaving a_1 unmatched. Thus an alternative definition would be the following: an allocation is IR if for every hospital the allocation matches a specific set which it can match on its own.

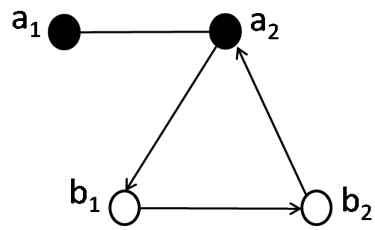

Figure 1: abc

Proposition 3.1. *For every $k \geq 3$, there exist a compatibility graph such that no k-efficient allocation is individually rational.*

Proof. The proof is given for $k = 3$ and follows from Figure 1. A similar proof holds for any $k > 3$. Formally, if there are two hospitals a and b where $V_a = \{a_1, a_2\}$, $V_b = \{b_1, b_2\}$, and the compatibility graph be as in Figure 1 then the only 3-efficient exchange is the cycle a_2, b_1, b_2, which violates individual rationality. ∎

By requiring only Pareto efficiency (rather than efficiency), one can also obtain individual rationality:

Proposition 3.2. *For every $k \geq 2$, and every compatibility graph there exists a k-Pareto efficient 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 by Edmonds (1965). The algorithm begins by finding an IR allocation in each hospital and searches for allocations that only enlarge the set of matched nodes:

Augmenting Algorithm:

1. Input: V_h for every h.

2. Find a k-efficient allocation in V_h for every h.

3. Repeat: (i) search for an allocation that increases the total number of matched pairs without unmatching a pair that was previously matched. (ii) if an allocation was found in (i) then replace the existing allocation with the new one. Otherwise terminate.

Choosing an individually rational allocation rather than an efficient allocation in the graph in Figure 1 costs one transplant. A slight extension of the example in Figure 1, yields a higher "price" of individually rationality:

Proposition 3.3. *For every $k \geq 3$, there exist a compatibility graph such that no k-Pareto efficient 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-Pareto efficient allocation is at least $\frac{1}{k-1}$ times the size of a k-efficient allocation.*

Proof. Fix some hospital a with k vertices, and suppose that a has a single internal exchange consisting of all of its pairs. The lower 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. Let V be a set of nodes and let M and M' be k-efficient and k-Pareto efficient allocations in V respectively. Fix an exchange c with size $1 \leq l \leq k$ in M' and suppose not all of the nodes in c are covered by M. Each node in c that is covered by M' is either not covered by M or part of an exchange of at most size k in M. Therefore at most $(l-1)k$ more nodes are covered by M in exchanges that at least one of their nodes belong to c. Since $(l-1)k/l$ is maximized at $l = k$ we obtain the result. ∎

For the case of $k = 2$, the sets of simultaneous matched nodes in an arbitrary graph form a matroid (see Edmonds et al. (1971)). Thus, by applying the augmenting algorithm we obtain:

Proposition 3.4. *If $k = 2$ there exists an individually rational allocation that is also k-efficient in every compatibility graph.*

Proposition 3.3 gives a worst-case result. But 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 2). So the cost of guaranteeing individual rationality is low, while (as we saw in Table 1) the cost of failing to guarantee it could be large if that causes hospitals to match their own internal pairs.

We will also prove this formally for large compatibility graphs in Section 6.1. This small efficiency loss motivates our concentration on IR and Pareto efficient allocations. In the next section we study mechanisms for Kidney Exchange Programs.

4 Kidney Exchange Mechanisms and Internal Hospital Allocations

A kidney exchange **mechanism**, φ, maps a profile of incompatible pairs $V = (V_{h_1}, V_{h_2}, \ldots, V_{h_n})$ to an allocation, denoted by $\varphi((V_h)_{h \in H})$. With a slight abuse of notation we will denote by

No. of Hosps.	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
Non IR, 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 2: Number of transplants achieved using individually rational allocations vs. using efficient (and not necessarily individually rational) allocations.

V both the profile and the union of its elements, $V = \cup_{h \in H} V_h$. A mechanism φ is IR if for every V, $\varphi(V)$ is IR. Efficient and Pareto efficient 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 $\sigma(V) = (\sigma_1(V_1), \ldots, \sigma_n(V_n))$ 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 $V = \cup_{h \in H} V_h$, 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 $V_h \setminus C(V_h, \varphi(\sigma(V)))$. In particular every hospital $h \in H$ 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 at strategy profile σ is defined by the number of h's pairs who are covered by the centralized match, plus the number of its remaining pairs that it can cover with an internal match:

$$u_h(\sigma_h(V_h), \sigma_{-h}(V_{-h})) = |C(V_h, \varphi(\sigma(V)))| + |C(V_h, \varphi_h(V_h \setminus C(V_h, \varphi(\sigma(V)))))|. \tag{1}$$

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

5 Incentives

Loosely speaking, most of the kidney exchange mechanisms presently employed choose an efficient allocation in the (reported) exchange pool.[10] But 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.

5.1 Incentives - Strategyproofness

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}

$$u_h(\varphi(V_h, V_{-h})) \geq u_h(\varphi(\sigma'_h(V_h), V_{-h})). \tag{2}$$

Roth et al. (2007a) showed:

Theorem 5.1 (Roth et al. (2007a)). *No IR mechanism is both Pareto-efficient and strategyproof.*

Proof. Consider two hospitals, a and b such that $V_a = \{a_1, a_2, a_3, a_4\}$ and $V_b = \{b_1, b_2, b_3\}$ and let $V = V_a \cup V_b$. Further assume the compatibility graph induced by V is given in Figure 2.

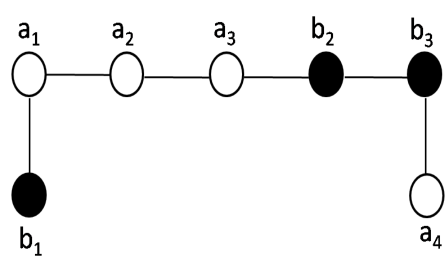

Figure 2

Note that every Pareto efficient allocation leaves exactly one node unmatched. Suppose φ is both Pareto efficient 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 $v \in V$ be

[10] 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)..

unmatched in $\varphi(V)$. If $v \in V_a$ then $u_a(\varphi(V)) = 3$. However, by withholding a_1 and a_2, a's utility is 4 since the Pareto efficient 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 be efficient:

Theorem 5.2. *No IR strategyproof mechanism can guarantee more than $\frac{1}{2}$ of the efficient allocation in every V.*

Proof. Consider the same setting as in the proof of Theorem 5.1 (see Figure 2) and suppose φ is an IR strategyproof mechanism which guarantees more than 1/2 of the efficient allocation in every possible V. 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 Theorem 5.1, 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 randomizing 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:

Theorem 5.3. *No IR strategyproof (in expectation) randomized mechanism can guarantee more than $\frac{7}{8}$ of the efficient allocation in every V.*

Proof. Consider the same setting as in the proof of Theorem 5.1 (Figure 2) 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. (2010) study dominant strategy mechanisms for $k = 2$ and provide a strategyproof (in expectation) randomized mechanism which guarantees 0.5 of the 2-efficient allocation.[11] Strategyproofness is independent of any probability distribution of the underlying compatibility graphs. However, in the case of compatibility of kidneys, a lot is known about the (approximate) distribution of compatibility graphs, that may be useful for finding mechanisms that can achieve (almost) efficient allocations as Bayesian equilibria.[12] 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) and Kojima and Pathak (2009)). First we study more carefully the structure of compatibility graphs.

6 Incentives - Bayesian Setting

6.1 Random Compatibility Graphs

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 their 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 $\mu_X > 0$. We will assume that (as in the U.S. population) $\mu_O > \mu_A > \mu_B > \mu_{AB}$. For any two blood types X and Y, we write $Y \rhd X$ if a donor of 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.[13]

A patient-donor pair have 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 that a donor and a patient will be compatible they should be **both** blood type compatible and tissue-type compatible. To test tissue type compatibility a *crossmatch* test is performed. In practice each patient has a different level of *percentage reactive antibodies* (PRA) which determines the likelihood that the patient will

[11] The 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.

[12] Showing 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.

[13] 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. 2005 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.)

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; the probability that a patient p with PRA $Q \in \{H, L\}$ and a donor are tissue type incompatible is given by γ_Q. Furthermore the probability that a random patient has PRA L is given by $\tau > 0$. Let $\bar{\gamma}$ denote the expected PRA level of a random patient, that is $\bar{\gamma} = \tau\gamma_L + (1-\tau)\gamma_H$.

Definition 6.1 (Random Compatibility Graph)**.** *A* ***random (directed) compatibility graph*** *of size m, denoted by $D(m)$, consists of m incompatible pairs, and a random edge is generated between every donor and each one of her compatible patients. 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 chosen independently according to the probability distribution $\mu = (\mu_X)_{X \in \{A,B,AB,O\}}$. The PRA of p, denoted by γ^p is also randomized (L with probability τ and H with probability $1-\tau$).*

   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 $z \leq \gamma^p$. Each realized node is assigned randomly to one of the hospitals in H.

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

To analyze random compatibility graphs we will use results and methods from random graphs based on the Erdos-Renyi model (see e.g. Erdös and Rényi (1959) and Erdös and Rényi (1966)). A **random graph** $G(m,p)$ is an *undirected* graph with m nodes such that between each two different nodes an edge exist with probability p (where p can be a function of m). A **bipartite random graph** $B(m,p)$ 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 edges 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. A **matching** in an undirected graph is a set of edges for which no two edges have a node in common.

Erdos-Renyi Theorem *Let $\epsilon > 0$.*

1. *Let $G(m,p)$ be a random graph where $p(m) \geq \frac{(1+\epsilon)lnm}{m}$. The probability that there exists a matching that matches all nodes but at most one approaches 1 as m tends to infinity.*

2. *Let $B(m,p)$ be a random graph where $p \geq 2\frac{(1+\epsilon)lnm}{m}$. The probability that there exists a matching that matches all nodes approaches 1 as m tends to infinity.*

For simplicity we adopt the following formalism from random graph theory: if the probability that a given property Q (e.g. a perfect matching) exists in $G(m,p)$ ($B(m,p)$) tends to 1, when $m \to \infty$ we say that Q holds in **almost every graph** $G(m,p)$ ($B(m,p)$).

We will be interested in properties of the random compatibility graph $D(m)$. Thus, we say that a property Q holds for **almost every** $D(m)$ if Q is satisfied almost surely when $m \to \infty$. Since we study large graphs we let γ (the probability for tissue type incompatibility) be a non-decreasing function of m, with the special interesting case at which γ is a constant.

One property that is immediate to derive will give a better idea on the relation between different pair types in $D(m)$. We will use the following notations. The (posterior) probability that an incompatible pair (p,d) has type X-Y be $\mu_{\text{X-Y}}$. In particular there exists $\rho > 1$ such that if X and Y are two blood types such that $\text{Y} \rhd \text{X}$ then $\mu_{\text{X-Y}} = \rho\mu_X\mu_Y\bar{\gamma}$ and otherwise $\mu_{\text{X-Y}} = \rho\mu_X\mu_Y$. The number of incompatible pairs of type X-Y in $D(m)$ is a random variable denoted by $Z_{\text{X-Y}}(m)$ (or just $Z_{\text{X-Y}}$ when it is clear form the context).

Lemma 6.2. *Let $0 < \delta < 1$. In almost every $D(m)$:*

$$\Pr(\forall\ X\text{-}Y \in \mathcal{P}\ (1-\delta)mp\mu_{X\text{-}Y} < Z_{X\text{-}Y} < (1+\delta)m\mu_{X\text{-}Y}) = 1 - o(m^{-1}).^{14}$$

Lemma 6.2 implies that in almost every $D(m)$ the number of O-X pairs is strictly larger than the number of $X - O$ pairs for $X \in \{A, B, AB\}$. Similarly is relation between $X - AB$ and $AB - X$ for $X \in \{A, B\}$.

In general Lemma 6.2 motivates the following partition of pair types (see also Roth et al. (2007b) and Ünver (2010)):

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

be the set of **over demanded** 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 **under demanded** types.

[14]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.

Let

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

be the set of **self demanded** types, and finally let $\mathcal{P}^{\mathcal{R}}$ 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 over demanded pair. Under-demanded types have the reverse property: they are seeking a kidney that is in greater demand than the one they are offering in exchange. The donor and a patient with a self demanded type have the same blood type.

We will next study efficient allocations in random compatibility graphs.

6.2 Efficient Allocations

We will make the following assumption which is compatible with real life tissue-type (sensitivity) frequencies. Zenios et al. (2001) reported that for non-related blood type donors and recipients $\bar{\gamma} = 0.11$. [15]

Assumption A [Non-highly-sensitized patients] $\bar{\gamma} < \frac{1}{2}$.[16]

Theorem 6.3. *Almost every $D(m)$ has an efficient allocation that requires exchanges of no more than size 3 with the following properties:*

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

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

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

4. *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 a X-Y pair is matched in 3-way exchanges with a O-Y pair and a X-O pair.*

5. *Every over demanded pair X-O ($X \neq O$) that is not matched above is matched to a O-X pair.*

[15] One can extend our results for larger tissue-type incompatibility probability.

[16] This assumption is also used for avoiding case-by-case analysis; one can extend the results to the opposite inequality.

The proof of Theorem 6.3 is deferred to the Appendix. Roth Sonmez and Unver (2007) show a similar result to Theorem 6.3 and a with a bit of effort a similar result can be derived also from Unver (2009). Both these works 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 works show that at most 4-way exchange are needed to find an efficient allocation (Unver (2009) analyzes a dynamic world). The difference from our result (we need at most 3-way exchanges) follows from the fact that they assume 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) support our findings. It is important to note in our model although $\mu_{\text{A-B}} = \mu_{\text{B-A}}$ the probability that the number of each of such pairs is different is positive, but the difference will almost always be sufficiently small to make 4-way exchange unproductive.

Sketch of proof of Theorem 6.3: We will use a simple extension of the Erdos-Renyi Theorem (Lemma 8.3) to l-partite 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 graph consists of a *perfect* allocation (an allocation which matches all the pairs in the smallest "part" of the graph).

Since $\mu_{\text{A-B}} = \mu_{\text{B-A}}$ one can show that with high probability (follows from Lemma 6.2) the difference between the number of A-B pairs and the number of B-A pairs is small. Suppose that the number of A-B pairs is a least the number of B-A pairs (the converse is symmetric). An application of the Erdos-Renyi Theorem provides that all self demanded 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 through 2-way exchanges. We choose such an allocation at random, say M_1.

Let $V_{\text{A-B}}$ be the set of A-B pairs that are not matched so far by M_1 (see Figure 3). With high probability the size of $V_{\text{A-B}}$ is smaller than both the size of the set of B-O pairs and the size of the set of O-A pairs. Again using an application of the Erdos-Renyi Theorem this graph almost always contains a perfect allocation, implying that all A-B pairs can be matched. Similarly one can show match with high probability all AB-O pairs using 3-way exchanges which contain B-AB pairs and O-B pairs. Using our assumptions on γ one can show that there are more (and a non trivial amount) O-B pairs and O-A pairs than B-O and A-O pairs respectively that are not yet matched. Therefore all remaining over demanded pairs can be matched to under demanded pairs (again by considering the bipartite graphs they induce).

By construction every pair whose type is colored in Figure 3 (as well as all self demanded pairs) is matched implying that we obtained a 3-efficient allocation. To see that that $k \geq 4$

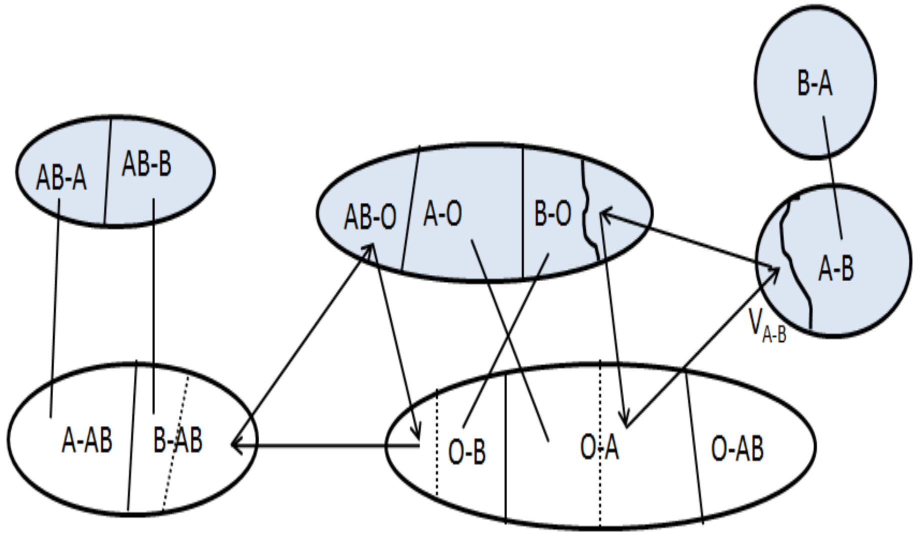

Figure 3: The graph $D(m)$ (excluding all self demanded pairs) is partitioned either conditionally on previous found allocations (scribbled lines) or independently (bold lines) in order to find exchanges. All pairs in colored regions will get transplants.

cannot increase the number of transplants we consider only the 4-way exchange with pairs AB-O,O-A,A-B and B-AB (see Figure 4). Such an exchange uses an AB-O pair and a A-B pair that is not matched to a B-A pair. But both of these pairs are all matched in $k = 3$ in 3-way exchanges implying that using such a 4-way will result in fewer transplants.

From Theorem 6.3 and its proof one can derive the efficiency loss between different k's in large graphs. Let $Z(k, m)$ be the expected size of a k-efficient allocation in $D(m)$.

Corollary 6.4. *1. For every $k \geq 2$, $\lim_{m\to\infty} \Pr(Z(3, m) \geq Z(k, m)) = 1$.*

2. For every $\epsilon > 0$, $\lim_{m\to\infty} \Pr\{Z(3, m) - Z(2, m) \leq (1 + \epsilon)(\mu_{AB-O})m + \epsilon\mu_{A-B}m\} = 1$.

3. In almost every graph $D(m)$, in all efficient allocations all over demanded pairs are matched.

To this point nothing is said about incentives in the Bayesian setting. In the next section we study the efficiency loss when requiring individual rationality.

6.3 Individually Rational Allocations

Throughout the remainder of the paper we will make following assumption:[17]
Assumption B: Each hospital can conduct exchanges of size at most 3.

[17]This assumption used to simplify our arguments and all our theorems hold without it.

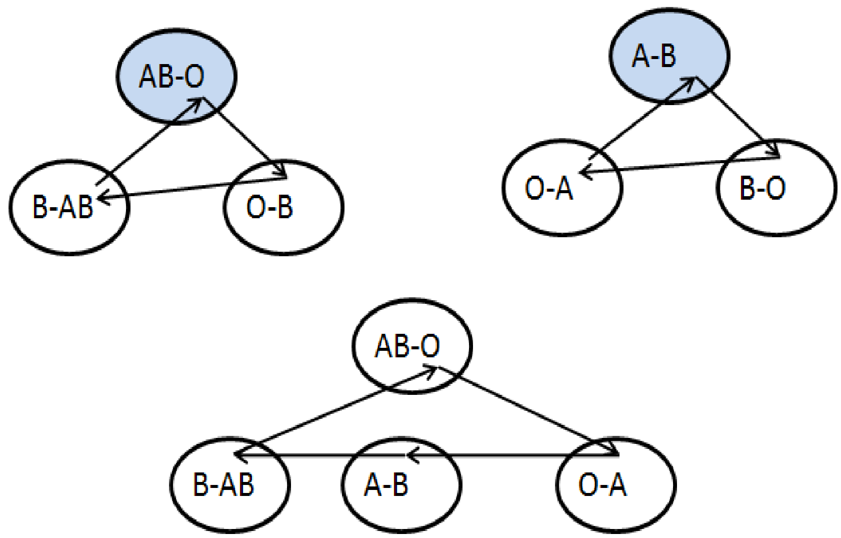

Figure 4: The possible 4-way exchange uses the bottlenecks of the 3-way exchanges - AB-O pairs and A-B pairs.

One way to bound the efficiency loss is by attempting to construct an efficient allocation as in Theorem 6.3, while making sure that the pairs each hospital can internally match are part of the efficient allocation. In such an efficient allocation only under demanded pairs are not matched. Unfortunately guaranteeing that *all* those under demanded pairs are part of the relevant allocation while matching all other types of pairs is not always feasible.

Consider the two types of 3-way exchanges (i) A-O,O-B and B-A, (ii) B-O,O-A and A-B which we will refer to by *special* 3-way exchanges. The first type, (i), contributes to a "wrong" asymmetry between the number of matched A-O and O-A pairs ("wrong" - since all A-O pairs will be easy to match in a large exchange pool), and similarly the second type, (ii), contributes to a "wrong" asymmetry between the number of matched B-O and O-B pairs. If there are many 3-way exchanges of type (i) as well as other exchanges that include O-B pairs but not B-O pairs (see e.g. Figure , we might run into a situation that more O-B pairs need to be matched than the total number of B-O pairs.

The next definition captures this extreme asymmetry. Let $c > 0$ and let $D(c)$ be a random compatibility graph. Denote by $R^c_{\text{O-A}}$ ($R^c_{\text{O-B}}$) the expected maximum number of O-A (O-B) pairs that can be matched through Pareto efficient allocations, where each chosen exchange in which a O-A (O-B) pair is part of contains also an A-O or a B-O pair. Similarly let $R^c_{\text{A-AB}}$ ($R^c_{\text{B-AB}}$) the expected maximum number of A-AB (B-AB) pairs that can be matched through Pareto efficient allocations, where each chosen exchange a A-AB (B-AB) pair is part of contains also a AB-A pair or a AB-B pair.

Definition 6.5. *We say that c is a* critical *size if for some under demanded pair X-Y$\in$*

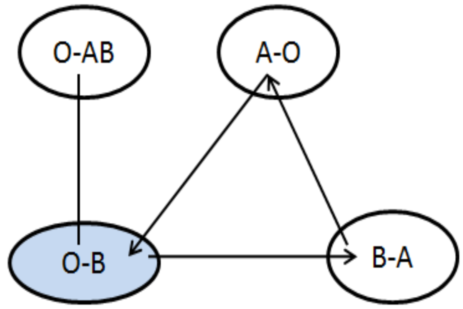

Figure 5: Example of exchanges that increase the number of B-O matched pairs and not the O-B matched pairs.

$\mathcal{P}^{\mathcal{U}} \setminus \{AB\text{-}O\}$, $R^c_{X\text{-}Y} > Z^c_{Y\text{-}X}$ *where* $Z^c_{Y\text{-}X}$ *is the expected number of Y-X pairs in* $D(c)$.

Calculating critical c's is difficult. Intuitively, for small c's are not critical since a hospital of such size will rarely have exchanges at all. Large enough c's are not critical since efficient matchings will have properties similar to those in Theorem 6.3. We ran simulations using approximated distributions and found that only for $c \geq 70$ it is critical.

In the next theorem we show that the efficiency loss is small given that hospitals are of a non-critical size.

Theorem 6.6. *If every hospital* $h \in H$ *is of a non-critical size c, then in almost every* $D(m)$ *the maximum size of an individually rational allocation is at most* $(1+\epsilon)(\mu_{AB-O})m$ *smaller than the efficient allocation for any* $\epsilon > 0$.

The proof of Theorem 6.3 is constructive and appears in the appendix.

Remarks: (a) The efficiency loss follows since some O-AB pairs will need to matched using AB-O which could have been used in 3-way exchanges. (b) Note that the lower bound we obtained in Theorem 6.6 is at least the size of a 2-efficient allocation. This is not obvious since there exist realizations for which the size of a maximum individually rational allocation is smaller than the size of a 2-efficient allocation. For example when one may need to choose a special 3-way exchange A-O,O-B,B-A over a couple of 2-way exchanges, one with A-B and B-A pairs and one with A-O, O-A pairs (see e.g. Figure 5). As the theorem asserts, this is very unlikely.

We want to bound the efficiency loss for any c. As we mentioned above values for critical c's are not small. In a hospital h with a large size c we expect to have more O-A pairs than A-O and AB-O pairs together with high probability. These extra pairs will help "fixing" the wrong asymmetry when deciding which pairs to match. We will call a hospital *regular* if it has fewer A-O and AB-O pairs than O-A pairs and fewer B-O and AB-O pairs than O-B pairs.

The bound will follow from the following intuition. Recall that if h is a hospital of critical size, then it may be possible to match more O-B (O-A) pairs than B-O (A-O) pairs. However if h is regular, it has enough spare under demanded pairs of types O-A and O-A that can be matched instead of those under demanded pairs. Thus if it has for example an internal special exchange A-O,O-B,B-A then the idea would be to guarantee to match the A-O,B-A and some other O-A pair. If a hospital is non-regular the allocation will match its nodes from the special internal exchanges, by choosing those exchanges if it is of critical size, and by just guaranteeing those nodes if it is not.

Denote by $\pi(c)$ the probability that a hospital is not regular. Observe that $\pi(c) \to 0$ as c approaches infinity. From Chernoff bounds (see Lemma 8.1) $\pi(c) > 1 - e^{-\beta c}$ for $\beta \geq \mu_{AB-O} + \mu_{B-O} + \mu_{O-B}$ (the inequality follows by restricting attention to the subgraph induced by only those three pair types).

Theorem 6.7. *Suppose every hospital is of size c. In almost every graph $D(m)$ there exist an individually rational allocation using exchanges of size at most 3, which is at most $(1+\epsilon)(\mu_{AB\text{-}O} + \mu_{AB\text{-}A} + e^{-\beta c}\min(\mu_{A\text{-}O}, \mu_{A\text{-}B}))m$ smaller than the efficient allocation for any $\epsilon > 0$.*

Again simulating with approximated distributions we obtained that $\pi(c) \sim 0.22$ for $c = 5$, $\pi(c) \sim 0.19$ for $c = 10$, and $\pi(c) \sim 0.14$ for $c = 20$. In particular the term $e^{-\beta c}\min(\mu_{\text{A-O}}, \mu_{\text{A-B}})$ is larger than the term $\mu_{\text{AB-O}} + \mu\text{AB-A}$ for already reasonably small c's.

Consider the following stronger notion of individually rationality which guarantees every hospital a maximum set of pairs that it can internally match:

Definition 6.8. *Let V_h be the set of pairs of each hospital h and let M be an allocation in $V = \cup_{h \in H} V_h$. M is **strongly individually rational** if for every h there exists an efficient allocation M_h in V_h such that $C(V_h, M_h) \subseteq C(V_h, M)$.*

Under strongly individual rationality the "trading" we done for normal hospitals is not valid, implying a larger bound.

Theorem 6.9. *Suppose every hospital is of size c. In almost every graph $D(m)$ there exist a strong individually rational allocation using exchanges of size at most 3, which is at most $(1+\epsilon)(\mu_{AB\text{-}O} + \mu_{AB\text{-}A} + \min(\mu_{A\text{-}O}, \mu_{A\text{-}B})m$ smaller than the efficient allocation for any $\epsilon > 0$.*

See Table 2 (in the previous section) for simulations results; in these simulations we use approximated blood type and tissue type distributions, and ran an exchange mechanism which outputs a strong individually rational Pareto efficient allocations. The average loss is very small for different sizes of graphs, for example, for 22 hospitals the average size of an IR Pareto efficient allocation is 160.74 and the average size of an efficient allocation is 161.1.

6.4 Mechanisms for Random Kidney Exchange Programs

As in the previous section hospitals are assumed to know their own compatibility graph but not the entire underlying graph. 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 $\Gamma(f) = (H, (T_h)_{h\in H}, (u_h)_{h\in H})$ where H is the set of hospitals, u_h is the utility function for hospital h, and T_h is the set of possible types for each hospital, where a type is drawn as in the compatibility graph generating process, that is the the underlying graph is generated, and each hospital observes its own subgraph.

The expected utility for hospital h at strategy profile σ is

$$E[u_h(\varphi(\sigma_h(V_h), \sigma_{-h}(\tilde{V}_{-h}))]. \tag{3}$$

Let σ be a strategy profile and let $\epsilon > 0$. Strategy σ_h is an ϵ-best response against σ_{-h} if for no σ'_h

$$E[u(\varphi(\sigma'_h(V_h), \sigma_{-h}(\tilde{V}_{-h}))] \geq E[u(\varphi(\sigma_h(V_h), \sigma_{-h}(\tilde{V}_{-h}))] - \epsilon. \tag{4}$$

σ is an ϵ-Bayes Nash equilibrium if 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 $\Gamma^n(\varphi)$ denotes a random kidney exchange game with $|H| = n$ hospitals.

We begin with analyzing current kidney exchange mechanisms.

6.4.1 The Status Quo

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

Theorem 6.10. *In the sequence of games $\Gamma^1(MT), \Gamma^2(MT), \ldots, \Gamma^n(MT), \ldots\ldots$ there exist no $\epsilon(n) = o(1)$ such that the truth-telling strategy for hospital h is an $\epsilon(n)$ best response against any strategy profile of all other hospitals in the kidney exchange game with n hospitals. Consequently, there exist no $\epsilon(n) = o(1)$ such that reporting truthfully is an $\epsilon(n)$-Bayes Nash equilibrium in $\Gamma^n(MT)$.*

Proof. It is enough to provide an example of a compatibility graph for some hospital in which it is better off not reporting truthfully. Fix some hospital h and suppose $V_h = \{v_1, v_2\}$ where

v_1 nd v_2 are of blood type pairs O-B and B-O respectively and v_1, v_2 is an internal exchange. Let $q = \frac{\mu_{\text{B-O}}}{\mu_{\text{O-B}}}$. There exist a large enough n such that the set of reported pairs T_n by all n hospitals satisfies the following: with probability at least $1 - o(1)$ every efficient allocation matches all over demanded pairs B-O. Furthermore each pair of type O-B is matched with probability $q + o(1)$. Therefore if hospital h reports both pairs v_1 and v_2, then the expected utility for h is $1 + q + o(1)$ and by not reporting both pairs h obtains a utility of 2. ∎

One may suggest that although a hospital might be better off withholding some pairs, efficiency would not be harmed. In the proof of Theorem 6.10 we showed that hospital h is better off withholding an exchange that consists of one over demanded pair and one under demanded pair. In a small exchange pool obviously such withholding can result in less transplants, but even in a large exchange pool this can cause efficiency loss since the over demanded pair can be part of a 3-way exchange rather than a 2-way exchange in the optimal allocation as Theorem 6.3 suggests. If all hospitals withhold such exchanges, this might lead to a substantial efficiency loss as will be illustrated below.

Essentially the asymmetry in blood type frequencies what gives Theorem 6.10. The MT might might further deepen the asymmetry: consider hospitals that withhold internal special 3-way exchanges. Since the expected number of each of the two special 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 over demanded type. Considering future exchanges, one wishes to overcome asymmetries rather than create new ones.

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 on itself. As depicted in Figure **??**, withholding provides on average more transplants than truth-telling for an arbitrary hospital given that all other hospitals are truth-telling. The benefit becomes even higher when all other hospitals also withhold internal matches (see Figure 6).

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 existing (non IR) mechanism. The efficiency loss is about 10% in both $k = 3$ and $k = 2$ (see Figure 7).

Consider the following mechanism.

Guarantee Mechanism (GM): *Let $V = (V_1, \ldots, V_h)$ be the profile of reported sets of incompatible pairs.*

Step 1*: For each h choose randomly a Pareto efficient allocation M_h which:*

(a) matches the largest number of under demanded pairs in V_h, and

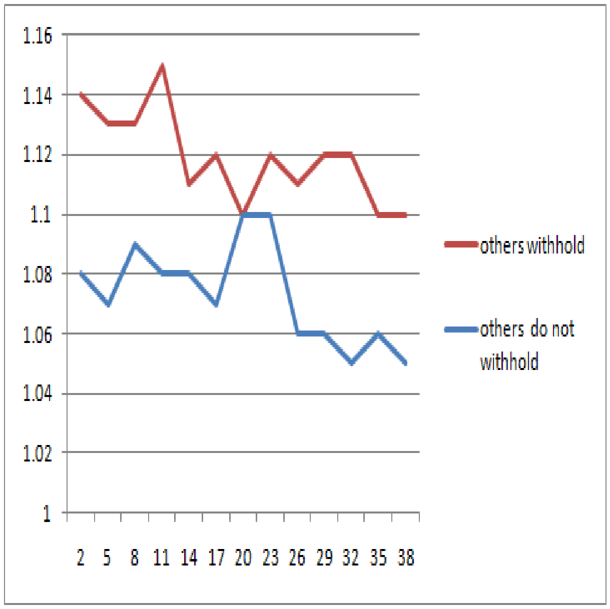

Figure 6: Withholding internal matches vs. reporting truthfully (k=3).

(b) matches the largest number of A-B or B-A pairs while satisfying (a).

Let S_h be the set of pairs that are matched under M_h.
Step 2*: Choose randomly an allocation with a maximum size for which matches all nodes in S_h for every h.*

Theorem 6.11. *The truth-telling strategy profile is an $\epsilon(m)$-Bayes Nash equilibrium for $\epsilon(m) = O(\frac{1}{m})$ in the game induced by the GM mechanism.*

The proof will appear in the full version.

7 Post Allocation Deviations

Currently hospitals need not commit to exchanges chosen by the centralized exchange. Indeed NEPKE reports show that hospitals withheld pairs after they have observed the allocation. We consider the following three stage model:

1. The mechanism chooses an allocation in the compatibility graph induced by the sets of reported pairs.

2. Each hospital chooses which incompatible pairs to withhold (in addition to those it already haven't reported in the first stage). All pairs in exchanges that at least one pair has been withheld, are considered as not matched, and the mechanism chooses another allocation in the graph induced by the set of reported pairs that are unmatched.

No. of Hospitals	Num Of Pairs	Profitable Naïve Strategy, k=2	IR, k=2	Profitable Naïve Strategy, k=3	IR, k=3	Non IR & non strategic hostpials, k=3
2	21	5.08	5.26	5.99	6.8	6.89
4	42	12.66	13.58	15.85	18.37	18.67
6	67	23.78	25.62	30.01	35.42	35.97
8	85	33.54	36.52	41.62	49.3	49.75
10	108	44.04	47.74	54.21	63.68	64.34
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

Figure 7: Withholding internal matches vs. reporting truthfully.

3. Each hospital chooses an allocation in the graph induced by its own pairs that are unmatched after the previous stage.

The third stage is similar to the second stage in the basic model. Intuitively, this model provides the hospitals: "see what you get, and then decide". To illustrate this model consider the following example.

Example 1. *Consider the compatibility graph in Figure 5 and assume a_1 and a_2 belong to hospital a and b_1, b_2, b_3 belong to hospital b. If the chosen allocation matches all nodes but b_3 then if b withholds b_1 and b_2 in the second stage then the mechanism will match b_3 and a_2, which increases the utility of b from 2 to 3.*

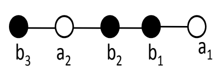

Figure 5.

Theorem 7.1. *Let $k = 2$. Consider the game induced by the GM mechanism (the same rule applies in both stages) and all hospital reports truthfully in the first stage. Then it is not beneficial for any hospital h to withhold any pairs in the second stage regardless what other hospitals withhold in that stage.*

Proof. Suppose some hospital h reports truthfully in the first stage but benefits from withholding a set of pairs $X_h \subseteq V_h$ in the second stage of the GM mechanism. Let X_h be a set with smallest size cardinality for which h benefits from withholding. Denote by M the allocation after the 1st stage of the mechanism and let $\bar{M}$ be the union of the allocations chosen either by the mechanism or hospital h in the entire graph at the end of all stages. By the minimality of X_h there is no exchange c in X_h such that $c \in M$, otherwise, if $c = v_1, v_2$ is such an exchange then h would benefit by withholding $X_t \setminus \{v_1, v_2\}$.

Since h benefits there exists a node $v \in V_h$ that is matched by $\bar{M}$ and not by M. Let $M \Delta \bar{M} = \{(v,u) : (v,u) \in M \setminus \bar{M} \quad \text{or} \quad (v,u) \in \bar{M} \setminus M\}$ be the symmetric difference of M and $\bar{M}$. Since h benefits there exist a sequence of distinct nodes $v = v_1, v_2, \ldots, v_m$ for some $m \geq 2$ such that $(v_{2i-1}, v_{2i}) \in \bar{M}$ and $(v_{2i}, v_{2i+1}) \in \bar{M}$ and at least for one i, $v_i \in X_t$. Choose such a sequence with a largest possible m. Note that m must be odd, otherwise M is not efficient: taking $(v_1, v_2), (v_3, v_4), \ldots, (v_{m-1}, v_m)$ to be in M instead of $(v_2, v_3), \ldots, v_{m-2}, v_{m-1})$ increases the size of M. By the minimality of X_h there are at most two nodes in the sequence that belong to X_t. Therefore by the construction of the mechanism $m \leq 5$, otherwise for some $1 \leq i \leq m$, $v_i, v_{i+1} \notin X_h$ and (v_{i+1}, v_{i+1}) is an exchange in M. Thus either $m = 3$ or $m = 5$. Suppose $m = 3$. If $v_2 \in V_h$ then $v_3 \in V_h$ otherwise this would contradict maximizing on internal pairs. But this means hospital h does not benefit. If $v_2 \notin V_h$ then v_1 and v_3 are both V_h which again means that h does not benefit. Finally suppose $m = 5$. Then $v_3, v_4 \in X_h$ and at least one of v_2 and v_5 are not in V_h. This contradicts that the mechanism maximized on internal pairs: the allocation $M' = M \cup \{(v_1, v_2), (v_3, v_4)\} \setminus \{(v_2, v_3), (v_4, v_5)\}$ is efficient and has more internal pairs than M. ∎

Note that the RM mechanism which chooses randomly between all efficient IR matchings, is manipulable in the second stage, as can be seen in Example 1.

For $k \geq 3$, IR and efficiency do not hold together. However even by requiring only Pareto efficiency Theorem 7.1 does not hold for $k \geq 4$:

Theorem 7.2. *For any $k \geq 4$ there exist no IR efficient mechanism that makes it a dominant strategy for every hospital to not withhold any pair in the second stage.*

Proof. The proof is given for $k = 4$. The proof is similar proof for any $k > 5$. Consider the compatibility graph in Figure 6. Suppose that for each $i = 1, 2, 3$, a_i belongs to hospital h, and b_i belong to hospital h'. Exactly one of the 4 length cycles will be chosen. If the left exchange is chosen then h' is better off withholding b_1 and if the right one is chosen h is better off withholding a_3. ∎

Whether Theorem 7.1 holds under (requiring only Pareto efficiency) for $k = 3$ remains an open problem. $k = 3$ is in particular interesting for the kidney exchange since in practice mechanisms for kidney markets allow exchanges to be of size at most 3.

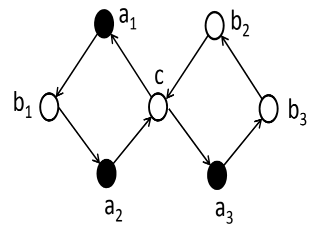

Figure 6.

8 Open Questions and Conclusion

We leave some interesting open questions: 1. Can the upper bounds we establish on the worst-case efficiency of individually rational kidney exchange mechanisms be achieved? (i.e. are these tight upper bounds?) 2. 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 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. It seems possible that in very large markets such patients can receive transplants with reduced cost to the total number of transplants.

Fewer than 1,000 transplants from kidney exchanges have been accomplished since the first kidney exchange in the United States in the year 2000, but well over half of those completed in the decade 2000-2009 were in 2008 and 2009, so kidney exchange is growing rapidly. As it grows, it faces new problems.

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 typically involve multiple hospitals that may have relatively little repeated interaction outside of kidney exchange. The present paper is meant to help establish a framework to study the kinds of problems that can be anticipated as the United States moves in the direction of nationally organized exchange.

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. The results suggest that, if care is taken in how kidney exchange mechanisms are organized, the problems of participation may be less troubling in very large exchange programs than they are observed to be in multi-hospital exchanges on the scale of those presently operating.

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, 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, 2010.

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.

P. Erdös and A. Rényi. On Random Graphs. I. *Publicationes Mathematicae*, 6:290–297, 1959.

P. Erdös and A. Rényi. On the existence of a factor of degree one of a connected random graph. *Acta Mathematica Hungarica*, 17:359–368, 1966.

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.

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

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.

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

8.1 Preparations

The following bounds will be useful in our proofs.

Lemma 8.1 (Chernoff bounds)**.** *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 $\delta \in (0, 1]$

$$\Pr\left(X < (1-\delta)np\right) < e^{\frac{-np\delta^2}{2}}. \tag{5}$$

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

$$\Pr\left(X > (1+\delta)np\right) < e^{\frac{-np\delta^2}{4}}. \tag{6}$$

Proof of Lemma 6.2:

Let $D(m)$ be a random compatibility graph and let $\delta > 0$. By Lemma 8.1 for every type X-Y

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

Therefore

$$\begin{aligned}
&\Pr\left[\text{for all X-Y} \in \mathcal{P},\ (1-\delta)m\mu_{\text{X-Y}} < Z_{\text{X-Y}} < (1+\delta)m\mu_{\text{X-Y}}\right] = \\
&\quad 1 - \Pr\left[\text{for some type X-Y} \in \mathcal{P} : Z_{\text{X-Y}} \notin ((1-\delta)m\mu_{\text{X-Y}}, (1+\delta)m\mu_{\text{X-Y}})\right] \geq \\
&\qquad 1 - \sum_{\text{X-Y} \in \mathcal{P}} \Pr\left[Z_{\text{X-Y}} \notin ((1-\delta)m\mu_{\text{X-Y}}, (1+\delta)m\mu_{\text{X-Y}})\right] = 1 - o(m^{-1}),
\end{aligned}$$

where the last inequality follows since there are a finite number of pair types. $\square$

Let $0 \leq \delta < 1$ and $m_0, m_1, \ldots, m_{l-1}$ be positive integers. For any integer $l \geq 0$, let $F_\delta^{(m_0,m_1,\ldots,m_{l-1})}$ be a distribution over l-tuples of integers that belong to $[(1-\delta)m_0, (1+\delta)m_0] \times [(1-\delta)m_1, (1+\delta)m_1] \times \cdots \times [(1-\delta)m_{l-1}, (1+\delta)m_{l-1}]$ and by $H_\delta^{(m_0,m_1,\ldots,m_{l-1})}$ a distribution over l-tuples of integers that belong to $[0, \delta m_0] \times [(1-\delta)m_1, (1+\delta)m_1] \times \cdots \times [(1-\delta)m_{l-1}, (1+\delta)m_{l-1}]$. We will write $F_\delta^m = F_\delta^{(m)}$.

Definition 8.2 (Uniformly Bounded Directed Random Graphs)**.** *A* ***uniformly bounded directed random graph*** *(UBDG) denoted by $D(m, p, \delta)$ and associated with a distribution F_δ^m is generated as follows: first the number of nodes is realized according to F_δ^m. Then for every two realized nodes v, w a directed edge is generated from v to w with probability at least p (note that for $\delta = 0$ the number of nodes is fixed).*

A ***l-partite random graph*** *(l-UBDG), denoted by $D(m_0, m_1, \ldots, m_{l-1}, p, \delta)$ and associated with distribution $F = F_\delta^{(m_0,m_1,\ldots,m_{l-1})}$ is generated as follows: First $l \geq 2$ distinct sets of nodes $V_0, V_1, \ldots, V_{l-1}$ are generated whose sizes are distributed according to F. Then for each $i = 0, 1, \ldots, l-1$, and each two nodes $v \in V_i, w \in V_{i+1}$ (i is taken modulo l) there is a directed edge from v to w with probability at least p, and between every two other nodes there exist no edge (again for $\delta = 0$, the size of each set V_i is deterministic).*

Finally a **quasi uniformly bounded directed l-partite random graph** *(l-QBDG), denoted by $\bar{D}(m_0, m_1, \ldots, m_{l-1}, p, \delta)$ is defined similarly to an l-UBDG only it is associated with a distribution $H_\delta^{(m_0,m_1,\ldots,m_{l-1})}$ rather then a distribution $F_\delta^{(m_0,m_1,\ldots,m_{l-1})}$.*

For clarity and brevity when there is no harm we will just refer to an l-UBDG and an l-QBDG by an l-partite graph. Note that in any l-partite graphs only exchanges of size $k = ql$ for positive integers q are feasible.

An allocation in $D(m, p, \delta)$ (for any $k \geq 2$) is **perfect** if it matches all but at most one node. Similarly an allocation in an l-partite graph $D(m_0, \ldots, m_{l-1}, p, \delta)$ is **perfect** if it matches all nodes in some set V_i for some $i \in \{0, \ldots, l-1\}$. Note that if there is a perfect allocation in $D(m_0, m_1, \ldots, m_{l-1}, p, 0)$ then it must match all nodes in a set V_i whose size is minimal.

Finally, **almost every** $D(m, p, \delta)$ has property Q if $P(Q) \to 1$ whenever $m \to \infty$ for any infinite sequence of distributions $F^1_\delta, F^2_\delta, \ldots$. For l-UBDG and l-QBDG graphs we use a similar definition requiring that for each $i = 0, \ldots, l-1$, $m_i \to \infty$.

Lemma 8.3. *Let $0 < p < 1$ and let $0 \leq \delta < 1$.*

1. *Almost every $D(m, p, \delta)$ contains a perfect allocation with $k = 2$ and an allocation that matches* all *nodes for any $k \geq 3$.*

2. *Let $0 < c_0 \leq c_1 \leq \cdots \leq c_{l-1} \leq 1$. Almost every $D(c_0 m, c_1 m, \ldots, c_{l-1} m, p, \delta)$ contains a perfect allocation which matches all nodes in some set V_i. Furthermore, if not all c'_i equal, $j' \geq 1$ is the least index for which $c_{j'} - c_{j'-1} > 0$ and $\delta < \frac{c_{j'} - c_{j'-1}}{c_{j'} + c_{j'-1}}$ then every perfect allocation matches all nodes in some V_i for some $i < j'$.*

3. *Let $0 < c_1 \leq \cdots \leq c_{l-1} \leq 1$ and let $\delta < \frac{c_1}{1+c_1}$. Almost every $\bar{D}(m, c_1 m, \ldots, c_{l-1} m, p, \delta)$ contains a perfect allocation which matches all nodes in V_0.*

Proof. First not that it is sufficient to prove the result for exact p rather than at least p (if the result holds for p than increasing the probability for the existence of some edges can only increase the probability of a perfect allocation). We begin with the first part. The proof for both $k = 2$ and $k \geq 3$ follows by applying the Erdos-Renyi Theorem to non-directed random graphs. We begin with $k = 2$. Let

$$p_m = \Pr\left[\text{there exists a perfect allocation in } \; G(m, p^2)\right].$$

Let $D(m, p, 0)$ be a UBDG. Since a cycle of length 2 has probability p^2 and because $k = 2$

$$\Pr\left[\text{there exist a perfect allocation in } D(m, p, 0)\right] = p_m.$$

Let δ be such that $\delta < 1$. We define a sequence $(x_m)_{m \geq 1}$ by choosing arbitrarily

$$x_m \in \arg \min_{x \in N \cap [(1-\delta)m, (1+\delta)m]} \Pr\left[\text{there exist a perfect allocation in } D(x, p, 0)\right]. \tag{7}$$

Note that the minimum is attained in some value since it is taken over a finite set. Therefore, for any distribution F_δ^m the graph $D(m, p, \delta)$ is associated with the following inequality holds:

$$\Pr\left[\text{there exist a perfect allocation in } D(m, p, \delta)\right] \geq \Pr\left[\text{there exist a perfect allocation in } D(x_m, \delta, 0)\right] = p_{x_m}.$$

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

We proceed to $k \geq 3$. Let $\tilde{m}$ be the realized number of nodes. Given that $\tilde{m}$ is even a perfect allocation can be found using only 2-way exchanges with probability $1 - o(1)$. Suppose $\tilde{m}$ is odd. Pick arbitrarily $\tilde{m}$-1 nodes and find an efficient allocation only with 2-way exchanges. Again, this can be found with probability $1 - o(1)$. Given such an allocation an efficient allocation exist if one can find a pair of nodes w and z that are matched to each other such that the single unmatched node can form a 3-way exchange with w, z. Such a pair of nodes v, z cannot be found with probability at most $(1 - p^2)^m$. This completes the first part.

The second part will follow from a reduction to a bipartite random graph and applying the Erdos-Renyi Theorem. Note that it is enough to prove the result for $k = l$, i.e. that there exist an efficient allocation using exchanges of at most (hence exact) size l.

Let be $D(m, m, \ldots m, p, 0)$ be an l-UBDG and let $V_0, V_1, \ldots, V_{l-1}$ be the sets of nodes in the graph as in Definition 8.2. For each $i = 0, \ldots, l-1$ and $j = 1, \ldots, m$ let $v_{i,j}$ be the j-th node in set V_i. We construct a bipartite graph $B(m, p^l)$ (with sets of nodes V and W) as follows. Let $V = V_0$ and for every $j = 1, \ldots m$, let the tuple $(v_{1,j}, v_{2,j}, \ldots, v_{l-1,j})$ be a node in W. (see Figure 8).

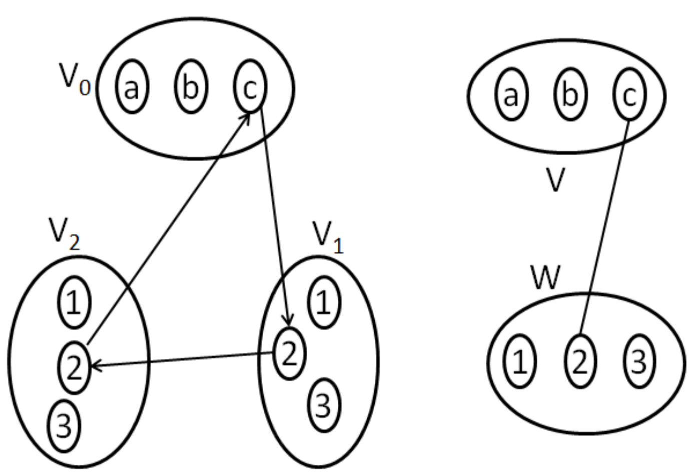

Figure 8: The graph on the left is a directed 3-partite graph and a the cycle corresponds to the non-directed edge on the bipartite graph on the right.

Let

$$q_m = \Pr\left[\text{there exists a perfect allocation in } \ B(m, p^l)\right].$$

Fix some $1 \le j \le m$ and some $v \in V_0$. Observe that the probability that $D(m, m, \ldots, m, p, 0)$ contains the cycle $v, v_{1,j}, v_{2,j}, \ldots, v_{l-1,j}$ is p^l. Moreover the probability that there exist an edge between $(v_{1,j}, v_{2,j}, \ldots, v_{l-1,j})$ and v is also p^l (see Figure 2). Therefore

$$\Pr\left[\text{there exist a perfect allocation in } D(m, m, \ldots, m, p, 0)\right] \ge q_m.$$

By only adding more nodes adding more nodes to sets of nodes while keeping the size of the smallest set

$$\Pr\left[\text{there exist a perfect allocation in } D(c_0 m, c_1 m, \ldots, c_{l-1} m, p, 0)\right] \ge \\ \Pr\left[\text{there exist a perfect allocation in } D(c_0 m, c_0 m, \ldots, c_0 m, p, 0)\right].$$

We define a sequence $(x_m)_{m \ge 1}$ by choosing arbitrarily

$$x_m \in \arg \min_{x \in N \cap [(1-\delta)m, (1+\delta)m]} \Pr\left[\text{there exist a perfect allocation in } D(x, x, \ldots, x, p, 0)\right]. \quad (8)$$

Therefore, for any $F_\delta^{(c_0 m, c_1, \ldots, c_{l-1} m)}$ the graph $D(c_0 m, c_1 m, \ldots, c_{l-1} m, p, 0)$ is associated with the following inequality holds:

$$\Pr\left[\text{there exist a perfect allocation in } D(c_0 m, \ldots, c_{l-1} m, p, \delta)\right] \ge \\ \Pr\left[\text{there exist a perfect allocation in } D(x_{c_0 m}, \ldots, x_{c_0 m}, p)\right] \ge q_{x_{c_0 m}}.$$

As in part one, $q_{x_{c_0 m}} \to 1$ as $m \to \infty$ by the Erdos-Renyi Theorem. We obtained that almost every graph $D(c_0 m, c_1 m, \ldots, c_{l-1} m, p, \delta)$ contains a perfect allocation.

Finally, if $c_0 = c_1 = \cdots = c_{l-1}$ we are done. Otherwise let j' be as in the hypothesis. Since $\delta < \frac{c_{j'} - c_{j'-1}}{c_{j'} + c_{j'-1}}$

$$(1-\delta) c_{j'} m > (1-\delta) c_{j'-1} m.$$

Therefore since each exchange contains exactly l pairs in every perfect allocation, for some $j \le j'$ all pairs in V_j are matched.

We proceed to the third part. Observe that if a l-partite graph contains a a perfect allocation then it also contains one after removing some nodes from the smallest set V_i. Therefore using similar arguments as in the second part one can show that almost every $\bar{D}(m, c_1 m, \ldots, c_{l-1} m, p, \delta)$ contains a perfect allocation, and since $\delta < \frac{1}{c_1}$, V_0 is the smallest set and all its nodes will be matched. ∎

8.2 Proofs

Proof of Theorem 6.3:

Let $D(m)$ be a random compatibility graph. Let B_δ be the event that $(1-\delta)m\mu_{\text{X-Y}} \leq Z_{\text{X-Y}} \leq (1+\delta)m\mu_{\text{X-Y}}$ for every type X-Y$\in \mathcal{P}$ and let Q be the event in which $Z_{A-B} \geq Z_{B-A}$, i.e. there are more A-B pairs than B-A pairs. Fix δ to be some constant $0 < \delta < \min\{\mu_{AB}^3, \frac{1-2.5\bar{\gamma}}{1+2.5\bar{\gamma}}\}$.

Given that Q occurs an upper bound on the size of an efficient allocation in $D(m)$ consists of the following two events (see also Roth et al. (2007)).

E_1: there exists an allocation at which every self demanded pair is matched either through a 2-way or through a 3-way exchange containing only self demanded pairs, and

E_2: there exist an allocation that matches every B-A pair in a 2-way exchange to a A-B pair; every *other* A-B pair is matched in 3-way exchange using a B-O pair and a O-A pair; every AB-O pair is matched in a 3-way exchange using a O-A pair and a A-AB pair; all remaining over-demanded pairs X-Y are matched in a 2-way exchange to a Y-X pair.

If Q does not occur then a similar then an upper bound for the size of an efficient allocation in $D(m)$ consists of similar sequence of events only each B-A pair that is not matched in a 2-way exchanges with a A-B pair is matched in a 3-way exchange together with a A-O and a O-B pair.

It is sufficient to show that $\Pr(E_1, E_2|Q) = 1 - o(1)$; a similar argument will imply that $\Pr\left(E_1, E_2|\bar{Q}\right) = 1 - o(1)$ which completes the proof.

Let Q_δ be the event that $0 \leq Z_{\text{A-B}} - Z_{\text{B-A}} < 2m\delta$. Note that $Q_\delta \subseteq Q$. By Lemma 6.2 $\Pr(B_\delta, Q_\delta) = 1 - o(m^{-1})$. Therefore it is sufficient to show that

$$\Pr(E_1, E_2|B_\delta, Q_\delta).$$

Therefore throughout the entire proof we will assume that both event B_δ and Q_δ occur (the probability that at least one of these events does not occur is very low). Observe that the event E_1 and E_2 are independent. Fix some self demanded type X-X. The graph induced by only X-X pairs is a UBDG graph $D(\mu_{\text{X-X}}m, \delta)$. Therefore

$$\Pr(E_1|B_\delta) = \Pr[\text{there exist a perfect allocation in } D(\mu_{\text{X-X}}m, \delta)],$$

which equals $1 - o(1)$ by the first part of Lemma 8.3.

It remains to show that $\Pr(E_2|B_\delta, Q_\delta) = 1 - o(1)$. Our proof will be constructive. We will consider a sequence of subgraphs and show that each one them contains a desired perfect allocation (see Figure 3).

The following partitions will be useful for the subgraphs to be considered. Partition the set of O-A pairs into two sets, $W_{\text{O-A}}$, $V_{\text{O-A}}$, such that $|W_{\text{O-A}}| = (1+\delta)\mu_{\text{AB-O}}m$; the set of

A-AB pairs into the sets $W_{\text{A-AB}}$ and $V_{\text{A-AB}}$ such that $|W_{\text{A-AB}}| = (1+\delta)\mu_{\text{AB-O}}m$; and finally the set of O-A pairs into the sets $W_{\text{O-A}}$ and $V_{\text{O-A}}$ such that $|W_{\text{O-B}}| = (1+\delta)\mu_{\text{B-O}}m$. The feasibility of these partitions follows from the following claim:

Claim 1

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

Proof: Observe that

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

where the last inequality follows since $\mu_{AB} < \mu_A$ and $\delta < \frac{1-2.5\gamma}{(1+2.5\gamma)} < \frac{1-2\gamma}{2(1+2\gamma)}$, completing the first part. Note that

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

where the last inequality follows since $\mu_O + \mu_A < 2.5\mu_A$ and $\delta < \frac{1-2.5\gamma}{(1+2.5\gamma)}$, implying the second part. Similarly the third part follows since

$$\mu_{O-A}(1-\delta)m = \rho\mu_O\mu_A(1-\delta)m > \rho\mu_O\bar{\gamma}(\mu_A + \mu_B)(1+\delta)m.$$

□

We are now ready to construct the efficient allocation. Consider first the subgraph induced by all B-A pairs and all A-B pairs, and denote this graph by D_1. Observe that D_1 is a 2-UBDG graph. Denote by M_1 a random efficient allocation in this graph and denote by $V_{\text{A-B}}$ the random set of pairs that are not matched by M_1. Similarly consider the random graph (its edges are random) denoted by D_2, induced by the set of pairs $V_{\text{A-B}}$, B-O pairs and by the set of pairs $W_{\text{O-A}}$, and let M_2 be a random efficient allocation in this graph.

We will show that the following probability is small:

$$\Pr\left[M_2 \text{ matches all pairs in } V_{\text{A-B}} \mid M_1 \text{ matches all B-A pairs}\right] \Pr\left[M_1 \text{ matches all B-A pairs}\right]. \quad (9)$$

By Lemma 8.3 the second term in (9) equals $1 - o(1)$. Given that Q_δ occurs and that all B-A pairs are matched in M_1, $|V_{\text{A-B}}| \leq 2m\delta$. Therefore the graph D_2 is a 3-QBDG $\bar{D}(m, \mu_{\text{B-O}}m, \mu_{\text{B-O}}m, 1-\gamma_H, \delta)$ and since $m\delta < m\mu_{\text{B-O}}(1+\delta)$ the first term in (9) equals $1 - o(1)$ by Lemma 8.3. This implies that $(9) = 1 - o(1)$.

Denote by $V_{\text{B-O}}$ the sets of B-O pairs that are not matched by M_2. Consider the following random graphs. Let D_3 be the graph induced by the sets of pairs $V_{\text{B-O}}$ and $V_{\text{O-B}}$; Let D_4 be the graph induced by all AB-O pairs and the sets $W_{\text{O-B}}$ and $W_{\text{B-AB}}$; D_5 be the graph induced by A-O pairs and the set $V_{\text{O-A}}$; Let D_6 be the graph induced by the AB-B pairs and the set $V_{\text{B-AB}}$, and finally let D_7 be the graph induced by AB-A pairs and A-AB pairs.

Observe that each of the graphs $D_3, \ldots, D_7$ is either a 2-UBDG or 3-UBDG graph and by Lemma 8.3 for each $i = 3, \ldots, 7$ almost every D_i contains a perfect allocation.

It remains to show that for $k \geq 4$ one cannot obtain more transplants than for $k = 3$. happens). Following Roth et al. (2007b), the only possible 4-way exchange that may possibly increase the number of transplant is the one in Figure 4. Such an exchange uses an AB-O pair and a A-B pair that is not matched to a B-A pair. Observe that all of these pairs are matched in $k = 3$ in 3-way exchanges implying that using such a 4-way will result in less transplants. □

Proof of Theorem 6.6:

The proof is similar to the proof of Theorem 6.3. The main difference is that in this proof we will need to choose more carefully the nodes of some of the induced subgraphs. For simplicity we will assume that also hospitals exchanges are limited to size 3.

Let E_1 and E_2 be as in the proof of Theorem 6.3 with the following modifications: in E_2 we allow every AB-O pair to be either matched in a 3-way exchange as described or in a 2-way exchange with an AB-X pair for any $X \neq O$, and A-B and B-A pairs are matched only in two way exchanges using another reciprocally type. It is sufficient to show that almost every graph has an individually rational allocation allocation that satisfies both E_1 and E_2.

First we handle E_1 similarly as in the proof of Theorem 6.3. We next deal with the AB-O pairs. Let M_h be an efficient allocation in the graph induced by the pairs of hospital h. Let V_{AB-O} the set of all O-AB, O-A and O-B that are matched under $\cup_{h \in H} M_h$ through exchanges that do not involve either A-O or B-O pairs.[18] If V_{AB-O} is smaller than the number of AB-O pairs then we add to this set O-AB pairs until we obtain the same number as the AB-O pairs (note that this is possible with high probability $1 - o(m^{-1})$. Consider the graph induced by the set of AB-O pairs and the set V_{AB-O}. This is a 2-UBDG graph and there exist a perfect matching with probability $1 - o(m^{-1})$.[19]

Next we find a perfect matching in the graph induced by only A-B and B-A pairs (again as in the proof of Theorem 6.3). Denote by W_{O-A} the set of all O-A pairs that are matched under $\cup_{h \in H} M_h$ and do not belong to V_{AB-O}. Note that $|W_{O-A}|$ is smaller than the number

[18]Here we use use the assumption that hospitals have exchanges of size at most 3.

[19]Here we constructed in a worse case manner; one might match a large fraction (if not all) of AB-O pairs using 3-way exchanges with where each exchange involves one pair from the set V_{AB-O} and one A-AB pair or one B-AB pair.

of A-O pairs with high probability since the hospitals have a non-critical size. We add to W_{O-A} an arbitrarty set of nodes such that the size has the same cardinality as the number of the set of A-O pairs. Consider the graph induced by the set of O-A pairs and W_{O-A}. Note that this is a 2-UBDG graph and contains with high probability a perfect matching.

The proof proceeds similarly with all over demanded pairs obtaining the desired result. □