Optimizing Kidney Exchange with Transplant Chains: Theory and Reality

John P. Dickerson　Ariel D. Procaccia　Tuomas Sandholm

Department of Computer Science
Carnegie Mellon University
{dickerson,arielpro,sandholm}@cs.cmu.edu

ABSTRACT

Kidney exchange, where needy patients swap incompatible donors with each other, offers a lifesaving alternative to waiting for an organ from the deceased-donor waiting list. Recently, *chains*—sequences of transplants initiated by an altruistic kidney donor—have shown marked success in practice, yet remain poorly understood. We provide a theoretical analysis of the efficacy of chains in the most widely used kidney exchange model, proving that long chains do not help beyond chains of length of 3 in the large. This completely contradicts our real-world results gathered from the budding nationwide kidney exchange in the United States; there, solution quality improves by increasing the chain length cap to 13 or beyond. We analyze reasons for this gulf between theory and practice, motivated by our experiences running the only nationwide kidney exchange. We augment the standard kidney exchange model to include a variety of real-world features. Experiments in the static setting support the theory and help determine how large is really "in the large". Experiments in the dynamic setting cannot be conducted in the large due to computational limitations, but with up to 460 candidates, a chain cap of 4 was best (in fact, better than 5).

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems; J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Economics, Theory, Experimentation

Keywords

Kidney exchange

1. INTRODUCTION

The role of kidneys is to filter waste from blood. Kidney failure results in accumulation of this waste, which leads to death in months. One treatment option is dialysis, in which the patient goes to a hospital to have his/her blood filtered by an external machine. Several visits are required per week, and each takes several hours. The quality of life on dialysis can be extremely low, and in fact many patients opt to withdraw from dialysis, leading to a natural death. Only 12% of dialysis patients survive 10 years [18].

Instead, the preferred treatment is a kidney transplant. Kidneys are by far the most common organ to transplant—more prevalent than all other organ transplants combined. Unfortunately, the demand for kidneys far outstrips supply. In the United States alone, in 2010, 4,654 people died waiting for a life-saving kidney transplant. During this time, 34,418 people were added to the national waiting list, while only 10,600 people left the list by receiving a deceased-donor kidney. The waiting list has 89,808 people, and the median waiting time is between 2 to 5 years, depending on blood type [16].

For many patients with kidney disease, the best option is to find a *living* donor, that is, a healthy person willing to donate one of his/her two kidneys. Although there are marketplaces for buying and selling living-donor kidneys, the commercialization of human organs is almost universally regarded as unethical, and the practice is explicitly illegal in most countries. However, in most countries, live donation is legal, provided it occurs as a gift with no financial compensation. In 2010, there were 5,467 live donations in the US.

The number of live donations would have been much higher if it were not for the fact that, in most cases, a potential donor and his intended recipient are blood-type or tissue-type incompatible. In the past, the incompatible donor was sent home, leaving the patient to wait for a deceased-donor kidney. This is where kidney exchanges come into play, in which patients can swap their incompatible donors with each other, in order to each obtain a compatible donor. While still in their infancy, kidney exchanges have now been fielded at the regional and national level.

In this paper, we consider altruistic chains, a recent innovation for barter exchanges that has been widely adopted for kidneys, but is poorly understood. Section 2 describes the formal exchange *clearing problem* and why chains exacerbate the already computationally intractable problem. Section 3 reports results from the first (and only) nationwide kidney exchange, using our fielded technology; these real-world results clearly show the benefit of integrating chains into the clearing process. Section 4 formalizes the theoretical benefit of chains as a kidney exchange scales to the large, and Section 5 experimentally determines exactly what "large" means. Section 6 studies the dynamics of kidney exchange over time, using an extension over the state-of-the-art model to more accurately represent the realities of modern kidney exchange.

2. THE CLEARING PROBLEM

One can encode an n-patient kidney exchange (and almost any n-agent barter exchange, such as Netcycler for used goods, Read-It-Swap-It for used books, the National Odd Shoe Exchange, and Intervac for exchanging time in holiday homes) as a directed graph $G(n)$ as follows. Construct one vertex for each patient. Add a weighted edge e from one patient v_i to another v_j, if v_j wants the item of v_i. In the context of kidney exchange, the item is a kidney

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

from a donor that v_i brings with him into the exchange; the donor is willing to give a kidney if and only if v_i receives a kidney. The weight w_e of edge e represents the utility to v_j of obtaining v_i's item. In kidney exchange, the methodology for setting weights is decided by the exchange design committee. The weights take into account such considerations as age, degree of compatibility, wait time, and geographic proximity. A cycle c in this graph represents a possible swap, with each agent in the cycle obtaining the item of the next agent. The weight w_c of a cycle c is the sum of its edge weights. An *exchange* is a collection of disjoint cycles. (They have to be disjoint because no donor can give more than one kidney.)

The vanilla version of the *clearing problem* is to find a maximum-weight exchange consisting of cycles with length at most some small constant L (typically, $2 \leq L \leq 5$). This cycle-length constraint is crucial. For one, all operations in a cycle have to be performed simultaneously; otherwise a donor might back out after his incompatible partner has received a kidney.[1] The availability of operating rooms, doctors, and staff thus constrains cycle length.

The clearing problem with $L > 2$ is NP-complete [1]. Yet significantly better solutions can be obtained by just allowing cycles of length 3 instead of allowing 2-cycles only [12]; in practice, a cycle length cap of 3 is typically used. Using a mixed integer program (MIP) where there is a decision variable for each cycle no longer than L and constraints that state that accepted cycles are vertex disjoint, combined with specialized branch-and-price MIP solving software, the (3-cycle) problem is solvable to optimality in practice at the projected steady-state nationwide scale of 10,000 patients [1]. In all our experiments, we use that algorithm as a subroutine.

A recent innovation in kidney exchange is *chains* [13, 9, 10]. Each chain starts with an *altruistic* donor—that is, a donor who enters the pool, without a candidate, offering to donate a kidney to any needy candidate in the pool. Chains start with an altruist donating a kidney to a candidate, whose paired donor donates a kidney to another candidate, and so on. Chains can be longer than cycles in practice because it is not necessary (although desirable) to carry out all the transplants in a chain simultaneously.[2] Already, chains of length ten or more have been reported in practice [10]. To our knowledge, all kidney exchanges in the US now use chains (in fact, the National Kidney Registry is using chains only and no cycles). In our experience, roughly 5% of the pool is altruistic.

There are many more feasible chains in a network than cycles—because one does not have to find a way to close a chain into a cycle. The straightforward way to incorporate chains into the optimizer is to add from the end of each potential chain a fake edge of weight 0 to every vertex that represents an altruist. This way, chains look exactly like cycles to the solver and are handled correctly. Unfortunately, due to the removal of the cap of 3 on cycle length, this approach does not scale even remotely to the nationwide level. Rather, it currently scale only to around 200 patients, depending on the cap on chain length. (Of course, if the chain length cap is lower than the cycle length cap, then chains do not significantly increase the complexity.)

3. NATIONWIDE KIDNEY EXCHANGE

Starting around 2003, several regional kidney exchanges have gone live in the US. Two examples include those run by the Alliance for Paired Donation and the Paired Donation Network. However, in 2008, the United Network for Organ Sharing (UNOS)—which controls all organ transplantation in the US—initiated the formation of a *nationwide* kidney exchange. The benefits of such a large-scale exchange are numerous (see, for instance, [5]), and it is ubiquitously accepted that one centralized exchange is better than fragmenting the market into separate exchanges. The UNOS nationwide kidney exchange pilot went live with 77 transplant centers in October 2010, and uses our algorithms and software to conduct a match run every month. Starting in May 2011, chains were incorporated into the UNOS pilot program. Currently, the cycle cap is 3, while the chain cap was 20 and is now being increased to infinity if it turns out to be computationally feasible.

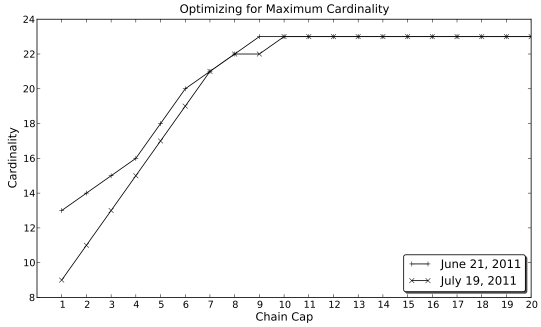

Figure 1: Real data from the June/July 2011 UNOS match runs, optimized for maximum cardinality.

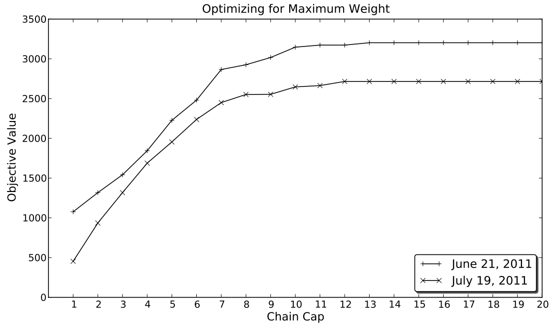

Figure 2: Real data from the June/July 2011 UNOS match runs, optimized for maximum total weight.

Figures 1 and 2 show results for two real matches, for June and July 2011. To show the efficacy of chains, we varied the chain cap from 1 (i.e., the altruistic donor donates directly to the deceased waiting list) to 20. In Figure 1, we maximize the cardinality of the final matching. That is, we ignore edge weights and assume all compatible matches are equally good, and determine the matching that allocates kidneys to the most candidates. The size of the matching increases significantly with chains up to length 9 (June) or 10 (July). Critically, with long chains we match 1.77 (June) and 2.55 (July) times the number of candidates than would have been matched with 3-cycles alone. We note that Ashlagi et al. [3] independently report similar findings from real-world data sets.

The improvement from long chains is even more drastic when the edge weights are taken into account, as is the case in the real UNOS match run. Figure 2 shows that in June, chains of length up to 13 increase the objective value, while chains of length up to 12 increase the objective of the matching in July. Overall, incorporating chains increases the objective value to 2.98 (June) and 6.00 (July) times that of chains only (with a cycle cap of 3).

It is important to note that the structure of the compatibility

---

[1] Such backing out cannot be prevented by legal means because it is illegal to contract for an organ in most countries.

[2] Unlike in a cycle, if a chain breaks by some donor backing out, the chain merely stops, but no patient-donor pair is out their “bargaining chip" (donor kidney).

graph, $G(n)$ in this early pilot program is special and, in many ways, computationally fortuitous. The current UNOS pool consists mainly of *highly sensitized* patients—that is, patients that are difficult to match based on their tissue type. Intuitively, these patients were too hard to match regionally and in prior runs of the national exchange—so the input graph is very sparse. Our other experiments have shown that with a less sensitized pool, we often cannot even solve the current problem size (with long chains) because the input graph $G(n)$ is not as sparse. Luckily, in the next section, we show theoretical results stating that in large kidney pools drawn from the full set of candidates (i.e., not just highly sensitized ones), long chains will have negligible effect on the overall cardinality of the matching with high probability. Therefore, one may not need to consider long chains in the clearing. This would be desirable in practice because short chains are (1) computationally dramatically more tractable for the clearing algorithm (there are fewer of them), (2) logistically easier to administer, and (3) less likely to fail due to a positive crossmatch or some non-simultaneous donor backing out (these two issues will be discussed later).

4. THEORETICAL BOUNDS ON CHAINS

In this section, we prove that using chains of length more than 3 provides no benefit in large, random, unweighted candidate pools. We will prove this result in the most common model of kidney exchange. We begin by describing the model.

4.1 Necessary background & model

The need for kidney exchange exists due to the myriad of immunological incompatibilities that can be present between a candidate and any potential donor. For instance, the *blood type* of a donor kidney can result in acceptance or outright rejection in a possible candidate. At a high level, human blood is split into four types—O, A, B, and AB—based on the presence or absence of the A and B proteins. While other complications may arise, a type O kidney can be transplanted into any candidate; type A and B kidneys can be transplanted into A and B candidates respectively, or an AB candidate; and type AB kidneys are limited to only type AB candidates. Therefore, some candidates are more difficult to match with a random donor than others. O-candidates are the hardest to match because only O-type kidneys can be given to them. Similarly, O-donors are the easiest to match.

With this in mind, candidate-donor pairs in the matching pool can be labeled based on their blood types using the *ABO model*; it is the *de facto* model for theoretical market design work on kidney exchange (see, e.g., [2, 5, 7, 15, 17]). An *under-demanded* pair is any pair such that the donor is not ABO-compatible with the candidate. Furthermore, if these pairs contain only type A and B blood (e.g., the candidate is type A and the donor is type B), the pair is called *reciprocal*. Any pair in the pool such that the donor is ABO-compatible with the candidate is called *over-demanded*. Furthermore, if a donor and candidate share the same blood type, they are a *self-demanded* pair. Intuitively, under-demanded and reciprocal pairs are "harder" to match than over-demanded and self-demanded pairs. In the ABO model, all compatible transplants are considered to be equally good (i.e., those edges have weight 1 each) and typically results in the ABO model are derived in the limit, when the number of pairs of each kind approaches infinity.

If blood type compatibility were the only requirement for a successful kidney donation, over-demanded and self-demanded pairs would have no need to enter the exchange pool because they could simply conduct the transplant within the pair. However, further complications force their hand: the people in a pair are usually incompatible due to tissue type. Tissue type, in particular what is known as HLA type, is measured as a combination of six proteins. Each potential candidate and potential donor must be tested for preformed antibodies against these six proteins; this needs to be done at least once a month because the antibody state of a person changes over time. An increase in the mismatches between donor and candidate HLA types decreases the likelihood of a successful kidney transplant, and can render a donor and candidate incompatible. These kinds of blood tests where measurements are taken separately from the donors and the patients are called *virtual crossmatch* for reasons that will become obvious in the next paragraph.

An important challenge is that medical knowledge is incomplete: even if a patient and donor are compatible based on the virtual crossmatch (so there is an edge in the input graph), in reality they might not be compatible (i.e., the edge might not be usable). This is determined days before the operation by conducting a test called a *crossmatch*: blood from the patient and blood from his/her planned donor are mixed together and if the mixture coagulates, they are incompatible. Such an unfortunate, but very common, occurrence is called a *positive crossmatch*. Positive crossmatch-sensitive models have only recently begun to appear in the literature, and have not included a study of chains [5, 15].

We will say that if an altruist donates directly to the deceased-donor waiting list, that constitutes a chain of length 1. If an altruist donates to a pair, whose donor donates to the deceased-donor waiting list, that constitutes a chain of length 2. If an altruist donates to a pair, whose donor donates to a pair, whose donor donates to the waiting list, that constitutes a chain of length 3, and so on. We are now ready to prove the main theoretical result of this paper.

4.2 Short chains suffice (in theory)

In this section, we use the canonical model for generating kidney exchange data [5]. It works as follows. We start with $G(n)$, a large compatibility graph representing a kidney exchange as described above. The set of n incompatible patient-donor pairs is partitioned into subsets $V_{X\text{-}Y}$ of type X-Y, for each combination of blood types X and Y of the patient and donor respectively. For each blood type X we denote the set of altruistic donors with that blood type by V_X, but make no assumptions about the size of these sets. We assume that a donor and a patient who are blood type compatible are tissue type incompatible with constant probability $\bar{\gamma}$, corresponding to the virtual crossmatch described above. The frequency of each blood type X is denoted by μ_X.

We are now ready to state our main theoretical result. It extends the recent results of Ashlagi and Roth [5] to the setting with chains.

THEOREM 1. *Assume that $\bar{\gamma} < 2/5$, $\mu_O < 3\mu_A/2$, and $\mu_O > \mu_A > \mu_B > \mu_{AB}$. Then with high probability $G(n)$ has an efficient allocation (i.e., one that saves as many patients as possible) that uses only cycles of length at most 3 and chains of length at most 3.*

The proof follows from three lemmas. The first lemma is a trivial simplification and extension of Lemma 9.5 of Ashlagi and Roth [5], which is a generalization of a classic theorem by Erdös and Rényi. To understand the lemma, denote by $G(n,p)$ a random graph with n vertices where an edge exists between two vertices with probability at least p. For a vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_r)$ where $\alpha_i \geq 0$ for $i = 1, \ldots, r$ let $G(\vec{\alpha}, n, p)$ be an r-partite graph with r sets of vertices $V_1, \ldots, V_r$ where $|V_i| = \alpha_i \cdot n$ for $i = 1, \ldots, r$, and a directed edge between $v \in V_i$ and $v' \in V_{i+1}$ for $i = 1, \ldots, r-1$, or between $v \in V_r$ and $v' \in V_1$, exists with probability at least p. A *perfect allocation* in a graph $G(n,p)$ matches all the vertices; a perfect allocation in $G(\vec{\alpha}, n, p)$ (consisting of cycles of length r) matches all the vertices in the smallest vertex set V_i for $i = \text{argmin}_j |V_j|$.

Deviating from [5], define $G'(\vec{\alpha}, n, p)$ similarly to $G(\vec{\alpha}, n, p)$, except that there are no edges between V_r and V_1. An allocation in $G'(\vec{\alpha}, n, p)$ consists of chains of length r that originate in a vertex in V_1. As before, a perfect allocation in $G'(\vec{\alpha}, n, p)$ matches all the vertices in the smallest vertex set V_i for $i = \operatorname{argmin}_j |V_j|$.

LEMMA 1 (ASHLAGI & ROTH [5]). *Let $p > 0$. Then $G(n, p)$ admits a perfect allocation that uses cycles of length at most 3 with high probability. In addition, for any vector $\vec{\alpha}$ as above, the random graphs $G(\vec{\alpha}, n, p)$ and $G'(\vec{\alpha}, n, p)$ admit a perfect allocation with high probability.*

Using Lemma 1, we can assume that if we single out several large groups of vertices (in a large random compatibility graph) that correspond to blood type compatible pairs, there will be sufficiently many edges to admit a perfect matching. For example, if there are large sets of AB-O pairs, O-A pairs, and A-AB pairs, then with high probability we can find an allocation that consists of 3-cycles that matches all the vertices in the smallest set. Even if we consider several such allocations sequentially, by applying the union bound we can see that they all exist with high probability. This essentially allows us to assume in the proof of the next lemma that any two vertices that are blood type-compatible are connected by an edge.

LEMMA 2. *Let $G(n)$ be a random graph that admits the following allocation:*

1. *Every self-demanded pair is matched in 2-way or 3-way cycles with other self-demanded pairs.*
2. *Every B-A pair is matched in a 2-way cycle with an A-B pair.*
3. *Every A-B pair that is not matched to a B-A pair is matched in a 3-way cycle with an O-A pair and an A-AB pair.*
4. *For $X \in \{A, B\}$, every over-demanded pair X-O is matched in a 2-way cycle with an O-X pair.*

Then with high probability $G(n)$ admits an efficient allocation that uses cycles of length at most 3 and chains of length at most 3.

PROOF SKETCH. We complete the allocation described in the lemma's statement to an efficient allocation. Figure 3 visualizes the augmented allocation; regular edges are assumed by the lemma's formulation while dashed edges are added during this proof. Let V^1 be the set of vertices not matched by the initial allocation. First, as many A-donors as possible donate to A-AB pairs and as many B-donors as possible donate to B-AB pairs (shown in Figure 3 by dashed edges from A-altruists to A-AB pairs and from B-altruists to B-AB pairs). In both cases, one of the two vertex sets will be exhausted. More formally, using Lemma 1 we find a perfect allocation for the subgraph induced by V_A^1 and $V_{\text{A-AB}}^1$, and similarly we find a perfect allocation for the subgraph induced by V_B^1 and $V_{\text{B-AB}}^1$.

Let V^2 be the vertices not matched by previous allocations. We find as many 3-way $(\text{AB-O}, \text{O-A}, \text{A-AB})$ cycles as possible, that is, we find a perfect allocation for the subgraph induced by $V_{\text{AB-O}}^2$, $V_{\text{O-A}}^2$, and $V_{\text{A-AB}}^2$. It may be the case that $V_{\text{A-AB}}^2 = \emptyset$. Let V^3 be the set of vertices not matched by previous allocations. Next we find a perfect allocation with 3-way $(\text{AB-O}, \text{O-B}, \text{B-AB})$ cycles. It may be the case that $V_{\text{AB-O}}^3 = \emptyset$ or $V_{\text{B-AB}}^3 = \emptyset$.

Let V^4 be the vertices not matched by previous allocations. The next component in the constructed allocation matches as many O-donors as possible in chains of length 3 of the form $(\text{O}, \text{O-A}, \text{A-AB})$ and then $(\text{O}, \text{O-B}, \text{B-AB})$. This is done sequentially as above. Finally, we match the remaining O-donors and AB-O pairs with remaining under-demanded pairs via chains of length 2 or 2-way cycles (not shown in Figure 3).

Each of the allocations constructed above exists with high probability; thus (by applying the union bound) they all exist with high probability. To complete the proof, we argue that our construction gives rise to an efficient allocation. Since under our construction all over-demanded, self-demanded, and reciprocally demanded pairs are matched, it is sufficient to show that no allocation can match more under-demanded pairs.

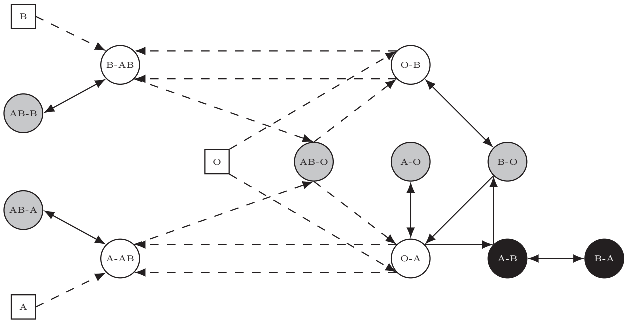

Figure 3: Accompanying figure to Lemma 2. Altruists are shown as rectangles; candidate-donor pairs as ovals. Over-demanded pairs are gray, under-demanded are white, and reciprocal pairs are black. Regular edges appear in the lemma's formulation and dashed edges are constructed in the proof.

Following Ashlagi and Roth [5], when vertex v participates in an exchange with under-demanded vertex v' we say that v *helps* v'. Self-demanded and reciprocally demanded pairs cannot help under-demanded pairs without involving donors or over-demanded pairs. Similarly, AB-donors cannot help under-demanded pairs. In addition, only two types of vertices can help two under-demanded pairs: AB-O pairs can participate in cycles with one of O-A and O-B and one of A-AB and B-AB, and O-donors can start a chain with the same types. Any other vertex can help at most one under-demanded pair, and in particular over-demanded pairs of type X-$Y \neq$ AB-O can only help under-demanded vertices of type Y-X.

Now, A-donors can only help A-AB pairs, and B-donors can only help B-AB pairs. Therefore, it is optimal to match these donors with their respective under-demanded pairs. Finally, in our constructed allocation as many AB-O pairs and O-donors as possible are helping two under-demanded pairs each, while the rest are helping one under-demanded pair each. ☐

The following lemma directly follows from Proposition 5.2 of [5], and holds under the assumptions of Theorem 1.

LEMMA 3 (ASHLAGI & ROTH [5]). *$G(n)$ has an allocation as in Lemma 2, up to symmetries between A-B pairs and B-A pairs, with high probability.*

4.3 Discussion

Theorem 1 follows from the proofs of the three lemmas in Section 4.2. The theorem itself is motivated by the recent work of Ashlagi and Roth [5]. One has to be careful, though, not to use the exact allocation constructed in Proposition 5.2 of their paper as a starting point for the efficient allocation that involves altruistic donors. Indeed, given that $|V_{\text{A-B}}| \geq |V_{\text{B-A}}|$, Ashlagi and Roth match AB-O pairs in cycles $(\text{AB-O}, \text{O-A}, \text{A-AB})$. However, because we are essentially making no assumptions regarding $|V_A|$ and $|V_B|$, it may be the (admittedly extreme) case that there are many (say an infinite supply) of A-donors, few B-donors, few O-donors, and a large number of unmatched under-demanded pairs of type O-B and B-AB. In that case we would rather have the A-donors donate to A-AB pairs while creating cycles $(\text{AB-O}, \text{O-B}, \text{B-AB})$. Therefore, we must match AB-O pairs only *after* matching altruistic donors.

The presence of (even short) chains allows us to avoid a negative property of the efficient allocation constructed by Ashlagi and Roth [5]: that it never matches O-AB pairs. These are, in a sense, the "most" under-demanded pairs in that their candidates are hardest to match, while their donors are least capable of finding a match. In our allocation, AB-O pairs and O-donors that cannot participate in 3-cycles can donate to O-AB pairs without affecting the size of the matching. More precisely, if there are sufficiently many donors to fully match one of the sets $V_{\text{O-A}}$ and $V_{\text{A-AB}}$, and one of the sets $V_{\text{O-B}}$ and $V_{\text{B-AB}}$, then an efficient allocation can match O-AB pairs.

Independent work by Ashlagi et al. [3] attempts to explain the observed benefit of longer chains by considering a theoretical model with highly sensitized patients. Specifically, the probability of tissue type compatibility is allowed to decrease with the size of the graph n. Among other results, it is shown that for any k there exists a small enough probability of compatibility such that chains of length $k+1$ are strictly better than chains of length k. However, to even derive such a statement for chains of length 5 versus chains of length 3, the probability must be as small as c/n for some constant c, whereas intuitively this probability should be a constant that does not depend on n. Hence, despite the elegance of their results, the assumptions underlying their model may be hard to justify.

5. EXPERIMENTAL VALIDATION

The theoretical results from Theorem 1 are strong in that they limit the utility of chains to those of length 3 or fewer—as the graph grows to infinity. In this section we study the disconnect between that theorem and the real-world results from the recent UNOS kidney match runs (Figures 1 and 2).

There are three potential reasons for this disconnect: (1) the theory applies in the large, and the UNOS exchange is not yet large enough for the theory to have taken hold, (2) the model that each blood type compatible edge fails tissue type compatibility independently and with equal probability is a poor model of the (highly sensitized) UNOS pool, and (3) the theory assumes all edges have equal weight, while in the UNOS exchange, edges are weighted.

The discrepancy between the theory and the fielded results cannot be explained solely by the fact that the theory model uses unweighted edges while the real UNOS data has edge weights. If that were the main difference, we would see the curves in Figure 1 reach their maxima at a chain cap of 3. This is not the case. So, we see that even if all the weights were binary, long chains would produce a significant benefit in practice. The difference can, in part, be attributed to the highly structured and very small UNOS pool. This is the product of the newness of the UNOS pilot program; as the exchange grows, we expect the compatibility graph's structure to converge to one similar to our theoretical model.

In reality, the input graph $G(n)$ cannot grow infinitely; specifically, in kidney paired donation, it has been estimated that in steady state the fully fielded nationwide exchange will have around 10,000 pairs at any one time. In this section, we experimentally determine just how large the candidate pool needs to be for the chain cap prescribed by Theorem 1 to apply.

The minimum size of this compatibility graph needed for the theory to take hold depends on the probability distribution of blood and HLA types in the candidate and altruist pools, the number of candidates in the graph, and the number of altruists. We will vary both the number of candidates and altruists, but choose to focus only on blood and HLA types representative of the US population (which serves the current nationwide kidney exchange).

Here we generate candidate-donor pairs and altruists via the most advanced and commonly used data generator for kidney exchange today, by Saidman et al. [14]. This generator incorporates the blood types from the ABO model discussed earlier. It also incorporates an abstract model of tissue types to compute a type of score that quantifies the likelihood of a specific candidate being tissue type compatible with a random donor. In other words, this tissue type model is more refined than assuming all blood type compatible edges are tissue type incompatible with equal probability.

5.1 Increasing the candidate pool size

In the first set of experiments, we explore the effect of a large number of *candidates* on the efficacy of long chains. We hold the number of *altruists* constant at 1, 5, or 10 for each experiment.

Figures 4, 5, and 6 show that larger pools match a higher percentage of candidates, leveling out at roughly 62% in compatibility graphs with a couple hundred candidates. At a high level, this is a strong argument for a national kidney exchange to replace the set of smaller regional exchanges; see [11] for similar arguments. These figures also make a case for the inclusion of chains in pools at both the regional and national level. Figure 5 shows that, for generated pools of size 256, the optimal matching with a chain cap of 1 (i.e., altruists donating directly to the deceased waiting list, avoiding the paired candidate pool entirely) matches nearly 4% fewer candidates overall than matching with a chain cap of 3. The case is more drastic as the number of altruists increases; for instance, Figure 6 shows a 5% decrease on compatibility graphs of the same size. The effect of altruists on the pool is discussed further in the next section.

From above, we can now ignore matchings that only include chains of length 1 and 2; capping chains at either of these levels would result in fewer candidates being matched. Figures 7, 8, and 9 show the expected number of extra transplants resulting from matches incorporating chains of length 4 and 5, compared to only considering chains of up to length 3. Clearly, the maximum number of additional transplants offered by increasing the chain cap by 1 is proportional to the number of altruists present in the graph. For example, for a graph with a altruists, incorporating 5-chains can provide a benefit of at most $2a$ matches over incorporating at most 3-chains; similarly, increasing the cap from 3 to 4 results in at most a extra matches. Figures 7 and 8 show that at pool sizes of 256 with $a = 1$ and $a = 5$, the expected number of additional transplants for either 4- or 5-chains is nil (over 100 generated compatibility graphs). Figure 9 shows similar results while exemplifying another behavior: as the number of altruists increases, the size of the pool required so that limiting the mechanism to 3-chains is satisfactory increases. This behavior is explored further in the next section.

Figures 8 and 9 initially show an *increase* in the utility of longer chains as the graph size moves from very small (e.g., 16 candidates) to slightly larger (e.g., 32–64 candidates).[3] This is a side effect of the number of altruists present relative to the size of the pool. With a high enough ratio of altruists to candidates, altruists can "flood" the matching, an idea explored further in the next section.

All of the experiments validate the theory: there seems to clearly be a pool size beyond which long chains do not help.

5.2 Increasing the number of altruists

In the previous subsection, we held the number of altruists constant while increasing the size of the candidate pool. We now explore the opposite, allowing ever increasing numbers of altruists to enter candidate pools of constant size.

As the number of altruists increases relative to the size of the candidate pool, the expected number of candidates matched rises to 100%, as shown in Figures 10, 11, and 12. This full flooding of the pool to create a complete matching, while interesting, is not

---

[3] In Figure 9, the computational demands of this experiment precluded us from extending the dotted line past 128 candidates.

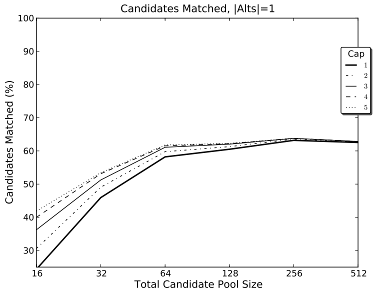

Figure 4: Total percentage of candidates matched as #candidates increases across various chain caps, #altruists=1.

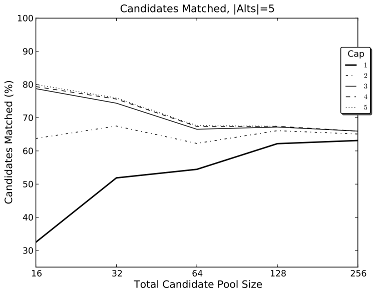

Figure 5: Total percentage of candidates matched as #candidates increases across various chain caps, #altruists=5.

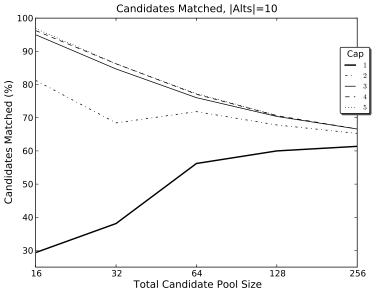

Figure 6: Total percentage of candidates matched as #candidates increases across various chain caps, #altruists=10.

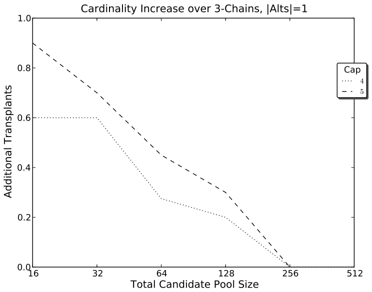

Figure 7: Cardinality increase over 3-chains for 4- and 5-chains, #altruists=1.

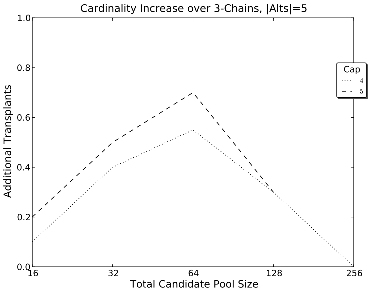

Figure 8: Cardinality increase over 3-chains for 4- and 5-chains, #altruists=5.

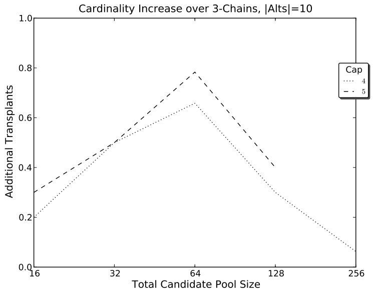

Figure 9: Cardinality increase over 3-chains for 4- and 5-chains, #altruists=10.

presently a realistic scenario; all three tested compatibility graph sizes would require around 50% as many altruists as candidates in the pool (Figure 12 has the x-axis cut short). In our experience with UNOS, the number of altruists is typically around 5% the size of the candidate pool. Increasing this number could feasibly change as the exchange grows in size and publicity, paying special notice to the ethical issues that arise in coercion of possible donors.

6. DYNAMIC KIDNEY EXCHANGE

In the paper so far, we have studied static models. We now discuss the dynamics of a kidney exchange running month to month.

6.1 Augmenting the model

We augment the model in several ways to make it capture the nuances that have arisen in practice.

Dynamics. Most of the work in kidney exchange has focused on a single-shot optimization on a static pool. This deviates from reality in that matching should occur *dynamically*. In reality, candidates arrive and depart from the pool. Even with dialysis, only 12% of patients survive 10 years [18]; this gives us the monthly death rate we use in our experiments. Timeliness in matching is clearly important. Our experimental results, discussed later, perform matching over 24 months using a changing kidney pool.

Some work in this area has been done already. Ünver [17] derives an efficient mechanism in the dynamic setting for a simplified model of kidney exchange that can be solved analytically. Awasthi and Sandholm [6] apply the model discussed above to the dynamic setting, using trajectory-based optimization to look into the possible futures and then use optimization technology to determine transplants for the current period, including chains.

Work by Gentry et al. [8] on simulated data and Ashlagi et al. [4] on real-world data explores the trade-offs between two types of chain execution polices. The first chain type is executed in its entirety in one time period, with the leftover donor donating to the waiting list. An alternative is to split long chains into segments with intra-segment simultaneous transplants, but the segments execute one after another. The left over donor (aka *bridge donor*) from one segment then serves as a virtual altruist for the next segment. These two types of chains perform differently under the presence of *renege rates*—that is, when a bridge donor decides to leave the pool before donating a kidney. However, no reliable quantification of a renege rate exists due to the infancy of kidney exchanges.

While Gentry et al. [8] do not explicitly consider chain caps, Ashlagi et al. [4] do; they experimentally show that longer (up to length 6) chains can, in fact, help. Our work uses a similar model with single-shot execution chains and, importantly, takes into account the policies of the UNOS nationwide kidney exchange. As we will show, this addition results in different matching behavior. We now discuss these UNOS-specific additions to the model.

Individual crossmatch sensitivity. As exemplified in the real, highly-sensitized UNOS candidate pool, candidates can have widely varying susceptibility to incompatibilities in kidney donation. The Saidman et al. model from the previous section has a rather realistic view of virtual crossmatch failures, and we use that model here.

In addition, here we do (non-virtual) crossmatches for all the planned transplants just before the transplant takes place, as in reality. This is again done using the Saidman et al. [14] generator. It provides for each candidate a probability that the candidate is tissue type compatible with a random person. We use that probability to draw crossmatch success versus failure. If the crossmatch fails, the transplant cannot proceed. If it is part of a cycle, the cycle does not execute; the pairs in the cycle go back in the pool. The failed edge is permanently removed from the compatibility graph $G(n)$.

Crossmatching has a significant effect on the size of the "real" matching. Assume an optimal matching (pre-crossmatch) yields a 3-cycle. If *any* crossmatch fails between a candidate and potential donor, the *entire* cycle must be thrown away—since we cannot force a donor to give a kidney if his accompany candidate does not

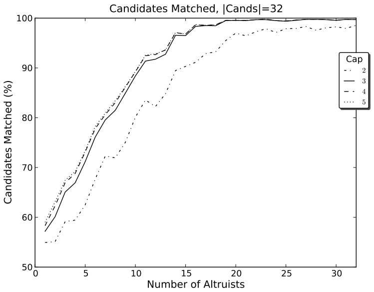

Figure 10: Total percentage of candidates matched as #altruists increases across various chain caps, #cands=32.

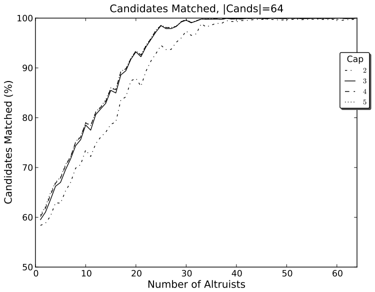

Figure 11: Total percentage of candidates matched as #altruists increases across various chain caps, #cands=64.

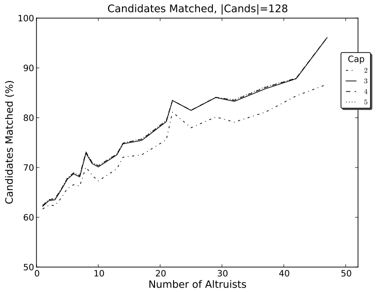

Figure 12: Total percentage of candidates matched as #altruists increases across various chain caps, #cands=128.

receive one. Even more drastic is the case of chains: if, for example, a pre-crossmatch matching yields a 20-chain, any transplants after the first crossmatch failure cannot be performed.

Because of this special case for chains, real-world exchanges have enacted policies for the acceptance or rejection of chains based on their length and the quality of the altruistic donor. O-type altruists are highly valued, as they can (potentially) donate to any blood type, so short chains enabled by O-type altruists should (potentially) be rejected in favor of longer chains in the future. Our experiments follow current UNOS policy which, along with some special cases discussed below, states that (i) chains started by non-O-type altruists are always executed, while (ii) chains triggered by an O-type altruist are executed only if they can be executed to length at least 5 (before there is a crossmatch failure). We will experiment with varying the value away from 5; we will call this parameter k.

Altruists are allowed choices. In the event that an O-type chain is shorter than length 5, the UNOS policy allows for the altruist to decide that the chain be executed anyway. This is due to the fact that altruists do not want to stay in the candidate pool indefinitely, but rather want to move on with their lives and other plans. In UNOS's experience running kidney exchange, altruists typically do not wish to stay active in the pool for more than three months—instead opting to donate directly to the deceased donor waiting list. While exact data on this phenomenon are too sparse at the moment, our experiments use the anecdotal rates (received through UNOS): 75% probability of an altruist requesting execution of a short chain, and a monthly altruist exit rate that corresponds to an expected presence of two months in the pool for each altruist. Our model executes each chain in a single time segment.

6.2 Experimental Results

We now present preliminary results simulating dynamic kidney exchange under the model described above. Figure 13 shows the expected increase in transplants when including chains over the cycles-only approach. The x-axis describes the total number of candidates available during at least one time period over the entire simulation; between 15 and 20 candidates arrive every time period and between 1 and 2 altruists arrive every time period. The initial pool (i.e., the pool at time $t = 0$) is seeded with between 50 and 100 candidates and 5 altruists. These settings roughly mimic the current state of the nationwide UNOS pilot program.

The results both remain true and (appear to) deviate from the theory in a number of ways. The benefit of using chains is immediately obvious; in all cases, even using only 2-chains increases the total number of transplants by 20 or more. However, in this new setting, chains of length at most 3 (at least for the tested pool sizes, number of altruists, etc) do *not* provide equivalent benefit to longer chains. While 3-chains do provide a net gain over 2-chains, considering longer chains helps—sometimes by nearly 10 additional transplants. This increase is surprising because, intuitively, longer chains are less likely to be executed in full (and thus likely to be canceled by the UNOS policy) due to low crossmatch probability. Not executing a chain is dangerous because altruists leave the pool entirely if they remain unmatched for more than a few months.

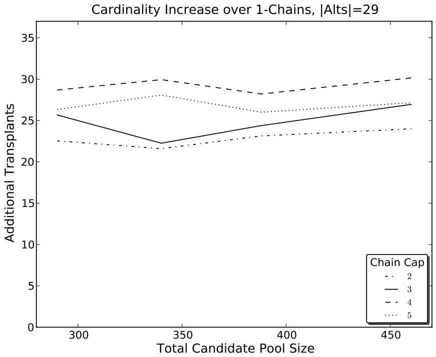

Figure 13: Expected improvement of n-chains over 1-chains (over 24 months).

The results above can be explained by considering the effect of time on an evolving small-scale pool of candidates. Over time, highly sensitized candidates will build up in the pool, since they are often significantly harder to match—both because they have fewer connected edges in the generated compatibility graph and because they are more likely to fail during the crossmatch. Through the real-world results detailed in Section 3, we have seen that the utility of (long) chains increases tremendously in the presence of a small, highly sensitized pool. In Figure 13, chains of length greater than 3 are able to serve highly sensitized candidates because they do not need to "close" the chain, as is the case with a cycle.

Surprisingly, allowing the optimizer to use chains of up to length 5 is strictly worse than constraining it to chains of length at most 4 (while a cap of 4 is better than 3). This suggests that there is diminishing benefit to longer and longer chains, and at the same time there is increasing risk of crossmatch failure (and thereby altruists leaving and candidates dying) with increasing chain cap. The experiments here suggest that in the dynamic setting with these pool sizes (i.e., not in the very large), a chain cap of 4 is best.

We now expand our preliminary experiments to include the chain execution policy from UNOS (see Section 6.1), and we will vary k (between 1 and the chain cap). Intuitively, a higher k will prevent "wasting" a valuable O-altruist on short chains, favoring waiting for a longer, higher-scoring chain instead. Figure 14 shows the effect of varying k as we increase the chain length cap. When considering only short chains, a higher k increases the total number of transplants. In contrast, when chains of length 4 and 5 are considered, it appears better to reduce k. The drop in overall utility from allowing *only* long chains to execute is due to altruists' propensity to leave the pool; if an altruist is not used in an executed chain within a few time period, he/she is likely to leave the pool (and thus be "wasted" by going straight to the deceased donor waiting list instead of saving some lives in the pool first).[4]

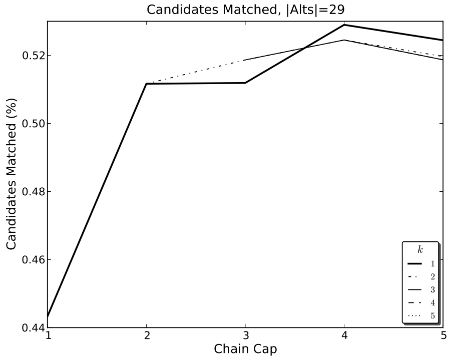

Figure 14: Expected percentage of candidates matched with 1- to 5-chains, varying k in the UNOS chain execution policy.

7. CONCLUSIONS & FUTURE RESEARCH

In this paper, we considered altruist-initiated chains, a recent innovation in barter exchanges that has seen wide adoption in regional and national kidney exchange, but has not been well understood. We described results gathered from the first nationwide kidney exchange in the US that show, for relatively small, highly-sensitized pools of candidates, the benefit of long chains. We then showed that, in the large, the benefit from chains longer than 3 becomes negligible (with high probability) on random compatibility graphs drawn from distributions that mimic the real world population. We supported these theoretical results by extensive experiments using the state-of-the-art instance generator to allow us to experiment on larger instances than exist in current kidney exchanges. The theoretical results take hold in exchanges orders of magnitude smaller than the expected steady-state of the nationwide kidney exchange; this provides evidence for considering only short chains in the large, real-world exchanges we expect to see.

Finally, we experimented in the dynamic setting where the exchange clears every month. We included in the simulations all the known (to us) considerations that have arisen through our work with real kidney exchanges. Computational complexity precluded experiments in the large for the dynamic setting, but in medium-sized pools a chain cap of 4 was best (and strictly better than 5). At any given point in our largest dynamic simulations, 100–150 candidates were present in the pool—others had already been matched, had died, or had not entered the simulation yet. We showed in Section 5 that, at such a small size and with so many altruists, we cannot expect 3-chains to suffice. We believe that, were the pool increased to hundreds of new candidates per month (as is projected to be the case in a fully fielded nationwide exchange), experiments in a dynamic setting would yield results similar to the static setting—with chains of length 3 sufficing.

Many avenues for future research arise from this work. Theoretical results in less abstract models would provide further insight into the efficacy of chains in real world exchanges. Ongoing work by Ashlagi et al. [3] is, to our knowledge, the only other push in this direction; they analyze chains in highly sensitized pools, but under arguable assumptions. Furthermore, advances in clearing algorithms are necessary to handle chains at even the moderate scale; the current state of the art can clear only small candidate pools with just a few altruists. Scaling to the expected size of the nationwide kidney exchange will require algorithmic and computational advances that allow clearing pools orders of magnitude larger than what can be solved today. Restricting attention to short chains may be a promising avenue for tackling that complexity.

Acknowledgments. This work is supported by the National Science Foundation under grants IIS-0905390, IIS-0964579, and CCF-1101668. We acknowledge Intel Corporation and IBM for gifts.

8. REFERENCES

[1] D. Abraham, A. Blum, and T. Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. *EC*, 2007.

[2] I. Ashlagi, F. Fischer, I. Kash, and A. Procaccia. Mix and match. *EC*, 2010.

[3] I. Ashlagi, D. Gamarnik, and A. Roth. The need for chains in kidney exchange. *NBER Market Design Workshop*, 2011.

[4] I. Ashlagi, D. Gilchrist, A. Roth, M. Rees. Nonsimultaneous Chains and Dominos in Kidney-Paired Donation—Revisited. *Am J Transplant*, 2011.

[5] I. Ashlagi and A. Roth. Individual rationality and participation in large scale, multi-hospital kidney exchange. *EC*, 2011.

[6] P. Awasthi and T. Sandholm. Online stochastic optimization in the large: Application to kidney exchange. *IJCAI*, 2009.

[7] I. Caragiannis, A. Filos-Ratsikas, and A. Procaccia. An improved 2-agent kidney exchange mechanism. *WINE*, 2011.

[8] S. Gentry, R. Montgomery, B. Swihart, D. Segev. The roles of dominos and nonsimultaneous chains in kidney paired donation, *Am J Transplant*, 2009.

[9] R. Montgomery *et al*. Domino paired kidney donation: a strategy to make best use of live non-directed donation. *The Lancet*, 2006.

[10] M. Rees, J. Kopke, R. Pelletier, D. Segev, M. Rutter, A. Fabrega, J. Rogers, O. Pankewycz, J. Hiller, A. Roth, T. Sandholm, U. Ünver, R. Montgomery. A nonsimultaneous, extended, altruistic-donor chain. *New Engl J Med*, 2009.

[11] A. Roth, T. Sönmez, and U. Ünver. Kidney exchange. *Q J Econ*, 2004.

[12] A. Roth, T. Sönmez, and U. Ünver. Efficient kidney exchange: Coincidence of wants in a market with compatibility-based preferences. *Am Econ Rev*, 2007.

[13] A. Roth, T. Sönmez, U. Ünver, F. Delmonico, L. Saidman. Utilizing list exchange and nondirected donation through 'chain' paired kidney donations. *Am J Transplant*, 2006.

[14] S. Saidman, A. Roth, T. Sömnez, U. Ünver, F. Delmonico. Increasing the opportunity of live kidney donation by matching for two and three way exchanges. *Transplantation*, 2006.

[15] P. Toulis and D. Parkes. A random graph model of kidney exchanges: efficiency, individual-rationality and incentives. *EC*, 2011.

[16] United Network for Organ Sharing. http://www.unos.org/.

[17] U. Ünver. Dynamic kidney exchange. *Rev Econ Stud*, 2010.

[18] United States Renal Data System. http://www.usrds.org/.

---

[4] Following the acceptance of this paper, UNOS removed the rule that chains triggered by an O-type altruist are executed only if they can be executed to length at least 5. Our experimental results were the reason for this change in policy.