FutureMatch: Learning to Match in Dynamic Environments

John P. Dickerson
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
dickerson@cs.cmu.edu

Tuomas Sandholm
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
sandholm@cs.cmu.edu

ABSTRACT

The preferred treatment for kidney failure is a transplant; however, demand for donor kidneys far outstrips supply. *Kidney exchange*, an innovation where willing but incompatible donor-patient pairs can exchange organs—via cycles and altruist-initiated chains—provides a life-saving alternative to long waiting lists for deceased-donor organs. Typically, fielded exchanges act *myopically*, considering only the current pool of pairs when planning the cycles and chains. Yet kidney exchange is inherently dynamic, with participants arriving and departing. Also, many planned exchange transplants do not go to surgery due to various failures. Thus, it is important to consider the future when matching.

Motivated by our experience running the computational side of the US nationwide kidney exchange—which already includes 56% of US transplant centers—we present FutureMatch, a framework for *learning to match in a general dynamic model*. It uses data from all live kidney transplants in the US since 1987 to learn the quality of each possible match; it then learns the *potentials* of elements of the current input graph offline (e.g., potentials of pairs based on features such as donor and patient blood types), translates these to weights, and performs a computationally feasible batch matching that incorporates dynamic, failure-aware considerations through the weights.

We validate FutureMatch on the real nationwide exchange data. Not only does dynamic matching result in more expected transplants than myopic matching, but also even dynamic matching under economically inefficient objectives that enforce equity can result in statistically significant increases in social welfare over efficient myopic matching.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Health; J.4 [**Social and Behavioral Sciences**]: Economics; J.3 [**Life and Medical Sciences**]: Medical information systems; I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

General Terms

Design, Economics, Experimentation

Keywords

Kidney exchange, dynamic matching, complex matching

1. INTRODUCTION

Chronic kidney disease is a life-threatening health issue that affects millions of people worldwide; its societal burden is likened to that of diabetes [22]. Damage from kidney disease can cause irreparable loss of organ function and, eventually, complete kidney failure. Such failure requires either continual dialysis or an organ transplant to maintain life.

The preferred treatment for kidney failure is transplantation. However, the demand for donor kidneys is far greater than supply. In the US alone, the waiting list for a kidney transplant had 99,276 patients as of February 17, 2014 [30]. Demand is increasing: for example, 34,834 people were added to the US national waiting list in 2012, while only 15,939 left it due to receiving a kidney.

Patients can receive a transplant organ from either a deceased or living donor. Roughly two thirds of transplanted kidneys are sourced from cadavers, while one third come from willing and healthy living donors. Patients who are fortunate enough to find a willing living donor must still contend with *compatibility* issues like blood type, tissue type, and other medical or logistical factors (as we discuss later in the paper). If a willing would-be donor is incompatible with a patient, the kidney cannot be transplanted.

Kidney exchange is a recent innovation that allows patients to swap willing but incompatible donors. Figure 1 shows a graphical representation of a pool consisting of three patient-donor pairs, where an arrow from pair i to pair j means the patient at j is compatible with the donor at i. Also shown is a donor without a paired patient who is willing to donate a kidney altruistically. The basic kidney exchange problem is then to recommend a "good" set of organ swaps.

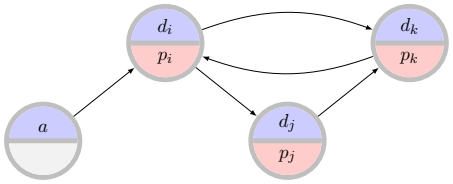

Figure 1: Example kidney exchange pool with three patient-donor pairs and one altruistic donor.

1.1 Our Contributions

Fielded kidney exchange algorithms match patients and donors myopically, considering only the current state of the pool when deciding on a matching. However, kidney exchange is inherently dynamic, with patient-donor pairs and altruists arriving and departing the pool over time. Intuitively, the matching algorithm should take distributional information about the future into account when deciding what action to take now. Indeed, recent research has shown that such *dynamic* matching should result in increased overall social welfare [4, 5, 10, 31]. However, most of that prior work has been either not scalable to the large in terms of the computational requirements [5], or studied in simplified theoretical models [4, 31], while here we study the problem in its full complexity and develop a computationally scalable methodology. Also, the prior work has focused mainly on maximizing the number of matches, while our approach is for any objective. Furthermore, many planned exchange transplants do not go to surgery due to various failures in the last few weeks before the operation [11]. That introduces a second form of dynamism. The approach in this paper is, to our knowledge, the first to address both forms of dynamism.

Motivated by our experience running the computational side of the United Network for Organ Sharing (UNOS) US nationwide kidney exchange, which has grown to include 133 transplant centers since its inception in Oct. 2010, we present FutureMatch, a general framework for learning how to match in dynamic environments. We validate the framework on real data drawn from the nationwide exchange. We find that using FutureMatch *even with economically inefficient objectives*—like maximizing the match size subject to equity constraints—results in significantly higher social welfare than efficient but myopic matching.

Our main contribution is the design, implementation, and validation of FutureMatch, the first data-driven learning framework for *complex* online matching. By "complex" we mean that we are not within the traditional framework of matching where the goal is simply to pair up vertices. We built FutureMatch to satisfy three primary desiderata:

Generality. It is important to provide medical professionals with a *general* framework that is adaptable to recent advances in medical knowledge and legal trends, as has been motivated in the clinical decision support system literature [32]. Being young, the kidney exchange community is still settling on ethical, economic, logistical, and legal best practices. Our framework allows for sensitivity analysis of methods for incorporating new factors as kidney exchange continues to grow. It also allows the objective to be changed, additional donor and patient attributes to be used in graft survival prediction, additional attributes of the input graph (including new donor and patient attributes) to be used in learning potentials, the simulation module to be made increasingly realistic as experience and medical knowledge builds, and of course, the framework can be used to re-learn as new kidney exchange data becomes available.

Scalability. True stochastic optimization would be completely computationally infeasible here. Even approximate variations thereof [5] do not scale adequately for the projected nationwide kidney exchange scale. With this in mind, our framework uses heavy-duty offline learning to collapse the future into a set of value estimates we call *potentials*, which are then implemented at no (or very low) cost to the online clearing engine.

Data centricity. Recent work uses "big data" methodologies for knowledge discovery in healthcare data [8, 33]; to our knowledge, such methodologies have never been applied to kidney exchange. FutureMatch draws on historical data (and optional domain expert input) during initialization to determine its objective function. As a proof of concept for this paper, we learn a novel transplant quality predictor from a dataset consisting of all living donor kidney transplant events in the US since Oct 1, 1987, and discover that some present-day features of living donor transplants do not align with older results from deceased donation in the medical literature [23]. This is noteworthy since today's priority policies have largely been inherited from the UNOS deceased-donor waiting list policies.

The framework also uses a realistic exchange generator during the learning process. Learning matching policies *using data* is especially important in kidney exchange, where the earliest state-of-the-art generative model [28] did not end up aligning with the composition of current fielded kidney exchanges, resulting in initial theoretical and empirical results that were suboptimal or not applicable in practice. We draw on all match runs from the US nationwide exchange.

The overarching goal of kidney exchange is to increase the opportunity for needy patients to find a life-saving kidney. We are actively involved in the development and implementation of new ideas and algorithms in fielded kidney exchanges, and drew on our experience when designing and testing FutureMatch, which combines optimization and data mining for significant societal benefit.

2. BACKGROUND

In this section, we define the standard computational model of kidney exchange, briefly overview related work from the matching theory literature, and discuss the state of the art in currently fielded kidney exchange.

2.1 Kidney Exchange Model

The standard model for kidney exchange encodes a kidney exchange as a directed *compatibility graph* $G = (V, E)$ by constructing one vertex for each patient-donor pair in the pool. An edge e from v_i to v_j is added if the patient in v_j wants the donor kidney of v_i. A paired donor is willing to give her kidney if and only if the patient in her vertex v_i receives a kidney.

The weight w_e of an edge e represents the utility to v_j of obtaining v_i's donor kidney. A cycle c in the graph G represents a possible kidney swap, with each vertex in the cycle obtaining the kidney of the previous vertex. If c includes k patient-donor pairs, we refer to it as a *k-cycle*. For example, the compatibility graph in Figure 1 includes two possible cycles: a 2-cycle between vertex v_i and v_k, and a 3-cycle consisting of vertices v_i, v_j, and v_k. In kidney exchange, cycles of length at most some small constant L are allowed—all transplants in a cycle must be performed simultaneously so that no donor backs out after his patient has received a kidney but before he has donated his kidney. In most fielded kidney exchanges, including the United Network for Organ Sharing (UNOS) nationwide kidney exchange, $L = 3$ (i.e., only cycles of length at most 3 are allowed).

Fielded kidney exchanges gain great utility through the use of chains [25, 27]. Chains start with an altruistic donor donating his kidney to a patient, whose paired donor donates her kidney to another patient, and so on. The compatibility graph in Figure 1 includes four possible chains: $\langle a, v_i \rangle$, $\langle a, v_i, v_j \rangle$, $\langle a, v_i, v_k \rangle$, and $\langle a, v_i, v_j, v_k \rangle$. Chains can be (and typically are) longer than cycles in practice because it is not necessary to carry out all the transplants in a chain simultaneously. There is a chance that a bridge donor backs out of his/her commitment to donate—which has happened (albeit rarely) already in the United States. Cycles cannot be executed in parts because if someone backs out of a cycle, then some pair has lost a kidney (i.e., their "bargaining chip"). In contrast, if someone backs out of a chain, no pair has lost their bargaining chip (although it is unfortunate that the chain ends).

A *matching* M is a collection of disjoint cycles and chains in the graph G. The cycles and chains must be disjoint because no donor can give more than one of her kidneys. Given the set of all legal matchings $\mathcal{M}$, the *clearing problem* in kidney exchange is to find a matching M^* that maximizes some utility function $u : \mathcal{M} \to \mathbb{R}$. Formally:

$$M^* = \arg\max_{M \in \mathcal{M}} u(M)$$

The standard clearing problem for finite cycle cap $L > 2$ is NP-hard [1]. Abraham, Blum, and Sandholm [1] took the first serious computational step toward solving the kidney exchange problem by providing a specialized branch-and-price-based [6] integer program solver; subsequent work by Dickerson, Procaccia, and Sandholm has increased solver speed and generality [11, 12]. We use an adapted version of that clearing algorithm as the batch clearing algorithm module in our framework (as we will discuss later).

In fielded kidney exchanges, one typically finds the maximum weighted cycle cover (i.e., $u(M) = \sum_{c \in M} \sum_{e \in c} w_e$). This *utilitarian* objective can favor certain classes of patient-donor pairs while marginalizing others, a phenomenon that we explore—and help alleviate—in this paper.

2.2 Prior Research on Dynamic Matching

Prior work exists on dynamic matching. However, our work differs from it significantly. For one, FutureMatch learns to match better using data. Also, it applies to a significantly richer set of problems, such as kidney exchange, than traditional matching where the goal is to simply pair up vertices. We now briefly overview some of the prior work.

In the dynamic matching problem, vertices and/or edges arrive and depart over time. Karp, Vazirani, and Vazirani showed that the competitive ratio of any randomized online bipartite matching algorithm is at best $1 - 1/e$, and gave an algorithm that achieves this [16]. The online bipartite matching problem has seen significant renewed interest from Internet advertising firms, since it relates to keyword and display ad allocation (e.g., [13, 19]).

In the *query-commit* problem [14, 21], the goal is to find a matching of maximum size in a graph where the set of edges is not known ahead of time. Instead, an actor may *query* an edge and, if present, is forced to *commit* to using that edge in the final matching. This is relevant to kidney exchange, where a matching algorithm first selects an edge, but that edge can fail for either a medical or logistical reason before transplantation. Blum et al. [7] address a similar problem, where at most two edges incident to any vertex can be selected in a possible matching. Those papers operate in significantly simpler models than real kidney exchange.

Dickerson, Procaccia, and Sandholm [11] show that making the optimization "failure-aware" by incorporating edge failure probabilities directly into the optimization process increases the expected number of transplants both in theory and in practice; we incorporate their (static) model into the FutureMatch framework. Dickerson, Procaccia, and Sandholm [12] also explored the *price of fairness* in kidney exchange, which is a measure of the relative loss in efficiency of the system under various "fair" matching rules. They found that while the theoretical price of fairness is low, in practice it can be non-trivial; we also incorporate this model into FutureMatch (and show that a favorable balance between fairness and efficiency can be struck in practice).

Dynamic kidney exchange is a largely unsolved problem. From the theory side, Ünver provided the first results in a model of dynamic kidney exchange that only includes 2-cycles and no chains [31]. Ashlagi, Jaillet, and Manshadi [4] and Anshelevich et al. [2] look at the *batch matching* problem, where a maximum cardinality matching is performed every k time periods, or whenever some feature of the graph crosses a predetermined threshold. Awasthi and Sandholm [5] approached the problem computationally, using trajectory-based optimization that samples potential future states to inform the present matching algorithm; unfortunately, this does not scale computationally to the projected steady-state size of the nationwide kidney exchange. Motivated by this, Dickerson, Procaccia, and Sandholm [10] proposed learning offline the *potentials* of different elements (e.g., types of vertices or edges) in the input graph, then subtracting out these potentials per element in the objective online. We incorporate this approach, too, into FutureMatch.

2.3 Fielded Kidney Exchanges

The idea of kidney exchange was presented in 1986 [24], while the first organized kidney exchange, the New England Paired Kidney Exchange (NEPKE), started in 2003–2004 [26]. It has since ceased operations and its pool was merged into the UNOS kidney exchange, which started in 2010 and now includes 56% of the US transplant centers. Our earlier matching algorithm is the decision-making engine for the UNOS exchange. All the decisions are transparent and purely computational without human intervention.

There are also two large private kidney exchanges in the US, the National Kidney Registry (NKR) and the Alliance for Paired Donation (APD). They typically only work with large transplant centers. Transplant centers can be part of multiple exchanges. NKR makes their matching decisions manually and APD uses a combination of algorithmic and manual decision making. There was also another large private kidney exchange, the Paired Donation Network (PDN), which has ceased operations. In addition, there are several smaller private kidney exchanges in the US. They typically only involve one or a couple of transplant centers. These include an exchange at Johns Hopkins University and a single-center exchange at the Methodist Specialty and Transplant Hospital in San Antonio. Furthermore, there are now established kidney exchanges in the Netherlands, Canada, and England, as well as nascent ones in Portugal and Israel.

Kidney exchanges started with just using 2-cycles before also allowing 3-cycles and altruist-initiated chains [27]. Since 2006, kidney exchanges have also incorporated never-ending

chains, where the last donor in a chain serves as an altruist in a later match run to initiate a new chain [25]. This approach is now included at least in the three leading kidney exchanges (UNOS, NKR, and APD).

Fielded kidney exchanges perform batch matching. The objective in the batch optimization engine is not to simply maximize the number of matches, but a weighted sum of the matches, where the weights—aka. priority points—are decided by committees of medical personnel, computer scientists, and economists. At UNOS the current prioritization scheme was largely inherited from the US deceased-donor waiting list (which UNOS also runs) prioritization policy. The UNOS priority points take into account the following factors: do the donor and patient have zero antigen mismatch in tissue type, sensitization of the patient, prior organ donor status, pediatric status, wait time, geographic proximity, and other antibody specificities.

3. PROPOSED METHOD

We now present the FutureMatch framework for learning to match in dynamic environments. We begin by motivating and describing the framework at a high level in the first subsection. In the following subsections we discuss the different parts of the framework in detail and how we instantiated them for kidney exchange.

3.1 The FutureMatch Framework

We are interested in learning from demographically accurate data how to match *in the present* such that some overarching objective function is maximized over time. Scalability is important: heavy offline statistics can be computed and periodically updated, but the fielded clearing algorithm must run quickly (within minutes or at most hours).

Figure 2 graphically depicts the FutureMatch framework. A domain expert (e.g., a committee of medical and legal professionals) begins by describing an overall objective function for the exchange. Even *measuring* this objective can be difficult: for example, if the goal is to maximize the number of days added to patients' lives via kidney transplantation, then calculating the relative quality of a proposed match requires knowing some notion of utility for each edge—representing a potential transplant—in the compatibility graph. We propose to learn this edge weight function $w : E \to \mathbb{R}^+$ from data, and give examples for a variety of potential objective functions in Section 3.2.

The learned weight function w is then fed into a parameterized instance generator that mimics the underlying distribution. This generator in turn feeds training and test sets into a system for learning the *potentials* of various element classes in the compatibility graph. Intuitively, given an element θ (e.g., vertex, edge, cycle, or chain type), a potential $P_\theta \in \mathbb{R}$ quantifies the expected utility to the exchange of that element in the future [10]. Potentials are combined with w to quantify an edge-specific quality rating. In Section 3.3, we learn potentials for the combinations of different blood types for patient-donor pairs under each of the weight functions defined in Section 3.2.

Finally, the fielded clearing algorithm incorporates the combined weight function w and set of potentials $\mathcal{P}_\Theta$ into its myopic weighted matching algorithm. This incorporation comes at very low or no cost to the runtime of the clearing algorithm; indeed, the final "potential-aware" input graph is simply a re-weighted version of the original compatibility graph, using the weights that encode the future.

In the rest of the section, we describe an in-depth implementation of FutureMatch. Our goals are twofold: first, to show the general applicability and tractability of the framework, and second, to mimic the UNOS nationwide kidney exchange, with which we are actively involved.

3.2 Encoding an Objective Function

We now discuss in depth the process of defining an objective for FutureMatch. We do this in the context of kidney exchange, but note that the process is general.

The medical and legal communities in kidney exchange are concerned about a wide variety of match characteristics. In our experience, the most frequently discussed include the number of overall matches, the number of overall transplants, the quality of transplants, and whether or not to prefer specific subgroups in the exchange (children, sensitized patients, underrepresented ethnicities) and by how much. Other concerns might include some sort of equitable treatment between participating transplant centers, minimizing legal exposure, and fair compensation.

In this paper, we consider two different kidney exchange models—deterministic, where post-algorithmic match failures are not quantified in the optimization problem and failure-aware, where they are—and three formal matching objectives in each of the two models:

1. MaxCard: Maximize the total number (i.e., cardinality) of patients who are algorithmically matched (in the deterministic model) or receive transplants in expectation (in the failure-aware model);
2. MaxCard-Fair: Maximize the total number of patients who are algorithmically matched (in the deterministic model) or receive transplants in expectation (in the failure-aware model), where "marginalized" patients are weighted in the objective by some constant factor β more than others; and
3. MaxLife: Maximize the total time algorithmically-matched (deterministic) or transplanted (failure-aware) donor organs will last in patients.

Each of these objectives amounts to setting weights on edges in the input graph (e.g., Figure 1). Next, we detail edge weighting algorithms for these example objectives.

In our experience, when committees debate priority points for today's exchanges, the discussions confound the goal and the means. For example, a goal could be to maximize matches and a means could be to prioritize sensitized patients because they are harder to match in the future. On the other hand, many argue that sensitized patients should be inherently preferred, and it seems that most do not make a clear distinction between means and ends. In contrast, FutureMatch clearly separates ends and means. Our objective lives in the space of weights on edges, which the committee can clearly debate. On the other hand, our framework automatically optimizes, via learning, the potentials that are used as the means for enabling the algorithm to make good future-aware failure-aware decisions. The committee does not need to debate these potentials, whose impact on performance is hard for a human to understand.

3.2.1 *Setting* MaxCard *and* MaxCard-Fair

We begin by addressing the MaxCard and MaxCard-Fair objectives. The MaxCard-Fair objective can viewed

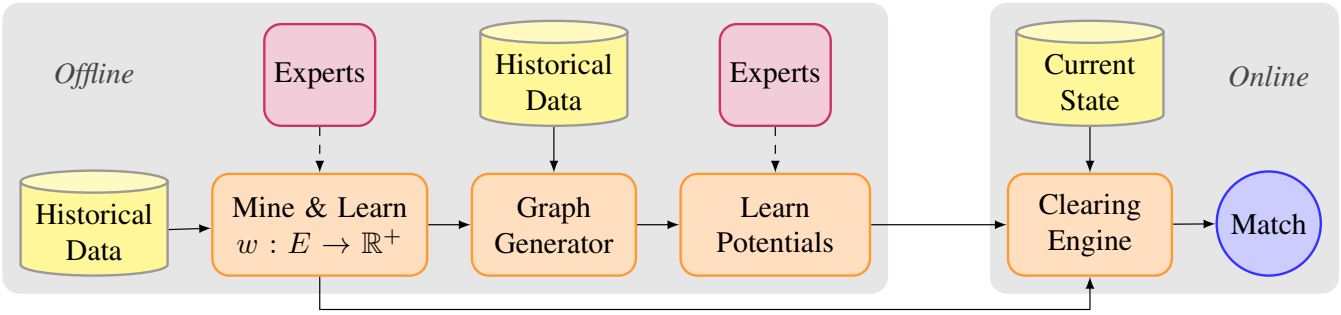

Figure 2: The FutureMatch framework.

as a generalized form of MaxCard (that is, MaxCard is just MaxCard-Fair with an empty set of vertices who are preferred by the objective).

Deciding which class of vertices are preferred is a complex ethical and medical decision. We use two common preference criteria in this paper: pediatric status and sensitization. Children (in the US, those who are under age 18) are typically treated preferentially in medical systems; we follow that rule here. For kidney exchange it has explicitly been articulated that such pediatric patients should be preferred not only because they have a lot of life left (barring their kidney disease) but also because having poor kidney function stunts growth. Some patients are *highly sensitized*, which means they are extremely unlikely to be medically compatible with a random organ. For these patients, finding a kidney is difficult [30]. The percentage of highly-sensitized patients in fielded kidney exchanges is high; over 60% of the patients in the UNOS kidney exchange are highly sensitized.

Figure 3 shows the evolution of the UNOS nationwide pool since inception, with each bar representing the pool at the time of a match run. Red bars show the portion of the pool that is marginalized (that is, either the patient is under the age of 18 or highly sensitized). UNOS currently prioritizes such patients significantly in its matching algorithm.

While defining fairness is a contentious issue in social science, a recent paper by Dickerson, Proccacia, and Sandholm [12] formalizes two natural "fair" utility functions and shows how to optimize either of these functions in the deterministic or failure-aware *static* models of kidney exchange. We adapt the *weighted* fairness rule from that paper to FutureMatch. The weighted fairness rule adjusts edge weights by some re-weighting function $\Delta : E \to \mathbb{R}^+$. A simple example re-weighting function is multiplicative:

$$\Delta^{\beta}(e) = \begin{cases} (1+\beta)w_e & \text{if } e \text{ ends in } V_M \\ w_e & \text{otherwise} \end{cases}$$

Here, $V_M \subseteq V$ is the set of vertices with marginalized patients. Intuitively, for some $\beta > 0$, this function scales the weight of edges ending in marginalized vertices by $(1+\beta)$. For example, if $\beta = 1.5$, then the optimization algorithm will value edges that result in a marginalized patient receiving a transplant at 250% of their initial weight (possibly scaled by factors like edge failure probability or chain position, as we discuss in the Section 3.4).

For any $M \in \mathcal{M}$, let M' be the matching such that every edge $e \in E$ has augmented weight $\Delta^{\beta}(e)$. Then the MaxCard-Fair utility function u_{Δ} is defined in terms of the utilitarian MaxCard utility function u applied to the augmented matching M', such that $u_{\Delta}(M) = u(M')$. In

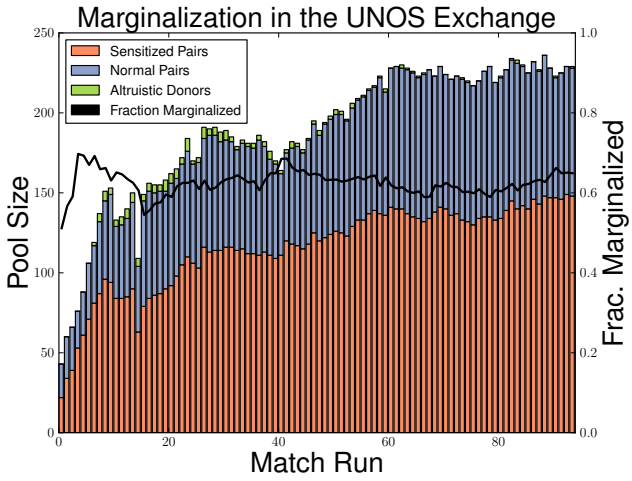

Figure 3: Evolution of the UNOS national kidney exchange. For each of 94 match runs (x-axis), the number of highly-sensitized or underage patients, non-highly-sensitized patients, and altruists are plotted (left y-axis), as well as the percentage of patients who are highly sensitized or underage as a percentage of the pool size (right y-axis).

Section 4, we vary the parameter β to empirically quantify its effects on each of the three objective functions.

3.2.2 *Optimizing for* MaxLife *via learning to predict graft survival from data*

With the MaxLife objective we are interested in maximizing how long the transplanted kidneys, in aggregate, survive in the patients.[1] To do so, we must first determine an empirically sound estimate of the lifespan of a transplant as a function of donor and recipient attributes.

Delen, Walker, and Kadam [9] compare a variety of techniques for predicting breast cancer survivability; unlike their study, we are interested in predicting the survival *length* of a kidney graft, as opposed to whether or not a patient survives treatment at all. Data mining models are also actively being developed to predict the risk of readmission for congestive heart failure patients [20]. Most related to our work is the

[1] Another objective would be to maximize aggregate increase in life duration. This would involve subtracting out the expected life duration without a transplant from the expected life duration with the transplant, and could incorporate the possibility of additional transplants after graft failure.

Kidney Donor Profile Index (KDPI), which is currently under development by UNOS for use in the deceased donor allocation process [18]. The KDPI score of a deceased donor kidney measures the estimated quality of the donor organ being allocating to the *average* recipient. In contrast, our predictor, which we will describe next, provides a unique quality score not just based on donor attributes but also based on attributes of the specific potential recipient.

We look at all 75,264 *living donor* transplant events in the US between October 1, 1987 and June 30, 2013. This data includes medical characteristics of the recipient and donor at the time of transplantation, as well as follow-up data regarding the health of the recipient and the recipient's new kidney; this follow-up data is updated at least annually.

Conditioned on a kidney graft being marked as failed in our dataset, the average graft lifetime is about 1912.7 days, or slightly over 5 years. However, due in large part to the marked increase in kidney failure since the late 1980s, nearly 75% of grafts in the dataset are not marked as failed. This occurs because either (i) the recipient is still alive with a functioning donated kidney or (ii) the recipient has died, but for a non-kidney-failure-related reason. Thus, we use *survival analysis* to estimate the lasting power of a graft, as we now proceed to explain.

Features of both the recipient and donor have a large effect on graft survival. For example, tissue type (HLA) testing measures the closeness of match between antigens in the cells of a donor and patient. Figure 4 gives a Kaplan-Meier estimator of the survival functions of (i) kidney transplants resulting from a donor and recipient being a perfect HLA match and (ii) those resulting from imperfect HLA matchings. Clearly, a kidney that is a perfect tissue type match is more desirable than an imperfectly matched one; indeed, the model estimates a median survival time of 5,808 days for a perfect match compared to 4,300 for an imperfect match. A log-rank test revealed that the difference between the two distributions was significant ($p \ll 0.0001$).

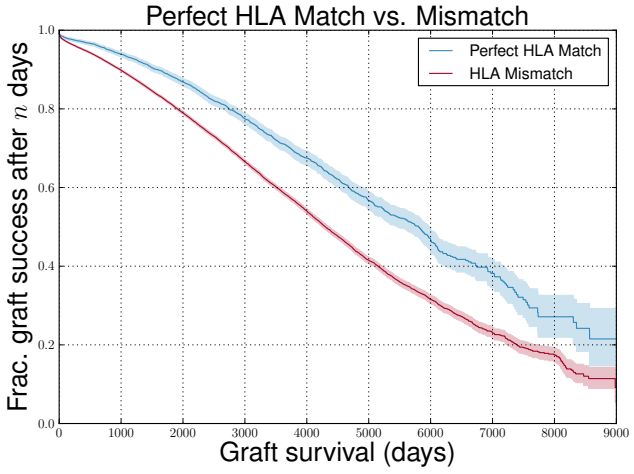

Figure 4: Kaplan-Meier estimator of the survival function for kidney transplants whose donors have zero HLA mismatches versus those with at least one HLA mismatch, with 95% confidence intervals.

In our experiments, we use a Cox proportional hazards regression analysis to explore the effect of multiple features on survivability. At a high level, this method regresses the survival time of the graft against explanatory features of the donor and recipient. More specifically, define the *hazard* H at time t days after a transplant as follows:

$$H(t) = H_0(t) \times \exp(b_1 X_1 + b_2 X_2 + \ldots + b_k X_k) \quad (1)$$

Here, each X_i is a predictor variable corresponding to a single feature of the donor or recipient, and $H_0(t)$ is a baseline hazard rate at time t for a recipient with $X_i = 0$ for $i \in \{1, \ldots, k\}$. Then $H(t)$ represents the instantaneous risk of graft failure at time t. We want to learn this function.

To begin, we include the following features: recipient age, difference in donor and recipient's age, donor HLA profile, recipient HLA profile, donor and recipient blood type compatibility. The HLA profile of a donor or recipient is separated into three integral features—HLA-A, HLA-B, and HLA-DR—that can take values in $\{0, 1, 2\}$, representing 0, 1, or 2 mismatches. By separating the general HLA mismatch feature into three separate mismatch features, we complicate (but increase the power of) the model [23]; this separation is motivated by evidence that mismatches at the HLA-A, -B, and -DR level have varyingly negative impact on survival.

We ran a Cox proportional hazards regression on this unpruned feature space. This used 74,244 live donor transplantations during which there were 18,714 graft failures; 920 live donation events were dropped due to one or more missing features. Our initial regression showed that increases in the HLA-B mismatch feature did not have a significant effect on the dependent variable ($p = 0.22$). Prior research from the mid-1980s on *cadaveric* donation found a significant relationship between the combined feature of HLA-B and HLA-DR mismatches on graft outcome; we find that this does not hold on living donor data in the present [23]. After selecting for significant variables in this initial run—that is, all of the attributes previously discussed *except* HLA-B mismatch—we re-ran a Cox proportional hazard regression. Results are reported in Table 1.

feature	$\exp(b_i)$	$\mathrm{SE}(b_i)$	z	p
recipient age	1.0075315	0.00077	9.715	$\mathbf{< 2 \times 10^{-16}}$
age difference	1.0052498	0.00067	7.766	$\mathbf{8.10 \times 10^{-15}}$
HLA-A	1.0527306	0.01195	4.297	$\mathbf{1.73 \times 10^{-5}}$
HLA-DR	1.0868021	0.01191	6.984	$\mathbf{2.86 \times 10^{-12}}$
ABO incomp.	1.3787143	0.07476	4.295	$\mathbf{1.74 \times 10^{-5}}$

Table 1: Learned weights via Cox regression after feature pruning for statistical significance.

Table 1 shows that each of our pruned features has a statistically significant effect on graft survival. To interpret the results, as an example first consider the HLA-DR feature. We see that $\exp(b_{\text{HLA-DR}}) \approx 1.087$; recalling Equation 1, a unit increase in the HLA-DR mismatch feature will result in a factor of 1.087 increase over the baseline hazard rate. Varying either recipient age or the difference in donor and recipient age was also statistically significant, with a unit increase in recipient age having a larger effect on the hazard rate. As might be expected, blood type compatibility plays the largest role in hazard rate, where an incompatible (and thus heavily immunosuppressed) transplant has a factor 1.379 increase over the base hazard rate.

Using this data, we can estimate $S_e(t)$, the survival probability at time t for a potential transplant $e \in E$ between a

recipient and donor with features x_i^e, as follows:

$$S_e(t) = \exp\left(-H_0(t) \times \sum_i x_i^e b_i\right) \quad (2)$$

Building on Equation 2, we can define a weight function $w : E \to \mathbb{R}^+$ as $w(e) \propto \exp\left(-\sum_i x_i^e b_i\right)$. Intuitively, the weight function w assigns higher relative weight to those edges with lower relative risk.

3.3 Learning Element Potentials

The weight function w learned in Section 3.2 quantifies how useful an edge is *in the present*. To complement this, *potentials* quantify how useful an element of the graph (an edge, 2-cycle, etc.) is expected to be in the future. This idea was initially proposed by Dickerson, Procaccia, and Sandholm [10]; we build on their work but use a better learning algorithm for determining the potentials, and a significantly more realistic distribution of training and test instances.

First, we select a set Θ of features representing different element types in the pool. Then, for each element type θ, assign some value $P_\theta \in \mathbb{R}$ that represents the expected potential usefulness of that kind of element to the pool over time. As an overly simplified pedagogical example, let $\Theta_{\text{ALT}} = \{\text{ALT}, \text{PAIR}\}$; potentials are assigned based on whether or not a vertex is an altruist or a patient-donor pair. Intuitively, $P_{\text{ALT}} \geq P_{\text{PAIR}}$; altruistic donors tend to be (much) easier to match because no returning edge is required to "close the cycle" at the end of a chain. These potentials are then subtracted out for each element in the objective.

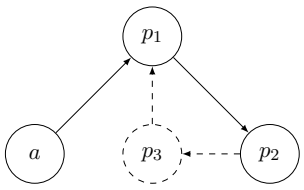

Figure 5: Example of potentials Θ_{ALT}. Pair p_3 appears in the second period. Myopic matching uses a to match two pairs; assigning a positive potential results in all three pairs matched without using a.

Figure 5 shows a two time period example under Θ_{ALT}. Vertices a, p_1, and p_2 arrive in the first time period, while vertex p_3 arrives in the second time period. Assigning a (large enough) potential P_{ALT} results in the chains $\langle a, p_1\rangle$ and $\langle a, p_1, p_2\rangle$ having negative weight and thus not being matched in the first time period. However, when p_3 arrives, the chain $\langle a, p_1, p_2, p_3\rangle$ potentially has positive value and can be matched, or the 3-cycle $\langle p_1, p_2, p_3\rangle$ has higher positive value and is matched instead, saving the altruistic donor a for a longer chain in the future.

In this paper, we use a richer feature space consisting of the *ABO blood types* of patients and donors. The blood type of a donor kidney can result in rejection in a potential patient. 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 reasons for incompatibility also exist (e.g., due to sensitization as discussed in Section 3.2), a type O kidney can be transplanted into any patient; type A and B kidneys can be transplanted into A and B patients, respectively, or an AB candidate; and type AB kidneys are limited to only type AB patients. This imposes a natural partial ordering on blood types; for example, it seems that an O-donor is somehow more "valuable" than an A-donor because she has no blood type restriction. Our automatically learned potentials agree with this, as we will now discuss.

Encode each of the $4 \times 4 = 16$ combinations of patient and donor blood types, as well as the 4 possible blood types of altruists, by $\Theta_{\text{ABO}} = \{\text{O-O}, \text{O-A}, \ldots, \text{AB-B}, \text{AB-AB}\} \cup \{\text{O}, \ldots, \text{AB}\}$. Then we want to learn $P_{X\text{-}Y} \in \mathcal{P}_{\text{ABO}}$, for each $X, Y \in \{\text{O}, \text{A}, \text{B}, \text{A-B}\}$. We combine the learned potentials $\mathcal{P}_{\text{ABO}}$ with the learned weight function w from Section 3.2 using a function $f_w : E \to \mathbb{R}$. For our experiments, we use $f_w(e) = w(e) \cdot (1 - P_X - P_Y)$, with X the donor blood type at e's source and Y the patient blood type at e's sink.

We use SMAC [15], a state-of-the-art model-based algorithm configuration tool that intelligently searches through a parameter space to optimize some objective function. SMAC guides its navigation in the space of parameter vectors by constructing a model that predicts algorithm performance as a function of the parameters. It uses the model to select promising parameter vectors and tests them against the incumbent parameter vector, possibly continuing that search.

In our setting, the parameter vector is the vector of potentials. At each parameter vector that SMAC navigates to, we run a large number of trials of our simulator of a kidney exchange (discussed in the next subsection) to see how the batch matching algorithm would perform in a dynamic setting using that parameter vector. That performance number is then fed back to SMAC, and SMAC navigates to the next parameter vector to continue the search.

3.4 Dynamic Simulator and Verification

When learning potentials offline, it is important to mimic closely the behavior of the fielded exchange online. If the distribution of incoming potential types is significantly different than expected, so too will be the estimates of potentials.

We built a dynamic simulator of kidney exchange using data from the UNOS exchange (and APD [3]). This work significantly extends that of Dickerson, Procaccia, and Sandholm [11], which defined and experimentally evaluated a model of the evolution of dynamic kidney exchange. Critically, they did not perform dynamic optimization in that model—just myopic optimization applied sequentially in a dynamic model. They also sampled from a basic generator that is no longer accepted in the kidney exchange community [28], while we sample from an accurate distribution—the historical UNOS exchange pool!

Figure 6 portrays two time steps in our dynamic kidney exchange simulator. New pairs and altruistic donors enter the pool at each time period, while some leave the pool due to a variety of non-exchange-related reasons (becoming too ill to transplant, death, having a donor back out, finding a donor elsewhere). A matching is performed at each time period, which results in a set of matched pairs leaving the pool for $t > 0$ time periods. This reflects the length of time required to medically and logistically verify the implementability of the planned match. Matched patients then either leave the pool permanently after successfully receiving a kidney, or return to the pool after failing to receive a kidney. We set the relevant entrance and exit probabilities based on the real UNOS kidney exchange data [17].

A matching is determined at each time period based on either a *deterministic* or *failure-aware* clearing algorithm, which we briefly describe here. Both models compute an

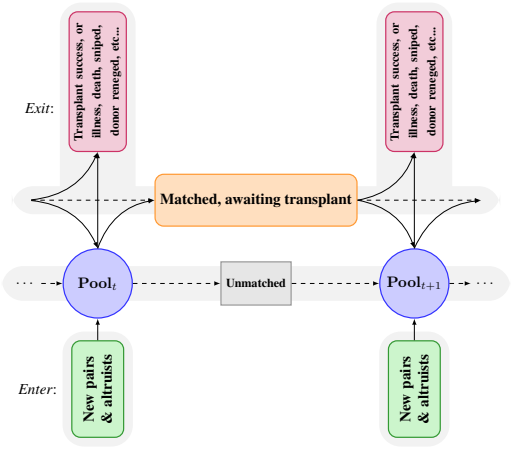

Figure 6: Dynamic kidney exchange.

optimal matching $M^* = \arg\max_{M \in \mathcal{M}} u(M)$, where $u(M) = \sum_{c \in M} u(c)$. Here, $u(c)$ represents the utility of a cycle or chain c. In the deterministic model, $u(c) = \sum_{e \in c} f_w(e)$: that is, the sum of the weights of the constituent edges in a cycle or chain subject to the weight function w and potential mapping f learned in Sections 3.2 and Section 3.3.

The deterministic model is susceptible to edge failures. For example, if a single edge in a 3-cycle fails, that entire cycle fails to execute. Similarly, if the third edge in a long chain fails, then the tail of that chain (after the failed edge) is cut off. Let $q_e \in [0, 1]$ be the probability that an edge succeeds. Then the failure-aware model defines the *discounted* utility for a cycle c as $u(c) = \left[\sum_{e \in c} f_w(e)\right] \cdot \left[\prod_{e \in c} q_e\right]$, and the discounted utility for a chain $c = \langle e_0, e_1, \ldots, e_{k-1} \rangle$ as

$$u(c) = \left[\sum_{i=1}^{k-1}(1 - q_i)\sum_{j=0}^{i-1} f_w(e_j)\prod_{j=0}^{i-1} q_j\right] + \left[\sum_{i=0}^{k-1} f_w(e_i)\prod_{i=0}^{k-1} q_i\right]$$

Next, we explore the effects of optimization under both models using each of the three objective functions defined earlier.

4. EXPERIMENTAL VALIDATION

In this section, we experimentally validate FutureMatch on data from the UNOS nationwide kidney exchange. We explore the effect each of the three objectives—MaxCard, MaxCard-Fair, and MaxLife—has on a variety of metrics under FutureMatch and under myopic deterministic matching, which is the fielded state of the art.

4.1 Method

Compatibility graphs are sampled with replacement from the set of all altruistic donors and patient-donor pairs who have ever participated (either successfully or unsuccessfully) in the UNOS exchange between Oct. 2010 and Jan. 2014, over a total of 94 match runs. Edges are drawn between two vertices in the graph if they pass the UNOS feasibility test, which determines compatibility with respect to patient and donor blood type, which kidney (left or right) is available, Hepatitis B/cytomegalovirus (CMV)/Epstein-Barr virus (EBV) positivity, creatinine clearance, blood pressure limits for the patient and donor, Body Mass Index (BMI) preferences, minimum and maximum age requirements for the donor, whether the donor and patient are willing to travel or accept a shipped organ, and other per-patient and per-donor requirements. The patient and donor features are used during the learning of the weight function for MaxCard-Fair (to determine marginalization status) and MaxLife (to determine relative risk), and during the learning of potentials $\mathcal{P}_{\text{ABO}}$ (to determine blood types). Which exact features are used, and how, was discussed in the respective subsections earlier in the paper.

We draw edge failure probabilities in accordance with those published in the medical literature [3]. Critically, according to that APD experience, sensitization plays a large role in the probability of a match failing, with higher sensitization correlating with higher failure probability. These failure probabilities are incorporated directly into the optimization under the failure-aware model, and are incorporated directly into the *simulation* under *both* models (deterministic and failure-aware). A matched patient-donor pair leaves the pool for $t = 8$ weeks before receiving a confirmation of transplant success (and thus leaving the pool) or match failure (and thus reentering the pool, if the pair did not leave for other reasons like death or finding a donor elsewhere). Matching is performed once per week as is the current practice in the UNOS exchange; the total simulation occurs over $T = 24$ weeks. Each data point is based on at least 140 separate runs on separate generated realistic graphs.

4.2 Results

We compare FutureMatch against a baseline of myopic deterministic matching under each of the objectives. Conservatively, statistical significance was determined using the Wilcoxon signed-rank test, which is a nonparametric alternative to the paired t-test. Table 2 shows the median expected gain in the overall number of transplants from using FutureMatch under each of the objectives. Each column labeled $|V| = k$ corresponds to a simulation over k patient-donor pairs and altruists sampled as described in Section 4.1.

Table 2 shows that the two objectives that do not regard fairness—MaxCard and MaxLife—significantly beat myopic deterministic matching under the same objective. Interestingly, so too does MaxCard-Fair for low values of β. As β increases, the gain in *overall* number of transplants decreases (although it never drops below the deterministic matching algorithm with significance). This decrease in overall gain is incurred because marginalized patients, who (i) generally have lower in-degree, and (ii) have a higher probability of match failure, are being weighted more than easier to match pairs.

Table 3 explores this tradeoff between fairness and efficiency explicitly. For the fairness-agnostic and lightly fairness preferring objectives, a relative loss of a few marginalized transplants is realized—although this loss of marginalized transplants is always less (typically much less) than the overall gain in transplants. Increasing the optimizer's preference for marginalized patients results in statistically significant gains in the number of marginalized transplants at *no* statistically significant loss in the overall expected number of transplants. In fact, for a middle ground around $\beta = 2$, FutureMatch often shows statistically significant gains in *both* overall transplant and marginalized transplant counts—a clear win over myopia.

Figure 7 shows the increase in number of marginalized transplants fielded as the β fairness factor is increased. It also shows the relative gain in the MaxLife objective compared to myopic matching. From Table 2, MaxLife yielded

	$\|V\| = 300$		$\|V\| = 400$		$\|V\| = 500$		$\|V\| = 600$		$\|V\| = 700$		$\|V\| = 800$		$\|V\| = 900$	
	Gain	*p*	*Gain*	*p*	*Gain*	*p*	*Gain*	*p*	*Gain*	*p*	*Gain*	*p*	*Gain*	*p*
MaxCard	+2	✓	+4	✓	+5	✓	+6	✓	+10	✓	+11	✓	+13	✓
MaxCard-Fair, $\beta = 1$	+1	✓	+4	✓	+6	✓	+8	✓	+9	✓	+11	✓	+12	✓
MaxCard-Fair, $\beta = 2$	+1		+2	✓	+3	✓	+3	✓	+5	✓	+6	✓	+10	✓
MaxCard-Fair, $\beta = 3$	+1		+0		+3	✓	+1		+1	✓	+3	✓	+2	
MaxCard-Fair, $\beta = 4$	-1		+1		+1		+1		+3	✓	+3		+2	
MaxCard-Fair, $\beta = 5$	+0		+0		+1		+1		+1		+2		+3	
MaxLife	+2	✓	+3	✓	+6	✓	+8	✓	+7	✓	+11	✓	+9	✓

Table 2: Median gains in expected total number of transplants under FutureMatch. A ✓ represents statistical significance (Wilcoxon signed-rank test, $p \ll 0.01$).

	$\|V\| = 300$		$\|V\| = 400$		$\|V\| = 500$		$\|V\| = 600$		$\|V\| = 700$		$\|V\| = 800$		$\|V\| = 900$	
	Gain	*p*	*Gain*	*p*	*Gain*	*p*	*Gain*	*p*	*Gain*	*p*	*Gain*	*p*	*Gain*	*p*
MaxCard	-2	✗	-2	✗	-3	✗	-4	✗	-6	✗	-7	✗	-9	✗
MaxCard-Fair, $\beta = 1$	-1	✗	-1	✗	-1	✗	-2	✗	-3	✗	-3	✗	-5	✗
MaxCard-Fair, $\beta = 2$	+0		+0		+1	✓	+1	✓	+2	✓	+1		+1	
MaxCard-Fair, $\beta = 3$	+1	✓	+1	✓	+3	✓	+3	✓	+3	✓	+5	✓	+4	✓
MaxCard-Fair, $\beta = 4$	+1	✓	+2	✓	+3	✓	+4	✓	+4	✓	+5	✓	+5	✓
MaxCard-Fair, $\beta = 5$	+1	✓	+2	✓	+3	✓	+4	✓	+5	✓	+7	✓	+5	✓
MaxLife	-1	✗	-3	✗	-3	✗	-5	✗	-6	✗	-6	✗	-9	✗

Table 3: Median gains in expected total number of marginalized transplants under FutureMatch. A ✓ or ✗ represents statistically significant gains or losses, respectively (Wilcoxon signed-rank test, $p \ll 0.01$).

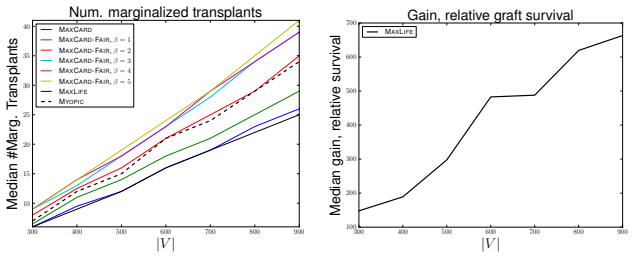

Figure 7: Median expected total number of marginalized transplants under FutureMatch (left), and median expected gain in MaxLife (right).

roughly the same number of transplants, just weighted slightly toward those with longer expected graft survival. This correlation aligns with the general notion that easier-to-match patients also tend to be healthier overall, and thus similar vertices will tend to be favored (or disfavored) by both the MaxCard and MaxLife objectives.

Our experiments support the following conclusions:

- FutureMatch under MaxCard and MaxCard-Fair with low $\beta = 1$ results in a significant increase in the *overall* number of transplants compared to myopic, at the cost of a smaller decrease in the number of *marginalized* transplants.
- FutureMatch under MaxCard-Fair with high β results in a significant increase in *marginalized* transplants, at *no cost* to the overall number of transplants under myopic matching.
- For a middle ground around $\beta = 2$, FutureMatch can result in both more overall expected transplants and more marginalized transplants.

We note that we are *not* making policy recommendations; rather, we are giving a proof of concept that the FutureMatch framework can effectively balance conflicting wants in an exchange. Indeed, the exact fairness quantification β that most effectively balances efficiency and fairness is a function of the underlying graph dynamics, which vertices are considered marginalized, and the ethical and legal wants of an exchange. All of these dimensions can be effectively encoded, validated, and compared through FutureMatch.

5. CONCLUSIONS AND FUTURE WORK

We presented FutureMatch, a framework for learning to do complex matching in a general dynamic model. The framework uses data mining in several of its modules. Motivated by our experience running the computational side of the US nationwide kidney exchange—which currently includes 56% of the US transplant centers—we showed how to instantiate FutureMatch to mimic the nationwide exchange under three different matching objectives and under two models of kidney exchange. We validated FutureMatch on real data drawn from 94 match runs of the US nationwide exchange between Oct. 2010 and Jan. 2014, and found that dynamic matching results in statistically significant increases in each of these objectives. Perhaps most critically, we showed that the framework yields better efficiency *and* better fairness than deterministic myopic matching algorithms (which are the status quo class of algorithm in kidney exchange practice).

The FutureMatch framework could be improved in a number of ways. While it was built explicitly to support input from domain experts, better automated methods for feature selection and classifier selection could be employed at both the weight function and potential learning stages; Auto-WEKA [29] could be promising in this direction. Similarly, better estimates of each parameter in our kidney exchange simulator—especially the probability of an edge fail-

ing after algorithmic match but before transplantation—would aid in more accurate simulation and thus ultimately in more/better transplants. The sort of data required to *accurately* estimate this probability is certainly "big", and would include not just rich medical histories for all patients and donors, but also exogenous factors that influence match failure like the histories of transplant centers and the dynamics of exchanges with the deceased-donor transplant wait list.

6. ACKNOWLEDGMENTS

This work was funded by an NDSEG fellowship awarded through the Army Research Office, NSF grants IS-1320620, CCF-1101668, and IIS-0964579, and used the Pittsburgh Supercomputing Center in partnership with XSEDE, which is supported by NSF grant OCI-1053575. The authors acknowledge Intel Corporation and IBM for gifts, and thank Christos Faloutsos, Katrina LaCurts, Evangelos Papalexakis, and Keith Winstein for comments.

This work was supported in part by Health Resources and Services Administration contract 234-2005-370011C. The content is the responsibility of the authors alone and does not necessarily reflect the views or policies of the Department of Health and Human Services. The authors also thank Ruthanne Leishman, Elizabeth Sleeman, Darren Stewart, and the rest of the UNOS KPD Pilot Program staff.

7. REFERENCES

[1] D. Abraham, A. Blum, and T. Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *EC-07*.

[2] E. Anshelevich, M. Chhabra, S. Das, and M. Gerrior. On the social welfare of mechanisms for repeated batch matching. In *AAAI-13*.

[3] I. Ashlagi, D. Gilchrist, A. Roth, and M. Rees. Nonsimultaneous chains and dominos in kidney-paired donation—revisited. *Am J Transplant*, 11(5):984–994, 2011.

[4] I. Ashlagi, P. Jaillet, and V. H. Manshadi. Kidney exchange in dynamic sparse heterogenous pools. In *EC-13*.

[5] P. Awasthi and T. Sandholm. Online stochastic optimization in the large: Application to kidney exchange. In *IJCAI-09*.

[6] C. Barnhart, E. L. Johnson, G. L. Nemhauser, M. W. P. Savelsbergh, and P. H. Vance. Branch-and-price: Column generation for solving huge integer programs. *Oper Res*, 46(3):316–329, 1998.

[7] A. Blum, A. Gupta, A. D. Procaccia, and A. Sharma. Harnessing the power of two crossmatches. In *EC-13*.

[8] V. Chandola, S. R. Sukumar, and J. C. Schryver. Knowledge discovery from massive healthcare claims data. In *KDD-13*.

[9] D. Delen, G. Walker, and A. Kadam. Predicting breast cancer survivability: A comparison of three data mining methods. *Artif Intell Med*, 34(2):113–127, June 2005.

[10] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Dynamic matching via weighted myopia with application to kidney exchange. In *AAAI-12*.

[11] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Failure-aware kidney exchange. In *EC-13*.

[12] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Price of fairness in kidney exchange. In *AAMAS-14*.

[13] J. Feldman, A. Mehta, V. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *FOCS-09*.

[14] G. Goel and P. Tripathi. Matching with our eyes closed. In *FOCS-2012*.

[15] F. Hutter, H. Hoos, and K. Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In *LION-11*.

[16] R. Karp, U. Vazirani, and V. Vazirani. An optimal algorithm for on-line bipartite matching. In *STOC-90*.

[17] Kidney Paired Donation Work Group. OPTN KPD pilot program cumulative match report (CMR) for KPD match runs: Oct 27, 2010 – Apr 15, 2013, 2013.

[18] Kidney Transplantation Committee. OPTN concepts for kidney allocation, 2011.

[19] V. H. Manshadi, S. O. Gharan, and A. Saberi. Online stochastic matching: Online actions based on offline statistics. *Math Oper Res*, 37(4):559–573, 2012.

[20] N. Meadem, N. Verbiest, K. Zolfaghar, J. Agarwal, S.-C. Chin, and S. B. Roy. Exploring preprocessing techniques for prediction of risk of readmission for congestive heart failure patients. In *Data Mining and Healthcare (DMH), at KDD-13*.

[21] M. Molinaro and R. Ravi. Kidney exchanges and the query-commit problem. Manuscript, 2013.

[22] B. L. Neuen, G. E. Taylor, A. R. Demaio, and V. Perkovic. Global kidney disease. *The Lancet*, 382(9900):1243, 2013.

[23] G. Opelz. Correlation of HLA matching with kidney graft survival in patients with or without cyclosporine treatment: for the collaborative transplant study. *Transplantation*, 40(3):240–242, 1985.

[24] F. T. Rapaport. The case for a living emotionally related international kidney donor exchange registry. *Transplantation Proceedings*, 18:5–9, 1986.

[25] M. Rees, J. Kopke, R. Pelletier, D. Segev, M. Rutter, A. Fabrega, J. Rogers, O. Pankewycz, J. Hiller, A. Roth, T. Sandholm, U. Ünver, and R. Montgomery. A nonsimultaneous, extended, altruistic-donor chain. *New Engl J Med*, 360(11):1096–1101, 2009.

[26] A. Roth, T. Sönmez, and U. Ünver. Kidney exchange. *Q J Econ*, 119(2):457–488, 2004.

[27] A. Roth, T. Sönmez, U. Ünver, F. Delmonico, and S. L. Saidman. Utilizing list exchange and nondirected donation through 'chain' paired kidney donations. *Am J Transplant*, 6:2694–2705, 2006.

[28] S. L. Saidman, A. Roth, T. Sönmez, U. Ünver, and F. Delmonico. Increasing the opportunity of live kidney donation by matching for two and three way exchanges. *Transplantation*, 81(5):773–782, 2006.

[29] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. In *KDD-13*.

[30] United Network for Organ Sharing (UNOS). http://www.unos.org/.

[31] U. Ünver. Dynamic kidney exchange. *Rev Econ Stud*, 77(1):372–414, 2010.

[32] F. Wang, J. Sun, J. Hu, and S. Ebadollahi. iMet: Interactive metric learning in healthcare applications. In *SDM-11*.

[33] K. Zolfaghar, J. Agarwal, D. Sistla, S.-C. Chin, S. Basu Roy, and N. Verbiest. Risk-o-meter: an intelligent clinical risk calculator. In *KDD-13*.