A Random Graph Model of Multi-Hospital Kidney Exchanges[☆]

Panos Toulis [a]

Department of Statistics, Harvard University,
1 Oxford St, Cambridge, MA, 02138

David C. Parkes[b]

SEAS, Harvard University,
33 Oxford St, Cambridge, MA, 02138

[a] *ptoulis@fas.harvard.edu*
[b] *parkes@eecs.harvard.edu*

Abstract

In kidney-paired donation, a patient can enter as a pair, with a donor, into an exchange and hope to be matched with a compatible kidney donor. In the simple case, matches can be made through 2-cycles (swaps) or 3-cycles, with the donor of each matched pair donating a kidney to a compatible patient of another matched pair. One challenge is to align incentives so that multiple hospital systems will choose to pool their local pairs into a larger exchange. Adopting a random-graph model, with the compatibility structure between pairs generated according to a probabilistic model of blood-type and tissue-type compatibility, we first establish a *square-root* law for the increase in transplants due to pooling as a function of the number of pairs in individual hospital pools. Second, we propose a Bayes-Nash incentive compatible, efficient exchange for 2-cycles and 3-cycles, under idealized assumptions. We validate the robustness of the design through a computational study, which demonstrates incentive-alignment and an efficiency loss of less than 1% or

[☆]This is the full version of an extended abstract that appeared in ACM EC'11. In addition to providing complete proofs, we introduce a complete statistical model that enables deeper insights and a natural extension to 3-way exchanges. Thanks to Itai Ashlagi, Ian Kash, Felix Fischer, Les Valiant, the anonymous reviewers from EC'11 and GEB, and the Associate Editor at GEB, for their valuable feedback on earlier versions of this work.

5%, with 2-cycle or 3-cycle exchanges, respectively.

1. Introduction

The scarcity of cadaver kidneys and significant medical benefits from live kidney donation has prompted the advancement of kidney-paired donation (KPD) in recent years [8]. In KPD, a patient with an incompatible donor can form a donor-patient pair, and by entering an exchange be matched such that the patient receives a compatible kidney and the donor donates his or her kidney to some other patient. In the simplest case this occurs through a 2-cycle (or swap), and involves the transplant of two kidneys and four simultaneous operations. Longer cycles, and in particular 3-cycles, are also practical, but beyond 3-cycles the logistics of the large number of simultaneous operations become difficult.[1] Kidney exchange programs exist around the world [14], although their large-scale expansion has been hindered by ethical, logistical and incentive issues [12, 4].

The matching problem can be modeled as a graph, with each donor-patient pair represented as a vertex and edges representing compatibility relationships between pairs. When identifying maximum matches composed of 2-cycles (or swaps), an undirected graph suffices with an edge between two pairs indicating mutual compatibility between donors and patients. Assuming all transplants provide the same value, the maximum cardinality match is welfare maximizing.[2] When allowing for simultaneous transplants of 3 kidneys (and thus looking for 2- or 3-cycles), a directed graph represents the matching problem with an edge from one pair to another to indicate that the donor of the first pair is compatible with the patient of the second pair. The welfare maximizing outcome is a set of maximum cardinality vertex-disjoint 2- or 3-cycles on this directed graph.

Multi-regional KPD programs have been recently established in the US [11], providing patients with access to larger pools of paired donations.[3] The main

[1] Since 2007, a growing pattern is to adopt *non-simultaneous altruistic donor chains*, in which an altruistic donor with no designated recipient can initiate a chain of transplants. For example, a chain of 10 kidney transplants was performed over 8 months from July 2007 [10]. The focus of our paper is on transplants that occur through simultaneous cycles rather than non-simultaneous altruistic donor chains.

[2] In practice, **xx weights and preferences.**

[3] At present, the largest exchange in the US is the National Kidney Register (NKR),

advantage that comes from larger exchanges is that market thickness can enable additional transplants to be identified. We are interested in providing an analysis of this gain that can arise from combining pools. A challenge that arises is that a hospital can be reasonably considered to be self-interested entities, seeking to maximize the number of its own patients that receive a transplant. For this reason, there is an interest in incentive compatible exchanges, in which a hospital maximizes the number of its own pairs involved in a matching by revealing its complete donor-patient list.

xx need to flesh this out This incentive challenge was first discussed by Unver and xx check, who xx. Ashlagi et al. [2] adopted the same example to prove (for two hospitals and 2-cycles) that no deterministic, dominant-strategy incentive compatible (DSIC) mechanism can obtain an approximation factor better than 2 on all possible instances, meaning that there exists instances where *any* such mechanism must forfeit at least half the matches possible in the welfare optimal outcome. The same authors presented a xx complete [and the guy from Columbia's paper]

Rather than adopt a worst-case analysis, we adopt an expected-case analysis and a random graph model, where the edge structure of this graph is generated according to a probabilistic model of the compatibility between donor-patient pairs that enter KPD exchanges, namely, blood-type and tissue-type compatibility. Our analysis makes the simplifying assumption of uniform probability of tissue-type compatibility, which has also been adoped in earlier papers [1, 13, 15], and indeed, our analytical model is able to explain many of the quantitative results in these papers. More significantly, it provides under an idealized model a simple *square-root law*, to characterize the welfare improvement obtained from a pooled, national exchange over separate exchanges, for matches comprised of 2-cycles. The expected number of additional transplants enabled per-hospital through the use of a centralized exchange scales according to the square-root of the number of pairs of each hospital **xx check at end (see Theorem 2, Corollary 2).** We prove that going from 2-cycles to 3-cycles suffices to maximize the expected number of transplants, in the same idealized model, such that cycle lengths greater than 3 are not required. **xx relate to earlier work. I think RSU already**

which has a pool of in excess of 500 pairs. Other large exchanges are operated by Alliance for Paired Donation (APD) and the United Network for Organ Sharing (UNOS). As of 2011, more than 700 transplants were made possible through the UNOS Organ Procurement and Transplanation Network.

showed that 3-cycles are all that are needed in a simple model? xx xx think about this some more In this light, the significance of moving from multiple separate exchanges to a single exchange for 3-cycles is to hasten the point at which the idealized assumptions hold with high probability.

Turning to mechanism design, we provide a welfare optimal, Bayes-Nash incentive compatible mechanism for 2-cycles under idealized conditions that we justify through simulations and theory in our domain. The `xCM` mechanism first determines a matching with a *regular* structure local to the pairs reported by each hospital (or an almost regular matching if there is no regular matching) and then augments this matching with a random maximum match on the collection of the remainders from each hospital.[4] Our technical results rely on a structural analysis of maximum matchings under the random-graph model. We extend the design to allow for 3-cycles, obtaining the $\texttt{xCM}^3$ mechanism, which is welfare optimal and BNIC under generalized regularity and bipartite perfect-matching assumptions.

In a computational study, we first contrast the efficiency and incentive properties of `xCM` with that of a simple, non incentive-compatible design and confirm that a self-interested hospital could achieve approximately 20% more transplants on average than a truthful hospital without careful mechanism design. **xx for what size pools?** Second, we confirm the robustness of the mechanism to a more realistic domain model with non-uniform tissue-type incompatibility. Although not perfectly incentive aligned, we show that the `xCM` mechanism achieves an average efficiency of 98% (se=1%) when there is a single self-interested hospital and two truthful hospitals. When all hospitals deviate, the `xCM` mechanism incurs a constant loss of an average of 2 matches, regardless of the underlying graph size. Considering 3-cycles, we verify that $\texttt{xCM}^3$ leaves essentially no incentive for a hospital to deviate across a range of pool sizes. Significanly, the idealized (generalized regularity) assumption appears to be a good approximation for larger pool sizes, with $\texttt{xCM}^3$ obtaining efficiency of 95% for hospital pools with 200 donor-patient pairs, compared to 81% efficiency in $\texttt{xCM}^3$ for hospitals with 40 local pairs.

Ashlagi and Roth [3], independently of our work, also adopt a random graph model and study the problem of multiple-hospital exchanges. They make a different regularity condition, assuming that the expected "utiliza-

[4]The same efficiency could be obtained without randomization in this second step, but we adopt randomization in order to provide ex ante fairness across participating hospitals.

tion" in maximum matchings of every under-demanded type X-Y, is higher than the expected "abundance" of the Y-X over-demanded type. Based on this, a maximum matching will be highly likely to match the over-demanded pairs to under-demanded ones. In comparison, our regularity assumption requires the existence of a maximum matching with a particular structure in every instance of the donor-patient random graph. **xx update at end xx** Our assumptions are validated for hospital pools as small as 20 donor-patient pairs for 2-cycles, but require larger pools such as 200 donor-patient pairs for 3-cycles given standard blood and tissue-type models. Similarly, the regularity assumption in Ashlagi & Roth is shown experimentally to hold well for pools with 100 or more pairs **xx 2 or 3 cycle?**. The mechanism proposed by Ashlagi and Roth is different from the one proposed here, and approximately efficient and BNIC in their idealized model. Ashlagi and Roth do not establish a result analogous to the square-root law in the present paper.

2. Preliminaries

Blood-type compatibility, ABO compatibility for short, is the first requirement for a kidney transplant. Tissue-type compatibility is the second requirement. Blood type is one of $\mathcal{B} = \{O, A, B, AB\}$, indicating the presence or absence of proteins A and B. Two people are blood-type compatible if the donor does not introduce a new protein to the patient. For example, a donor with blood type A can donate to patients with type A or AB and a donor with type O can donate to any blood type. We associate the donor-patient pairs with four distinct subsets based on ABO type, namely *under-demanded* UD, *over-demanded* OD, *reciprocal* R and *self-demanded* S (see Table 1) [15]. We denote the set of the types as $\mathcal{T} = \{UD, OD, S, R\}$. Donor-patient pairs of type SD and OD are ABO-compatible, and so we only expect to see these pairs in an exchange if they are tissue-type incompatible.

In regard to tissue-type incompatibility, this requires that a donor and patient share as many human leukocyte antigens as possible, to prevent a positive crossmatch and organ rejection. The probability of a positive crossmatch (and thus incompatibility) is variable across donor-patient pairs, and depends on the tissue-type sensitivity of the patient [14].[5] In developing our

[5]The Panel Reactive Antibody (PRA) of a patient measures the percentage of the population with whom the patient will be tissue-type incompatible. Higher PRA values result in a higher probability of incompatbility. Roughly speaking, PRA is low, medium

patient	O	A	B	AB
donor				
O	S	OD	OD	OD
A	UD	S	R	OD
B	UD	R	S	OD
AB	UD	UD	UD	S

Table 1: The association of donor-patient blood types with classes under-demanded UD, reciprocally demanded R, over-demanded OD and self-demanded S.

theoretical results we adopt the common assumption of uniform probability of tissue-type compatibility, and adopt the probability of positive crossmatch $p_c = 0.2$.

In summary, we adopt the following two basic modeling assumptions:

(a) the blood type of an individual participant (patient or donor) is i.i.d. sampled from a blood-type distribution,

(b) the tissue-type compatibility of any patient for any donor is an independent Bernoulli trial with fixed probability $1 - p_c$.

In general, the O blood group is the most common, with A, B and AB each less common in turn. In particular, for our numerical results we assume a distribution of O (50%), A(30%), B(15%) and AB(5%), which is close to the world average [16].[6]

2.1. The Random Graph Model: 2-cycles

For exchanges that utilize only 2-cycles (or swaps), we adopt an undirected graph to model the compatibility between pairs and thus ability to match. Donor-patient pairs are vertices and an undirected edge between two vertices indicates that a match is possible between the two pairs. A *random*

or high, corresponding to a probability 0.05, 0.45 or 0.9 of a positive cross match (and incompatibility) with a random donor, and with 70.2%, 20% and 9.8% of the population low, medium and high PRA respectively.

[6]Blood-type distribution among human population is quite variable around the world [16]. For example, Koreans are evenly distributed in types O, A and B (with less AB) while the Bororo people in Brazil are 100% of type O.

donor-patient graph with n pairs is denoted by $\widetilde{G_n}$, and is generated according to the generative process in Algorithm 1. We denote a donor-patient pair as u, where u_d, u_p is the donor and patient respectively, and adopt F_{ABO} to denote the distribution on ABO types. Predicate $ABO(x, y)$ is true when participants x and y are ABO compatible, and false otherwise. Last, we denote with U an independent draw from the uniform distribution on [0,1]. A new pair is introduced if the donor-patient pair is ABO incompatible or tissue-type incompatible. For any two ABO-compatible pairs in the pool, they are (mutually) tissue-type compatible with probability $(1-p_c)^2 = 0.64$, since each donor has to be compatible with the other patient.

Algorithm 1 Generative process of undirected donor-patient graph on n pairs

```
G ← empty graph
while |V(G)| < n do
  u ~ F_ABO  // sample pair from blood type distribution
  if not ABO(u_d, u_p) or (ABO(u_d, u_p) and U ≤ p_c) then
    V(G) ← V(G) ∪ {u}
  end if
end while
for distinct u, v ∈ V(G) do
  if (ABO(u_d, v_p) and ABO(v_d, u_p) and U ≤ (1 − p_c)^2) then
    E(G) ← E(G) ∪ (u, v)
  end if
end for
```

xx what do we want to call a component, a subcomponent and a subgraph ? xx

It is useful to identify subcomponents in $\widetilde{G_n}$, corresponding to types in $\mathcal{T}$. Let q_x denote the probability that a random vertex in $\widetilde{G_n}$ belongs to the $x \in \mathcal{T}$, and f_y the probability of blood type y, for $y \in \mathcal{B}$. By elementary analysis, we have:

$$q_{OD} = \frac{f_{OD} \cdot p_c}{V},\ q_S = \frac{f_S \cdot p_c}{V},\ q_{UD} = \frac{f_{UD}}{V},\ q_R = \frac{f_R}{V}, \tag{1}$$

where $f_{OD} = f_{\mathrm{O}}(f_{\mathrm{A}} + f_{\mathrm{B}}) + f_{\mathrm{AB}}(f_{\mathrm{O}} + f_{\mathrm{A}} + f_{\mathrm{B}})$, $f_{UD} = f_{OD}$, $f_S = f_{\mathrm{O}}^2 + f_{\mathrm{A}}^2 + f_{\mathrm{B}}^2 + f_{\mathrm{AB}}^2$, and $f_R = 2f_{\mathrm{A}} \cdot f_{\mathrm{B}}$, and denominator $V = f_{OD} \cdot p_c + f_S \cdot p_c + f_{UD} + f_R$ provides normalization. This gives the relative expected size of each

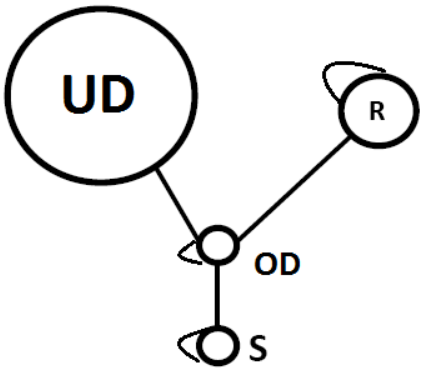

Figure 1: Subcomponents representing different ABO-types and the relative expected frequency of types in an exchange. The absence of an edge indicates that no pair in the incident subcomponents can be ABO-compatible.

subcomponent, with $f_{OD} = f_{UD} = 0.2725$, $f_S = 0.365$, and $f_R = 0.09$, and together with $p_c = 0.2$ we have, $\frac{q_{OD}}{q_{UD}} = p_c = 0.2$ and $\frac{q_S}{q_{UD}} \approx 1.34 \cdot p_c \approx 0.27$, and $\frac{q_R}{q_{UD}} = \frac{f_R}{f_{UD}} \approx 0.33$.

Based on this, the high-level structure and relative expected size of subcomponents is illustrated in Figure 1. The absence of an edge (including the absence of a self-edge) between a pair of subcomponents indicates that there can be no pairs, one in each subcomponent, that are ABO-compatible. For example, a donor-patient A-O pair (UD) can match with a donor-patient O-A pair (OD) but no UD pair can match with a R pair.

2.2. The Random Graph Model: 3-cycles

For exchanges that allow both 2-cycles and 3-cycles the KPD matching problem can be modeled as a directed graph. Each vertex continues to correspond to a donor-patient pair, and there is a directed edge (u, v) from u to v if the donor of u is compatible with the patient of v. For an example of a 3-cycle, consider the possibility of A-B $\rightarrow$ O-A $\rightarrow$ B-O and then back to A-B, or cycle B-A $\rightarrow$ O-B $\rightarrow$ A-O and then back to B-A. The generative model can be easily modified: the same generative process determines which pairs enter, and for every distinct pair of vertices u and v in the graph with $ABO(u_d, v_p)$, a directed edge is introduced from u to v with probability $(1 - p_c)$. Let $\widetilde{G}_n^3$ denote a random (directed) donor-patient graph with n pairs.

3. A Square-Root Law for the Benefit from Pooling

3.1. Idealized Model

The maximum matching on an undirected graph is a set of edges with no common vertices and maximum cardinality. A *perfect matching* is a matching that includes every vertex. We are interested in regular matchings:

Definition 1 (Regular Matching). A regular matching M_n on graph $\widetilde{G_n}$ is a matching in which every *OD* pair is matched with some *UD* pair, and the S and R subcomponents assume a maximum matching internally.

Note that a regular matching need not always exist because it is possible that there are some *OD* pairs that cannot be matched to *UD* pairs. We return to this below. For now, we prove the following:[7]

Lemma 1. *A regular matching M_n on graph $\widetilde{G_n}$ is maximum if it exists.*

Proof. If a regular matching M_n exists but is not maximum, then there has to be an *augmenting path* by Berge's lemma [9]. An augmenting path is a path that starts and ends on an unmatched vertex and alternates between edges that are are absent from the matching and edges that are included in the matching. Suppose there exists an augmenting path given a regular matching. Certainly, any such path must involve an *OD* vertex v because the path cannot be internal to S or R (the matching is maximum on S and R) and there are no edges to traverse within *UD*. Moreover, because all *OD* pairs are matched in M_n then v must not be the start of end of the augmenting path. In particular, the augmenting path through v must traverse an edge in the matching to a *UD* pair w. But w is matched and so the path must continue from w on an edge not in M_n back to some $v' \neq v$ in *OD*. Again, the path must traverse from v' back to a matched pair $w' \neq w$ in *UD*. The path must continue, zigging back and forth between *OD* and *UD* and eventually exhaust all pairs in *OD* before it is able to terminate. This establishes that there can be no augmenting path, completing the proof. □

We turn now to providing an analysis of when regular matchings exist. For this, the following is a useful property for random bipartite graphs $G_{n\times n,p}$

[7] The proof in Roth et al. [13] Proposition 1 is incomplete because it does not consider the possibility of an augmenting path in a matching. **xx dcp still to check this**

with n nodes in each part such that each node is connected to each node of the other part with uniform probability p. Erdos and Rényi [6] proved a stronger version (for $p = \frac{log(n)+c_n}{n}$):

Lemma 2. *Graphs in model $G_{n\times n,p}$ with any constant $p > 0$ assume a perfect matching w.h.p. for sufficiently large n.*

We will use this to understand the structure of maximum matchings within the S and R subcomponents and also in establishing the existence of a regular matching in large graphs:

Lemma 3. *A regular matching in $\widetilde{G_n}$ exists w.h.p. as n goes to infinity.*

Proof. In establishing this, we first establish a property about the number of *OD* pairs relative to the number of *UD* pairs.

Lemma 4. *Consider X-Y any OD pair type and Y-X the corresponding UD pair and adopt N_{X-Y} and N_{Y-X} to denote the number of each pair type in random graph $\widetilde{G_n}$. Then,* $\Pr(N_{X-Y} \leq N_{Y-X}) = 1 - O(n^{-2})$ *and* $\lim_{n\to\infty}[\Pr(N_{X-Y} \leq N_{Y-X})] = 1$.

Proof. Denote with q the probability that a random pair in some $\widetilde{G_n}$ will be *UD* type Y-X. Given a graph with n pairs, our generative model has $N_{Y-X} \sim Bin(n, q)$, and for the symmetric *OD* pair $N_{X-Y} \sim Bin(n-N_{Y-X}, q\cdot p_c)$, sampled from the remainder hospital size once the number of N_{Y-X} is known. **xx don't get why there is an asymmetry here. why not have *OD* be $N_{X-Y} \sim Bin(n, q \cdot p_c)$?** From this, we have that random variable $D \equiv (N_{Y-X} - N_{X-Y})$ has mean and variance $\propto n$. Denote this as $D \sim [an, bn]$ for some constants $a, b > 0$. By Chebyshev's inequality $P(|D - an| \geq k \cdot bn) \leq 1/k^2$ for any $k > 0$. Set $k = (a/b) \cdot n$, then we get: $P(|D - an| \geq an) \leq (b/a)^2 \cdot n^{-2}$. But it holds that $P(D < 0) < P(|D - an| \geq an)$ and therefore $P(D < 0) < (b/a)^2 \cdot n^{-2}$ which means that $P(D < 0) \to 0$ when n increases. We thus proved that $P(N_{Y-X} \geq N_{X-Y}) \to 1$ for n increases. □

By Lemma 4, we obtain that any *OD* type will have fewer pairs than its corresponding *UD* type, in the large n limit. Let us now consider a random bipartite graph involving an *OD* type in one part and its corresponding *UD* type in the other part, and with each vertex in one part connected to each vertex in the other part with uniform probability $(1 - p_c)^2 = 0.64$. Based

on this, and using Lemma 2, a regular matching exists w.h.p in the random patient-donor graph for large enough n. □

Having established the existence of regular graphs in the large n limit, we make two strong, simplifying assumptions, which together comprise what we refer to as an *idealized* model:

Assumption 1 (Regularity)**.** *For every $\widetilde{G_n}$, there exists a regular matching.*

Assumption 2 (Perfect matching (PM))**.** *For every $\widetilde{G_n}$*

(1) the bipartite subcomponent R has a perfect matching when there is an equal number of A-B and B-A pairs;

(2) each subcomponent of the S subgraph (e.g., the subcomponents of A-A pairs, B-B pairs, O-O pairs and AB-AB pairs) has a perfect matching when there is an even number of pairs in the subcomponent.

xx subcomponent to subgraph?

To validate the Regularity assumption, there exists a regular matching with probability 0.85, **xx and xx** in graphs with 40, xx and xx pairs, and **xx don't follow: regular matchings are maximum!!** there is a maximum matching that differs from a regular matching by only a single swap with probability 0.98. Allowing for non-uniform tissue-type compatibility, the corresponding probability (for 40 pairs) is 0.71. **xx more graph sizes**

xx don't follow the next: regular are maximum by construction! Last, the probability $P_r(n)$ that a regular matching is also maximum is about 82% for graph sizes of $n = 20$ pairs, and increases almost linearly with the graph size. A typical linear regression of $P_r(n)$ over the # pairs n gives $P_r(n) = 0.04 \cdot n + 83.28$ with $R^2 = 95\%$ and p-value=3.1×10^{-5} for the regression coefficient. By extrapolation, the probability that a regular matching is also maximum reaches 99% when $n \approx 400$ pairs, and this is also supported by our simulations. These probabilities are even higher if we compute near-optimal regular matchings, i.e. considering regular matchings having at most 1 or 2 matches fewer than a maximum one (e.g. for $n = 20$ pairs the corresponding probability is equal to 97.4%). **xx**

In regard to the PM assumption, we can verify in simulation that the probability of a perfect matching for a bipartite subcomponent R with an equal number of pairs in each part and $p_c = 0.2$ (and thus the probability of an edge between any vertex in one part and a vertex in the other part of $p = (1 - p_c)^2 = 0.64$) is 0.985 and 0.999 respectively for 4 and 6 pairs in each

part, respectively. For subcomponent S, these probabilities are even higher since it is not bipartite, and equal to 1 for practical numerical purposes.

xx need related all validation wrt graph size not size of subcomponents

3.2. The Expected Size of Maximum Matchings

The size (or cardnality) of a matching is the number of vertices incident to edges in the matching. The following results quantify the expected size of a maximum matching in subcomponents R and S under the PM assumption.

Lemma 5. *The expected cardinality of a maximum matching M_s on random subcomponent $\widetilde{S}$ induced by random graph $\widetilde{G_n}$ is*

$$E[|M_S|] = |\widetilde{S}| - 2, \tag{2}$$

under the PM assumption, where $|\widetilde{S}|$ is the number of vertices in $\widetilde{S}$ and $|M_S|$ is the size of matching M_S.

The proof of this result is technical and deferred to the appendix. We prove this through recursion, but a more intuitive idea can be given here. Recall that S has four components and each has 0 or 1 pair unmatched depending on whether that component has an even or odd number of pairs. Therefore, the number of unmatched pairs can be throught of as a binomial random variable $Bin(4, 0.5)$ and this has an expectation of $4 \cdot 0.5 = 2$.

Lemma 6. *Consider a random $\widetilde{R}$ subcomponent corresponding to $\widetilde{G_n}$. Let M_R denote a maximum matching of R. Under the PM assumption, it holds:*

$$E[|M_R|] = |\widetilde{R}| - \sqrt{\frac{2}{\pi}|\widetilde{R}|}, \tag{3}$$

where $|\widetilde{R}|$ is the number of vertices in $\widetilde{R}$ and $|M_R|$ the size of matching M_R.

xx first of all: a sentence of interpretation of the result xx

The proof of this result is technical and deferred to the appendix. For an informal argument, note that that the number of A-B pairs (conditioned on $N_R = |\widetilde{R}|$ pairs in total) is a random variable $N_{A-B} \sim Bin(N_R, 0.5)$ since $A - B$ and $B - A$ are equally likely. Given this, the number of unmatched pairs under the PM assumption is $Q = |N_{A-B} - N_{B-A}| = |2N_{A-B} - N_R|$. If we model N_{A-B} as a Normal random variable, then Q follows a Half-Normal

distribution and has expected value exactly equal to $\sqrt{\frac{2}{\pi}} \cdot \sigma$ where σ is its standard deviation. We know that the variance of N_{A-B} is proportional to N_R, therefore the expected value of the unmatched pairs is proportional to $\sqrt{\frac{2}{\pi} N_R}$.

Let $\mu(n)$ denote the expected size of a maximum matchings in $\widetilde{G_n}$. **xx rewrote this to adopt q_R explicitly. OK? xx**

Theorem 1. *Given the Regularity and PM assumptions, it holds that:*

$$\mu(n) = \gamma n - \sqrt{\frac{2q_R}{\pi} n} - 2, \qquad \gamma = 1 - q_{UD} + q_{OD},$$

where q_R, q_{UD} and q_{OD} denote the probability of reciprocal, under-demanded and over-demanded pair types, respectively.

Proof. By the Regularity assumption, there exists a regular maximum matching, where all OD pairs are matched with pairs in UD and subcomponents S and R assume maximum matchings internally. Under the PM assumption, we can apply Lemmas 5 and 6, and calculate the expected size of this matching as **xxxx why not** $|OD| + |OD| + |S| - 2 + |R| - \sqrt{\frac{2}{\pi}|R|}$**? xxxx**

$|UD| - |OD| + \Theta(\sqrt{n}) + 2$, where we overload notation and write $|UD|$ and $|OD|$ to denote the sizes of the subcomponents corresponding to UD and OD pair types. By (1) and Lemma 4 we have that $E[|UD| - |OD|] = (q_{UD} - q_{OD})n$, and hence the expected size of the matching is,

$$\mu(n) = n - ((q_{UD} - q_{OD})n + \Theta(\sqrt{n}) + 2). \tag{4}$$

Setting $\gamma = 1 - q_{UD} + q_{OD}$, we conclude that,

$$\mu(n) = \gamma n - \Theta(\sqrt{n}) - 2 \tag{5}$$

xx change proof to match new statement of thm xx □

Substituting values for q_{UD}, q_{OD} and q_R, we obtain the following quantified expression for the expected cardinality of the maximum matching in $\widetilde{G_n}$ under the idealized model:

$$\mu(n) \approx 0.556n - 0.338\sqrt{n} - 2 \tag{6}$$

This analysis explains many simulation results from previous work, which in turn validates the regularity and PM assumptions. For example, the cardinality of the matchings in Roth et al. [13] (see Propositions 1 and 2 and Table 2), are explained.[8] Similarly, the results in Table 1 of Abraham et al. [1], can be explained.

3.3. The Square-Root Law

We can now quantify the increase in the expected number of transplants from pooling the donor-patient pairs from multiple hospitals into a single exchange. For now we perform this analysis without considering incentives and the possibility that a hospital might benefit by hiding some of its local pairs. For this, let $W(m, n)$ denote the *welfare gain*, which is the expected increase in the number of transplants from pooling together m hospitals each of size n:

$$W(m, n) = \mu(mn) - m\mu(n) \tag{7}$$

The following result follows immediately from Theorem 1, substituting the expression for the expected cardinality of a maximum matching into $W(m, n)$:

Theorem 2. *For m hospitals each with donor-patient pools of size n, and under the PM and Regularity assumptions, the welfare gain when all hospitals participate truthfully in a larger exchange is*

$$W(m, n) = (m - \sqrt{m})\sqrt{\frac{2q_R}{\pi}n} + 2(m - 1), \tag{8}$$

where q_R is the probability of R pair types entering the exchange.

By symmetry, the incremental matches that are achieved are distributed evenly amongst participating hospitals: **xx refer to this as sq law. OK? xx**

Corollary 1 (square-root law)**.** *For m hospitals each with donor-patient pools of size n, and under the PM and Regularity assumptions, the expected gain in the number of transplants to an individual hospital when all hospitals participate truthfully in a larger exchange is,*

$$W_h(m, n) = (1 - 1/\sqrt{m})\sqrt{\frac{2p_R}{\pi}n} + 2(1 - 1/m), \tag{9}$$

[8] With $n = 100$, expression (6) yields an expected number of matched vertices of $55.6 - 3.38 - 2 = 50.22$, and close to the reported value of 49.7.

where p_R is the probability of R pair types entering the exchange. As the number of hospitals $m \to \infty$, the expected benefit to the patients of an individual hospital increases as:

$$\lim_{m\to\infty} W_h(m,n) = \sqrt{\frac{2p_R}{\pi}n} + 2. \tag{10}$$

A nationwide kidney exchange program has the capacity to yield individual benefits to participating hospitals that follow a square-root law to the size of their patient lists.

4. An Efficient Mechanism for KPD Exchanges with 2-Cycles

We now turn our attention to the problem of efficient mechanism design, first of all considering the case of exchanges that consider only 2-cycles (or swaps) when identifying matchings.

Let G_h denote the true donor-patient graph of hospital h, and consider hospitals $h \in \{1,\ldots,m\}$. Let $G_\oplus = \bigoplus_h G_h$ denote the joint graph that is obtained by:

- Adopting as the vertices of $G_\oplus$ all vertices of $G_1,\ldots,G_h$
- Adopting as the edges of $G_\oplus$ all edges of $G_1,\ldots,G_h$ and also additional edges between ABO and tissue-compatible pairs in the graphs G_h and $G_{h'}$ $(h \neq h')$ of two different hospitals.

Let $G = (G_1,\ldots,G_m)$. Given reported graphs $G' = (G'_1,\ldots,G'_m)$, a KPD exchange mechanism determines a *matching* $\Gamma(G')$ on $\bigoplus_h G'_h$, and we adopt $\Gamma_h(G')$ to denote the pairs in graph G'_h that are matched in $\Gamma(G')$.

The *strategy* s_h of hospital h determines which pairs, and thus subgraph $s_h(G_h)$ to report. We restrict attention to strategies that can only hide donor-patient pairs, and for any pair reported assume that the compatibility relationship between the pair and anyother pair (whether reported by h or some other hospital) cannot be misrepresented. **xx justify this!! xx** Let $s = (s_1,\ldots,s_m)$ denote the joint strategy profile of all m hospitals, and adopt $s(G) = (s_1(G_1),\ldots,S_m(G_m))$ to denote the vector of reported graphs.

We assume that a hospital that reports $G'_h = s_h(G_h)$ will compute a maximum matching on the graph associated with any pairs that are reported

but unmatched by the mechanism and those pairs that are not shared with the mechanism. The *utility* to hospital h, given strategy profile s, is:

$$u_h(s, G) = |\Gamma_h(s(G))| \ + \ |M(G_h \setminus \Gamma_h(s(G)))|, \tag{11}$$

where $M(G)$ is a maximum matching on graph G. This utility represents the total number of its pairs that are matched, and thus the total number of its patients that receive a transplant.

Given strategy s, let $Q(s, G) = \sum_h u_h(s, G)$ denote the total number of pairs that are matched internal to the mechanism Γ and also considering that each hospital determines a maximum matching on its own residual.

Definition 2 (efficiency). A mechanism Γ is *efficient* under strategy profile s if the total number of pairs matched $Q(s, G) = |M(G_\oplus)|$ for all possible graph profiles $G = (G_1, \ldots, G_m)$, given m hospitals, where M is a maximum matching and $G_\oplus$ is the combined graph associated with G.

The definition of efficiency is slightly non-standard from the perspective of mechanism design because it allows for the agents (the hospitals in this case) to take actions after the outcome of the mechanism is determined that affect welfare. In this case, those actions correspond to a hospital computing a maximum matching on its own residual graph. As an example, a mechanism that computes a maximum matching based on reported graphs is efficient under the truthul strategy profile.

Definition 3 (EPIC). A mechanism Γ is *ex post incentive compatible* if it is utility maximimzing for every hospital to report its true graph, whatever the graphs of other hospitals, as long as the other hospitals report truthfully.

4.1. The rCM Mechanism

Continuing, we first establish that a simple mechanism `rCM` in which an ex post equilibrium, in the idealized model, is for each hospital to compute a regular matching locally, hold the pairs involved in this matching back, and then send the remainder to the mechanism. We establish that the mechanism is efficient in this equilibrium.

Definition 4. Given reported graphs G', mechanism `rCM` outputs a random, maximum matching $M(G')$, sampled uniformly at random from the set of maximum matchings.

xx adopted new language here. **xx** Given graph G_h, say thar a matching is *almost regular* if:

(1) it is regular when a regular matching exists,

(2) otherwise, it is a matching that is composed from a maximum matching on the subgraph consisting of the OD and UD components, and then a maximum matching on the remainder graph once matched OD and UD pairs are removed.

Say that a hospital's strategy is a *canonical deviation* if it computes an almost regular matching given its local pool, and then shares the unmatched pairs with the mechanism.

Lemma 7. *It is an ex post Nash equilibrium in the* `rCM` *mechanism for every hospital to follow a canonical deviation in the idealized model.*

Proof. Focus on hospital h, and assume (consistent with the regularity assumption) that the rest of the hospitals report graphs that contain only UD, R and S pairs. Let G_{-h} denote the composite graph formed by the reports of hospitals except h. Fix a regular maximum matching of h and denote with $P\#Q$ the fact that pair P is matched to Q in this matching. Last, when pairs P and Q are of the same type we write $P \equiv Q$.

First, notice that a hospital h has the incentive to report both pairs in some $P\#Q$ if it expects that P will be matched to some other internal pair P' and Q will be matched to a pair in G_{-h} (or vice versa). If $P \equiv OD$, then $Q \equiv UD$ and Q cannot be matched in G_{-h} since there are no OD pairs in G_{-h}. So, hospital h prefers to clear all OD#UD matches internally. Hence $P \equiv R$ or $P \equiv S$. Recall that, in a regular maximum matching pairs in the R and S subgraphs are matched internally. Therefore, $P' \equiv Q$ and by the PM assumption, P' could be matched in G_{-h} as well as Q can be. Therefore, hospital h would get the same utility if it cleared P and Q internally and just shared P', and so weakly prefers to clear all $P\#Q$ internal matches. For the remainder unmatched pairs, h weakly prefers to share them into the central pool. This concludes the proof that canonical deviation is best response for hospital h and thus canonical deviation is an ex post Nash equilibrium of this game. □

Theorem 3. *The mechanism* `rCM` *is efficient in an ex post Nash equilibrium in which every hospital follows a canonical deviation in the idealized model.*

Proof. First we prove for two hospitals. Denote with T_i the number of pairs of type $T \in \mathcal{T}$ for hospital $i \in \{1, 2\}$, and with $T_{xy,i}$ the number of those

pairs who are of type $X-Y$. Since $X=Y$ for the S subgraph, we only write $S_{x,i}$ for type $X-X$. For example, OD_1 is the number of over-demanded pairs of hospital 1 and $R_{ab,2}$ is the number of $A-B$ pairs of hospital 2. Also denote with $I(n)$ the function ($n \mod 2$), i.e. $I(n)$ is 1 if n is odd and 0 if it is even.

We compare the case of full sharing with the case where hospitals canonically deviate and show that both produce the same # of exchanges under the idealized mechanism:

(a) By the regular assumption, full sharing performs $2 \cdot (OD_1 + OD_2)$ by completely matching OD pairs with UD pairs. The same matches for these pairs is achieved through canonical deviation.

(b) A total of $I(S_{a,1}+S_{a,2}) + I(S_{b,1}+S_{b,2}) + I(S_{o,1}+S_{o,2}) + I(S_{ab,1}+S_{ab,2})$ exchanges of type S are also achieved by full sharing. However, for the function $I(.)$ it holds that $I(n+m) = I(I(n)+I(m)), \forall n, m$ integers, and hence the same number of matches is achieved through canonical deviation.

(c) $|R_{ab,1} + R_{ab,2} - R_{ba,1} - R_{ba,2}|$ exchanges of type R remain *unmatched* by full sharing. For brevity denote $r_1 = R_{ab,1} - R_{ba,1}$ and $r_2 = R_{ab,2} - R_{ba,2}$. WLOG we can assume $r_1 \geq 0$ and $r_1 \geq |r_2|$ (if this doesn't hold we can relabel hospitals or blood types). Then, the total unmatched pairs under full sharing is equal to $u = |r_1 + r_2|$. There are two cases: (i) $r_2 > 0$ and so $u = r_1 + r_2$ and this is the case where both hospitals have unmatched pairs of same R type (e.g. both hospitals have an excess of $A-B$ pairs and so cannot benefit from sharing), and (ii) $r_2 < 0$ and so $u = r_1 - r_2$ remains unmatched. When sharing under canonical deviations, the number of unmatched pairs is $||r_1| - |r_2||$ but this is equal $r_1 - r_2$ by the assumptions on r_1, r_2.

This completes the proof that canonical deviations achieve the same amount of exchanges as full sharing in the case of two hospitals. To prove for m hospitals we can simply use induction. For example, when $m=3$ we can treat hospitals 1 and 2 as one big hospital and apply the result for $m=2$.

□

4.2. The xCM mechanism

The `xCM` mechanism follows from `rCM` by the revelation principle:

Definition 5. The `xCM` mechanism operates in two stages:

- Given reports $G' = (G'_1, \ldots, G'_m)$, it first determines an almost-regular matching for each hospital h
- Second, it runs `rCM` on the composition of the remainder graphs from each hospital, removing all pairs involved in these almost-regular matchings.

The matching returned to hospitals on the composite graph $G'_{\oplus}$ formed from the reported graphs is the union of the edges in the hospital-specific almost-regular matchings and the matching identified in stage two.

A uniform sample from the maximum matchings on the remainder graphs in stage two can be implemented through a relabeling of the vertices and a generic maximum matching algorithm. **xx need to say just a bit more about this— which generic algorithm? xx**

Theorem 4. *The `xCM` mechanism is efficient and EPIC in the idealized model.*

Proof. **xx sketch, but really all the work should be for rCM! xx** efficiency and EPIC should be basically immediate from the analysis of rCM! maybe the only thing to argue is why this is EPIC whereas PERHAPS the last one was DSE?? □

It is important for truthfulness that the first step of `xCM` selects a regular matching rather than an arbitrary matching that provides a hospital with at least its stand-alone quantity of pairs matched (given its reported graph.) To see this, assume an *OD* pair is matched internally to a *R* pair instead of a compatible *UD* pair. In the central pool the *UD* pair has much lower probability of getting matched since it is matched *only* to a particular type of *OD* pairs. In contrast, an *R* pair can be matched to the symmetric *R* pair and also an *OD* pair and therefore has higher chances of getting matched in a central pool.

5. Generalizing to Allow for 3-cycles

xx didn't spend much time on this section yet. will wait until the paper before here is fixed up. ideally it will adopt the same kind of revelation principle flow as the previous section. xx

5.1. Idealized Model

With 3-way exchanges and under the PM assumption, then as the size of a hospital pool increases the S subcomponent is matched perfectly by internal matching with high probability. To see this, if $|S|$ is even then we can apply the PM assumption to get perfect matching. If $|S|$ is odd, then we only need to find a 3-way cycle in the S graph involving the unmatched S pair. The probability of 3 random pairs forming such a cycle is some $q, q > 0$ and thus the probability that no 3-way cycle can be found is $(1-q)^{\binom{|S|}{3}}$, which decreases exponentially as the size increases. Going forward, we modify the PM assumption to include this special case that an S subcomponent is matched perfectly.xx

Definition 6 (Regular M_n^3 matching). Let M_n^3 denote a xx directed matching of $\widetilde{G_n}^3$ which is obtained successively by:

1. First applying a regular two-way matching M_n on the graph.
2. Matching all unmatched pairs A-B from Step 1, in 3-way exchanges involving pairs, A-B, O-A, B-O and similarly matching all unmatched pairs B-A in 3-way exchanges involving B-A, O-B and A-O pairs, breaking (if necessary) any 2-way matches that were created in Step 1.
3. Matching all O-AB pairs in 3-way exchanges involving pairs, O-AB, A-O and AB-A, breaking (if necessary) any 2-way matches that were created in Step 1.[9]
4. Using 3-way exchanges involving only S pairs (e.g. A-A, A-A, A-A) as necessary to match all S pairs internally, breaking any 2-way exchanges in S as necessary.

Theorem 5. *On a $\widetilde{G_n^3}$ graph allowing 3-way exchanges, a regular matching M_n^3 is maximum if it exists.*

Proof. First, suppose for contradiction that it was possible to introduce the same number of 3-exchanges but also combine this with more 2-exchanges. But then we would have a contradiction because each 3-exchange knocks out exactly one 2-exchange, and thus we could get a 2-way matching from the 3-way matching that is better than the M_n computed in Step 1 (and thus a

[9]Symmetrically, one could match all O-AB pairs in 3-way exchanges involving O-AB, B-O, and AB-B, or some mixture.

contradiction). Note also that one can't introduce more 3-exchanges, because M_n^3 is already maximizing this quantity.

Last, we need to establish that doing less 3-way exchanges is not helpful. For this, note that M_n^3 is matching all OD, R and S pairs which means that if there exists a better matching, it would match more *UD* pairs. Say M_n^3 is using n_1 *OD* pairs for 2-way exchanges (matching to a single *UD* pair), n_2 3-way exchanges that involve 1 *UD* pair (Step 2) and n_2 pairs for 3-way exchanges that involve 2 *UD* pairs (Step 3). Therefore, M_n^3 matches $n_1 + n_2 + 2n_3$ *UD* pairs in total. However, since all *OD* pairs are matched, $|OD| = n_1 + n_2 + n_3$ and therefore M_n^3 is matching $|OD| + n_3$ pairs in UD. Clearly, since M_n^3 is greedily maximizing n_3, it is using the maximum number of *UD* pairs that is possible and hence no other allocation can match more *UD* pairs. □

Theorem 6. *A regular matching M_n^3 in $\widetilde{G_n^3}$ exists w.h.p as n goes to infinity.*

Proof. Recall that all pairs in the S subcomponent are matched internally w.h.p. by the PM assumption. Furthermore, in Step 2 we have pairs from the R subcomponent of the same class, because of the Regularity assumption and the fact that we perform a 2-way regular (and thus maximum) matching M_n at Step 1. Assume without loss of generality that this pair type is the A-B. Next assume N_{O-A} and N_{A-B} to be the random variables representing the number of these respective types in the $\widetilde{G_n}$ and recall that $N_{O-A} \sim Bin(n, q_{O-A})$ where q_{O-A} is the probability of a random pair being of type O-A. This means that

$$E[N_{O-A}] \propto n, \quad Var(N_{O-A}) \propto n \tag{12}$$

In addition, recall that the size of the R subcomponent can be modeled as $|R| = N_R \sim Bin(n, p_R)$ and that N_{A-B} (the remainder) follows $|2Y - N_R|$ where $Y|N_R \sim Bin(N_R, 0.5)$, which follows directly from Lemma 6. Notice that

$$\begin{aligned} E[2Y - N_R] &= 2 \cdot E[E[Y|N_R]] - E[N_R] = 2E[0.5 \cdot N_R] - E[N_R] = 0 \\ Var(2Y - N_R) &= E[Var(2Y - N_R|N_R)] + Var(E[2Y - N_R|N_R]) \\ &= 4 \cdot E[Var(Y|N_R)] + 0 = E[N_R] \propto n \end{aligned}$$

Then we only need to prove that $P(N_{O-A} > N_{A-B})$ goes to 1 for sufficiently large values n of the graph size. To see this for sufficiently large

n, note that N_{A-B} approaches a Half-Normal and therefore it has expectation $\propto \sqrt{n}$. Then, the standard deviations of both N_{O-A} and N_{A-B} r.v.s are $\propto \sqrt{n}$ while the locations of N_{O-A} and N_{A-B} are $\propto n$ and $\propto \sqrt{n}$ respectively. This implies that for sufficiently large n we have $P(N_{O-A} > N_{A-B}) \to 0$ for large n. For the same reason, note that $P(N_{O-A} + N_{A-B} < n) \to 1$ and hence our modeling through *independent* normals is justified.

We thus proved that 3-way exchanges of Step 2 occur with high probability as n goes large. The analysis for Step 3 can follow directly from Lemma 4. □

According to Theorem 6, there are almost always more O-A pairs than A-B for sufficiently large n (the same holds for O-B and B-A pairs). Validating this, when $n = 50$ $P(N_{O-A} > N_{A-B}) = 0.72$, for $n = 100$ then $P(N_{O-A} > N_{A-B}) = 0.86$ and for $n = 200$ the same probability is almost 0.95. For the purpose of our analysis, we abstract away from this variation by generalizing the regularity assumption to cover 3-way exchanges:

Assumption 3. (Generalized Regularity) *For every $\widetilde{G_n^3}$ there exists a regular matching M_n^3.*

Theorem 7. *The expected number of matches under the generalized regularity assumption, denoted by $\mu_3(n)$, are given by*

$$\mu_3(n) = (\gamma + \epsilon) \cdot n, \qquad \epsilon > 0 \tag{13}$$

Proof. Observe that under the generalized regularity assumption, all reciprocal pairs are matched and hence there is no square-root remainder as in the 2-way exchange case. Furthermore, if we denote with $N_{O\text{-}AB}$ the total number of O-AB pairs in a $\widetilde{G_n}$, then the mechanism is allocating $N_{O\text{-}AB}$ more matches than in the 2-way exchange case. This means that in expectation it is allocating $E[N_{O\text{-}AB}] = q_{OD}\left(\frac{f_{AB}}{f_{AB}+f_A+f_B}\right) n = \epsilon n$ additional transplants. For our assumptions so far (see Lemma **??**), $\epsilon = 0.011$. □

5.2. Incentive Analysis

We now examine the incentives problem when 3-way exchanges are allowed. We first extend the `rCM` mechanism:

Definition 7. `rCM`^{3} outputs a uniformly sampled, maximum directed matching $M(G')$ given reports G' where exchanges up to 3-way are allowed.

Definition 8. xCM^3 is a mechanism that operates in two stages:

1. Given reports $G' = (G'_1, \ldots, G'_m)$, then for each hospital h internally,
   (i) Compute a maximum 2-way matching on the $UD \cup OD$ subcomponent, the a maximum 2-way matching on the remainder graph of G'_h including the S and R subcomponents.
   (ii) Considering only 3-way exchanges, maximize the number of pairs matched by considering unmatched A-B and B-A pairs (relevant 3-cycles being A-B, O-A, B-O and B-A, O-B, A-O), and all O-AB pairs (relevant 3-cycles being O-AB, A-O, AB-A and O-AB, B-O, AB-B). Finally, use 3-way exchanges as necessary on S pairs in order to maximize the number of S pairs that can be matched by a combination of 2-exchanges and 3-exchanges.
2. Second, it runs rCM^3 on the union of the remainder graphs formed by removing the matched pairs from Step 1 from G'. Eventually, the matching output by xCM^3 is the combination of the matching identified in steps 1 and 2.

The clearing rule for xCM^3 is designed to ensure that it will select a regular 3-matching whenever one exists.

We analyze *idealized* rCM^3 and *idealized* xCM^3, meaning that we assume Perfect Matching (slightly extended to handle 3-exchanges in the S subcomponent as discussed above), and generalized regularity.

Theorem 8. *A fully-hiding strategy forms an ex post Nash equilibrium of idealized rCM^3, and the mechanism is efficient in this equilibrium.*

Proof. Consider hospital h. When hospitals other than h fully deviate, then under generalized regularity and the PM assumption then each send only UD pairs. Suppose that hospital h had exclusive access to these pairs. The maximum matching for h would remain a regular matching, and thus the hospital should fully deviate and implement this regular matching locally (and is indifferent between sending residual UD pairs or not.) Efficiency follows by observing that the regular matching selected locally by each hospital is also regular on the combined graph and thus maximum on that graph. □

Theorem 9. *The idealized mechanism xCM^3 is efficient and EPIC.*

Proof. For the same reason as in rCM^3, recognizing that the other hospitals only share UD pairs after stage 1 when following canonical strategies in

xCM^3, then the maximum matching h could hope for even given full access to these *UD* pairs is the just its regular 3-way matching. Unlike in rCM^3, in the xCM^3 mechanism the hospital can achieve this through any canonical strategy. IR and efficiency are both immediate, efficiency since the union of regular matchings determined by each hospital is also a regular matching on the combined graphs. □

On one hand, under our generalized regularity and PM assumptions, there are no interactions between hospitals in obtaining efficiency because each hospital is able to obtain a maximum matching internally that leaves only *UD* pairs.[10] The 3-way matchings are able to "mop-up" any remaining inefficiency that was occurring due to the imbalance of R pairs. Given this, we obtain a simple explanation of how to obtain incentive compatibility with 3-way matches, through imposing IR constraints in the center.

6. Experimental results

In this section we report the results of extensive computational analysis of the properties of the xCM and xCM^3 mechanisms. In particular, we are interested in understanding the robustness of the theory under more realistic probabilistic models.[11] We first consider 2-way cycles and then exchanges with both 2-way and 3-way cycles.

6.1. 2-way Cycles

In our first experiments, we generate random graphs $\widetilde{G_n}$ through the process defined in Section 2.1. The first two tables refer to the matching

[10] Ashlagi & Roth [3] get a similar positive result in the context of 3-way matchings for their "Bonus" mechanism. In fact, the first steps of their algorithm are similar to the first step of xCM and xCM^3, in the way that matches in the S and R subcomponents are implemented. From that point, their algorithm diverges as it is using a separate lottery procedure to allocate pairs in the *UD* subcomponent by splitting the *OD* subcomponent of each hospital into specific subsets.

[11] We used Perl to implement the $\widetilde{G_n}$ random graph model. Maximum matchings were determined by an implementation of the Galil algorithm [7] that runs in $O(|V|^3)$ time, which proved faster than open-source linear programming libraries. Currently, on a dual-core 2.4GHz PC, the donor-patient graph of $1,000$ pairs is cleared in approximately 10 seconds. The Perl source code is available for download from `http://www.eecs.harvard.edu/econcs/code/rgke.zip` Detailed instructions on how to reproduce the results of this section can be found in the package.

properties of the S and R subcomponents, in which we report the number of matched patients compared to the individual sizes of the subcomponents averaged over $1,000$ samples.

#pairs	#matches	$(\lvert S\rvert - \beta)$	sd
20	18.12	18.0	1.15
40	38.1	38.0	1.06
60	57.91	58.0	1.14
80	77.93	78.0	1.11
100	97.92	98.0	1.07

Table 2: Maximum matching in S.

The expected values for optimal matchings of the subcomponent S (Table 2) are given by Lemma 5 and for the expected matchings in R by Lemma 6. The realized values are very close to what our analysis predicts. Notice also the increasing standard deviation for matchings in R, which stems from the properties of the folded binomial. For Theorem 1, which computes the expected cardinality of the maximum matchings of $\widetilde{G_n}$, we simply take 1,000 samples of $M(\widetilde{G_n})$ for various values of n. The results are shown in Table 4, and validate the analysis.

#pairs	#matches	$(\lvert R\rvert - \sqrt{\frac{2}{\pi}\lvert R\rvert})$	sd
20	16.64	16.43	2.64
40	34.94	34.95	3.73
60	54.16	53.82	4.79
80	72.97	72.86	5.67
100	91.76	92.02	6.12

Table 3: Maximum matching in R

Let's now examine the welfare gain of `rCM` with 3 truthful hospitals, each reporting some $\widetilde{G_n}$. Again, we take 1,000 samples for various sizes of n. The left part of Table 5 refers to the simulation results and the right one gives the theoretical predictions. The column titled *"Selfish"* gives the welfare when hospitals are not reporting anything to the centralized mechanism. Our theory predicts with high accuracy the expected welfare gain. For example,

#pairs	#matches	$\mu(n)$	sd
20	7.83	7.57	3.16
40	18.12	18.04	4.46
60	28.56	28.66	5.66
80	39.2	39.35	6.69
100	49.51	50.09	7.33
200	104.75	104.18	11.22

Table 4: Maximum matching in $\widetilde{G_n}$.

n	H-1	H-2	H-3	Total	Selfish	$W(s, G)$
20	9.63	9.36	9.45	28.44	22.77	5.94
40	20.26	20.05	20.31	60.62	54.24	6.74
60	31.05	31.22	31.23	93.49	86.14	7.36
80	41.99	41.97	42.31	126.26	118.27	7.88
100	52.97	52.37	52.59	157.93	150.55	8.33

Table 5: Welfare gain from `rCM` with truthful hospitals.

for $n = 60$ we have that $W = 7.36$ using Theorem 2, when the real value is $93.49 - 86.14 = 7.35$. Last, notice that the welfare gain is distributed evenly among all hospitals with values that are very close to $\mu(n) + W/3$, as predicted in Theorem 2. For example, if $n = 60$, the predicted utility is $28.71 + 2.45 = 31.16$, which is very close to what all hospitals receive (see H-1, H-2, and H-3 for $n = 60$)

Next, we proceed to study cases in which not all hospitals are truthful. Here we will consider only full deviations, i.e. hiding the entire M_n graph for internal matching; for brevity we will refer to this as simply "a deviation." Our analysis in Section 4 showed that deviating strategies are provably hurtful to the utility of truthful hospitals under the non-IR `rCM`. We modify our simulation so that H-1 deviates fully. We first run on `rCM`. See Table 6.

First, notice that H-1 matches consistently more pairs any other truthful hospital. Second, it is interesting to see that total welfare is not hurt by the deviation.[12] We run `xCM` under the same setting to get the results in Table 7.

[12]This should not come as a surprise, since Theorem **??** actually builds upon this prop-

n	H-1$_d$	H-2	H-3	Total
20	10.67	8.60	8.65	27.92
40	23.08	18.39	18.52	59.99
60	35.45	28.68	28.81	92.93
80	47.71	38.62	38.55	124.88
100	60.68	49.18	48.94	158.8

Table 6: H-1 deviates in `rCM`

There are two important remarks. First, the mechanism produces efficient outcomes if we compare with Table 6 (under `rCM`, which generates efficient outcomes under truthful reports), and as predicted by Theorem **??**. Second, imposing IR is effective in achieving BNIC, if one looks at the individual utilities received by each hospital (also compare with Table 5).

n	H-1$_d$	H-2	H-3	Total
20	9.28	9.46	9.27	28.01
40	19.74	19.79	20.03	59.57
60	30.65	30.51	30.54	91.69
80	41.66	41.91	41.13	124.7
100	52.02	52.39	52.36	156.76

Table 7: H-1 deviates under `xCM`

Notice that there is a small difference in the efficiency achieved in Table 7 compared to Table 5 (156.76 vs. 157.93 respectively for $n = 100$), whereas our theory predicts that these two numbers should be the same. In addition to sampling issues, the graph size if not large enough for the regularity assumption to hold almost surely.

6.2. Robustness of results

The $\widetilde{G_n}$ model makes several simplifying assumptions, by using a specific blood type distribution and treating the tissue-type compatibilities as independent random events with probability $p_c = 0.2$. Here we test how relevant our model remains with the introduction of "nuisance parameters."

erty to deliver a mechanism that is IR and efficient on average.

In particular, we test two different blood-type distributions, one O(27%), A(32%), B(31%) and AB(10%) as in Korea, and one set to O(44%), A(42%), B(10%) and AB(4%) as in the USA. Each distribution induces different parameters for the calculation of the $\mu(n)$ formula in Lemma **??**. Let us denote with $\mu_K(n)$ and $\mu_{US}(n)$ the formulas for Korea and US respectively. Furthermore we endow p_c with a variability as in Table **??**. In particular, every pair draws a low, medium or high Panel Reactive Antibody (PRA) type with prob 0.72, 0.2 and 0.098 respectively, defining a probability p_c of 0.05, 0.45 or 0.9 for a positive cross match and incompatibility. Two pairs i, j are then tissue-type compatible with probability $(1 - p_{c,i})(1 - p_{c,j})$

#pairs	#(Korea)	$\mu_K(n)$	#(USA)	$\mu_{US}(n)$
20	7.18	8.67	5.83	7.62
40	17.21	20.6	14.4	18.11
60	28.27	32.74	23.86	28.75
80	39.54	44.99	33.89	39.46
100	51.24	57.3	44.01	50.22
200	111.16	119.41	97.6	104.39

Table 8: Matches with different blood type (ABO) distributions and with non-uniform tissue-type incompatibility.

We reproduce Table 4 in this more realistic scenario. The results on 1,000 samples are shown in Table 8. The theoretical predictions are less accurate, due mostly to the introduction of non-uniform tissue type compatibility. However, our theoretical model remains robust, and can be seen to approach the results of the simulation for large n.

We examine if `xCM` remains effective in mitigating the impact of strategic behavior under non-uniform tissue type compatibility. To stress our analysis further, we will examine the US blood type distribution and will denote this combination of parameters as "PRA+US". We first present the welfare when all hospitals are truthful under `rCM`; i.e., the welfare in an efficient outcome. The results are shown in Table 9. Next we let hospital H-1 deviate. The results shown in Table 10 reveal that H-1 is doing consistently better than truthful hospitals in `rCM`, and at the same time hurts the overall efficiency ($\sim$ 1-2%, comparing with Table 9).

Let's now consider the effect of H-1 deviating under mechanism `xCM`. The results are shown in Table 11. First observe that `xCM` restores fairness,

n	H-1	H-2	H-3	Total
20	8.05	7.88	7.75	23.67
40	18.37	18.08	18.04	54.49
60	29.15	29.13	28.99	87.27
80	39.82	39.85	39.4	119.06
100	50.99	50.71	51.21	152.92

Table 9: All truthful in `rCM` (PRA+US blood freqs.)

n	H-1$_d$	H-2	H-3	Total
20	9.18	7.45	7.33	23.96
40	20.41	16.52	16.68	53.61
60	32.19	26.16	25.74	84.09
80	45.29	36.25	35.98	117.52
100	57.15	46.0	45.97	149.12

Table 10: H-1 deviating in `rCM` (PRA+US)

since H-1$_d$ now receives approximately the same number of transplants with truthful participants. Furthermore, the cost in efficiency is small and stems mostly from the O and S subcomponents, which under non-uniform tissue-type compatibility are more likely to contain highly sensitized patients. As a result, our connectivity assumptions are violated, thus leading to a loss in overall efficiency. For $n = 40$, this loss is $\sim 3.9\%$ of the total efficiency, which can be seen by comparing the value 52.37 in Table 11 to the 54.49 in Table 9. Again we see that the total efficiency, under deviations by hospital one, is slightly better in `rCM` than `xCM`. This seems to be the necessary trade-off, in order to restore proper incentives.

6.3. 3-way Cycles

Let us now consider exchanges with 3-way cycles and the `xCM`^{3} mechanism. In all of the following experiments we allow for non-uniform tissue-type compatibility. We first present the number of matched pairs when all hospitals are truthful under `rCM`^{3}. The results are shown in Table 12. [13] We

[13] We used the software in `http://web.mit.edu/iashlagi/www/papers/KECyclesAndChains.zip`, which is using CPLEX to solve maximally for 3-way matches. It is also being used for

n	H-1$_d$	H-2	H-3	Total
20	7.78	7.3	7.56	22.64
40	17.54	17.36	17.47	52.37
60	27.66	27.72	27.27	82.65
80	37.2	37.53	37.48	112.20
100	48.33	47.81	48.02	144.16

Table 11: H-1 deviating in `xCM` (PRA+US)

now let hospital H-1 deviate in `rCM`^{3}. As in the 2-way exchanges setting, the results shown in Table 13 reveal that H-1$_d$ is doing better than truthful hospitals under `rCM`^{3}.

n	H-1	H-2	H-3	Total
20	8.68	8.61	8.53	25.81
40	21.63	21.39	21.66	64.68
60	35.35	35.25	35.24	105.85
80	49.14	48.87	48.80	146.81
100	62.45	62.09	62.68	187.23
200	127.11	126.17	126.53	379.81

Table 12: All truthful in `rCM`^{3} with non-uniform tissue type compatibility

n	H-1$_d$	H-2	H-3	Total
20	9.35	7.81	7.57	24.72
40	22.81	19.35	18.45	60.62
60	38.03	31.71	30.08	99.81
80	53.99	44.29	42.29	140.57
100	69.63	56.52	54.40	180.55
200	147.97	118.27	113.65	379.89

Table 13: H-1 deviating in `rCM`^{3} with non-uniform tissue type compatibility

similar 3-way experiments in Ashlagi and Roth [3].

n	H-1$_d$	H-2	H-3	Total
20	7.34	7.13	7.08	21.55
40	17.69	17.43	17.41	52.53
60	29.05	28.67	29.12	86.84
80	40.78	40.45	40.93	122.16
100	53.18	53.10	53.09	159.36
200	119.62	119.88	121.08	360.57

Table 14: H-1 deviating in $\texttt{xCM}^3$ with non-uniform tissue type compatibility

For the effect of H-1 deviating under mechanism $\texttt{xCM}^3$ see Table 14. Mechanism $\texttt{xCM}^3$ is able to restore fairness, and H-1$_d$ is no longer receiving consistently more transplants than the truthful participants. Note also that $\texttt{xCM}^3$ has a noticeable impact on welfare. This is because the generalized regularity assumption is violated for small to moderate sized pools (30% of the times when $n = 50$). The efficiency loss for $n = 40$ pairs is $1 - 52.53/64.68 = 0.187$, i.e. almost 19% compared to the truthful $\texttt{rCM}^3$ (see Table 12).[14] However in Theorem 6 we proved that the assumption conditions will hold for large enough n, and this becomes evident even when $n = 200$, in which case the conditions of the assumption are met with probability 95%. Indeed, the efficiency loss when $n = 200$ is calculated to $1 - 360.57/379.81 = 0.0506$, i.e. down to a 5% loss.[15]

7. Conclusions

We address efficiency and incentives in kidney exchanges through a quantified random graph analysis. Our model allows us to derive an analytic expression for the expected number of patients matched in a maximum match-

[14]It also bears emphasis that the efficiency of $\texttt{rCM}^3$ under deviation by a single hospital is not representative of the efficiency we would obtain in $\texttt{rCM}^3$ if all hospitals deviated, as one would expect.

[15]Note that Table 14 reports fewer transplants than Table 7 although 3-way exchanges are used. This is because the latter was generated through simulation code and the former using the code from Ashlagi and Roth [3]. We performed additional tests (not included here) which reproduced the experiment of Table 7 and which compared well with our findings. For example, the number of transplants for $n = 80$ and hospitals 1,2 and 3 were 33.99 33.38 34.12 respectively.

ing. This allowed us to explain much of the early experimental results, and to study the welfare gains that can be achieved through pooled exchanges. In particular, we quantify the expected individual benefit to a hospital participating in 2-way exchanges as being proportional to the square-root of the size of each hospital's donor-patient pool. The `xCM` mechanism (2-cycles) and the $\texttt{xCM}^3$ mechanisms are efficient and EPIC in an idalized model. The robustness of the economic properties of the mechanisms are validated in simulation in settings with non-uniform tissue type sensitivity and alternate blood-type distributions.

References

[1] Abraham, D., Blum, A., Sandholm, T., 2007. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges, in: Proceedings of the 8th ACM conference on Electronic commerce, ACM. p. 304.

[2] Ashlagi, I., Fischer, F., Kash, I., Procaccia, A., 2010. Mix and match, in: Proceedings of the 11th ACM conference on Electronic commerce, ACM. pp. 305–314.

[3] Ashlagi, I., Roth, A., 2011. Individual rationality and participation in large scale, multi-hospital kidney exchange. Under review, brief abstract appeared in EC'11. .

[4] Delmonico, F., Arnold, R., Scheper-Hughes, N., Siminoff, L., Kahn, J., Youngner, S., 2002. Ethical incentivesnot paymentfor organ donation. New England Journal of Medicine 346, 2002–2005.

[5] Edmonds, J., 1965. Paths, trees, and flowers. Canadian Journal of Mathematics 17, 449–467.

[6] Erdos, P., Rényi, A., 1964. On random matrices. Magyar Tud. Akad. Mat. Kutató Int. Közl 8, 1964.

[7] Galil, Z., 1986. Efficient algorithms for finding maximum matching in graphs. ACM Computing Surveys (CSUR) 18, 23–38.

[8] Montgomery, R., Zachary, A., Ratner, L., Segev, D., Hiller, J., Houp, J., Cooper, M., Kavoussi, L., Jarrett, T., Burdick, J., et al., 2005. Clinical

results from transplanting incompatible live kidney donor/recipient pairs using kidney paired donation. JAMA 294, 1655.

[9] Papadimitriou, C., Steiglitz, K., 1998. Combinatorial optimization: algorithms and complexity. Dover Pubns.

[10] Rees, M., Kopke, J., Pelletier, R., Segev, D., Rutter, M., Fabrega, A., Rogers, J., Pankewycz, O., Hiller, J., Roth, A., 2009. A nonsimultaneous, extended, altruistic-donor chain. New England Journal of Medicine 360, 1096.

[11] Rees, M., Pelletier, R., Mulgaonkar, S., Laskow, D., Nibhanupudy, B., Kopke, J., Roth, A., Unver, U., Sandholm, T., Rogers, J., 2008. Report From A 60 Transplant Center Multiregional Kidney Paired Donation Program: 2. Transplantation 86, 1.

[12] Ross, L., Woodle, E., 2000. Ethical issues in increasing living kidney donations by expanding kidney paired exchange programs. Transplantation 69, 1539.

[13] Roth, A., Sönmez, T., Ünver, M., 2007. Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. American Economic Review 97, 828–851.

[14] Saidman, S., Roth, A., Sönmez, T., Ünver, M., Delmonico, F., 2006. Increasing the opportunity of live kidney donation by matching for two- and three-way exchanges. Transplantation 81, 773.

[15] Ünver, M., 2010. Dynamic kidney exchange. Review of Economic Studies 77, 372–414.

[16] Wikipedia, 2010. Blood type. [Online; accessed 22-December-2010].

[17] Zenios, S., Woodle, E., Friedman Ross, L., 2001. Primum non nocere: Avoiding harm to vulnerable wait list candidates in an indirect kidney exchange. Transplantation 72, 648.

Appendix A. Additional Proofs

Lemma 6. *Consider subcomponent S contained in some $\widetilde{G_n}$. Also denote with M_S the size of a maximum matching of S. Under the PM assumption, it holds:*

$$E[M_S] = |S| - 2 \tag{A.1}$$

Proof. By Lemma **??** the subcomponent is composed by four disconnected subcomponents each one adhering to the $G_{n,p}$ random graph model with $p = (1-p_c)^2$. Also, by the PM assumption every subcomponent with even number of pairs will have a perfect matching and it will have one pair unmatched in the case of an odd size. If S has k pairs and a single pair has probability q_x of being in subcomponent (X, X) then the average number of unmatched pairs in the subcomponent is equal to:

$$D_{k,x} = \sum_{i\ odd \leq k} \binom{k}{i} \cdot q_x^i (1-q_x)^{k-i} \tag{A.2}$$

If β is the average unmatched pairs in S then it holds:

$$\beta = \sum_{x \in \{O,A,B,AB\}} D_{k,x} \tag{A.3}$$

Let us make the following definitions:

$$C_d(n,q) = \sum_{i\ odd \leq n} \binom{n}{i} \cdot q^i (1-q)^{n-i} \tag{A.4}$$

$$C_e(n,q) = \sum_{i\ even \leq n} \binom{n}{i} \cdot q^i (1-q)^{n-i} \tag{A.5}$$

Obviously, for the aforementioned process, the expression $C_d(n,q)$ counts how many times on average, the process ends up in odd number of successes. Similarly, $C_e(n,q)$ counts how many times on average the process ends up in even number of successes. It therefore holds:

$$C_d(n,q) = q \cdot C_e(n-1,q) + (1-q) \cdot C_d(n-1,q)$$

$$C_e(n,q) = q \cdot C_d(n-1,q) + (1-q) \cdot C_e(n-1,q)$$

We can eliminate the function C_e and get a closed recursive relation for C_d. For convenience let us denote $C_d(n, q)$ with c_n. Then we have for $n \geq 2$:

$$c_n - 2(1-q) \cdot c_{n-1} + (1-2q) \cdot c_{n-2} = 0 \tag{A.6}$$

We define

$$\xi = 1 - 2q \tag{A.7}$$

Then the characteristic equation of A.6 is:

$$x^2 - (1+\xi) \cdot x + \xi = 0 \tag{A.8}$$

The roots of the characteristic equation A.8 are $\sigma_0 = 1$ and $\sigma_1 = \xi$. The general solution of recursive relation A.6 is thus:

$$c_n = a \cdot \sigma_0 + b \cdot \sigma_1^n = a + b \cdot \xi^n \tag{A.9}$$

We now solve for the special cases:

$$c_2 = C_d(2, q) = 2q(1-q) \tag{A.10}$$

$$c_3 = C_d(3, q) = q^3 + 3q(1-q)^2 \tag{A.11}$$

Combining equations A.7, A.9, A.10 and A.11, we finally get:

$$a = 0.5$$
$$b = -0.5$$

Hence, we conclude that:

$$C_d(n, q) = 0.5 - 0.5 \cdot \xi^n$$

Since $|\xi| < 1$ for $q \neq 0$, we have that:

$$lim_{n \to \infty} C_d(n, q) = 0.5 \tag{A.12}$$

Also, for the subcomponent S that has k pairs the average number of matched pairs is thus equal to:

$$E[M_S] = \sum_x \left(|X| - C_d(k, q_x) \right) = \sum_x |X| - \beta$$

Finally, we conclude:

$$E[M_S] = |S| - \beta$$

To compute the value of β, using result in A.12 we get:

$$\beta = \sum_x D_{k,x} = \sum_x C_d(k, q_x) = \sum_x 0.5 = 2$$

□

Lemma 6. *Consider the R subcomponent contained in some $\widetilde{G_n}$. Also denote with M_R the size of a maximum matching of R. Under assumption PM, it holds:*

$$E[M_R] = |R| - \Theta(\sqrt{|R|})$$

In particular,

$$\frac{|R| - E[M_R]}{\sqrt{\frac{2}{\pi} \cdot |R|}} \to 1 \quad \textit{for } |R| \to \infty$$

Proof. We make use of the PM assumption and assume that all pairs in the class with the fewer nodes will be completely matched on the other class. If i is the number of elements in the smaller class and R has a total k pairs, then we wish to calculate the average value of $k - 2i$. Assume now that k is even and let us make the following definitions:

$$A_k = \sum_{i=0}^{k/2} \binom{k}{i} \tag{A.13}$$

$$B_k = \sum_{i=0}^{k/2} i \cdot \binom{k}{i} \tag{A.14}$$

$$\gamma_k = \binom{k}{k/2} \tag{A.15}$$

Because of symmetry, observe that every random pair in R has 0.5 probability of belonging in any of the two classes. Therefore, the probability that the lesser class has i elements, $i \leq k/2$, is equal to $\binom{k}{i} \cdot 2^{-k+1}$. Denote with $\tilde{u}$ the average unmatched pairs. Then $\tilde{u}$ is equal to:

$$\sum_{i=0}^{k/2} (k - 2i) \cdot \binom{k}{i} \cdot 2^{-k+1} = 2^{-k+1} \cdot (k \cdot A_k - 2B_k) \tag{A.16}$$

For A_k we have:

$$A_k = 2^k - \sum_{i=\frac{k}{2}+1}^{k} \binom{k}{i} = 2^k - (A_k - \gamma_k)$$

Hence,

$$A_k = 2^{k-1} - \frac{\gamma_k}{2} \tag{A.17}$$

Also B_k can be re-written as:

$$B_k = k \cdot \sum_{i=0}^{\frac{k}{2}-1} \binom{(k-1)}{i} = k \cdot 2^{k-2} \tag{A.18}$$

Using A.17 and A.18 to equation A.16 we get:

$$\tilde{u} = k \cdot 2^{-k+1} \cdot (2^{k-1} + \frac{\gamma_k}{2}) - 2^{-k+2} \cdot 2^{k-2} \cdot k$$

That is,

$$\tilde{u} = k \cdot 2^{-k} \cdot \gamma_k \tag{A.19}$$

Using Stirling's approximation, it is easy to see that

$$\gamma_k \sim 2^k \cdot \sqrt{\frac{2}{\pi k}} \tag{A.20}$$

Substituting for equation A.16, we finally have

$$\tilde{u} \sim \sqrt{\frac{2}{\pi}} \cdot \sqrt{k}, k\, even \tag{A.21}$$

A similar result holds for k odd. Therefore since the average number of matched pairs are given by $E[M_R] = k - \tilde{u}$ we finally get:

$$E[M_R] = |R| - \Theta(\sqrt{|R|}) \tag{A.22}$$

□