Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

DISTRIBUTIONALLY ROBUST CYCLE AND CHAIN PACKING WITH APPLICATION TO ORGAN EXCHANGE

Duncan C. McElfresh
John P. Dickerson

Department of Applied Mathematics
Department of Computer Science
University of Maryland
125 Paint Branch Drive
College Park, MD 20742, USA

Ke Ren
Hoda Bidkhori

Department of Industrial Engineering
University of Pittsburgh
3700 O'Hara Street
Pittsburgh, PA 15261, USA

ABSTRACT

We consider the cycle packing problems motivated by kidney exchange. In kidney exchange, patients with willing but incompatible donors enter into an organized market and trade donors in cyclic structures. Exchange programs attempt to match patients and donors utilizing the quality of matches. Current methods use a point estimate for the utility of a potential match that is drawn from an unknown distribution over possible true qualities. We apply the conditional value-at-risk paradigm to the size-constrained cycle and chain packing problem. We derive sample average approximation and distributionally-robust-optimization approaches to maximize the true quality of matched organs in the face of uncertainty over the quality of potential matches. We test our approach on the realistic kidney exchange data and show they outperform the state-of-the-art approaches. In the experiments, we use randomly generated exchange graphs resembling the structure of real exchanges, using anonymized data from the United Network for Organ Sharing.

1 INTRODUCTION

Cycle packing is a fundamental problem in algorithmic graph theory, with classical results dating to at least 1959 (ErdHos and Gallai 1959; Lovász 1968). Many variants of the cycle packing problem exist—on directed and undirected graphs, with and without edge weights, and with and without minimum or maximum cycle lengths—and many variants are NP- and even APX-hard (Pyber 1996). Cycle packing problems arise in many real-world settings, including scheduling, transportation, healthcare, and finance. In these settings, models parameters—vertices, edges, weights—are uncertain but can be estimated from data, at some cost. The interplay between this estimation and the final combinatorial optimization provides a rich, and important role for modern data-driven decision-making.

In this paper, we are motivated by the application of cycle packing to a specific problem with societal impact at the intersection of machine learning (ML), healthcare, and economics: *kidney exchange* (Rapaport 1986; Roth, Sönmez, and Ünver 2004). A kidney exchange is an organized barter marketplace where patients with end-stage renal disease trade willing donors in cyclic or chain-like transactions. Exchanges are facilitated by a centralized clearinghouse, which aims to find the "best" disjoint set of such swaps—i.e., to solve a cycle and chain packing problem. While relatively new, kidney exchange already accounts for over 12% of living kidney donations in the US, and is growing worldwide (Biró, Burnapp, Haase, Hemke, Johnson, van de Klundert, and Manlove 2017)—including via extensions to liver and lung (Ergin, Sönmez, and Unver 2017), and even multi-organ (Dickerson and Sandholm 2017), exchange. Furthermore, the AI/ML and operations research communities play a central role in fielding these exchanges (Abraham, Blum, and

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

Sandholm 2007; Manlove and O'Malley 2015; Dickerson, Manlove, Plaut, Sandholm, and Trimble 2016; Li, Lieberman, Macke, Carrillo, Ho, Wellen, and Das 2019).

Potential transplants vary in quality. A donor's kidney quality and the medical similarity to the patient both impact expected graft survival (Massie, Leanza, Fahmy, Chow, Desai, Luo, King, Bowring, and Segev 2016). Similarly, policymakers express value judgments that quantitatively prefer traits (e.g., pediatric patients) to varying degrees (Freedman, Borg, Sinnott-Armstrong, Dickerson, and Conitzer 2020). Neither the true medical quality of a transplant nor the "true" reflection of policymakers' value judgments is known. In practice, instead of distributional information, simple point estimates are used (UNOS 2015). This leads to brittle solutions that may be far from optimal when true quality is realized post-transplant.

Our contributions. We consider edge weight uncertainty over the *quality* of the individual match (henceforth, *weight uncertainty*) in the general cycle packing setting. We integrate risk assessment into the cardinality-constrained cycle packing model. The proposed mean-risk model considers both the *expected* edge weight value of the solution—the focus of classical models (Chen, Tanahashi, and Wang 2008) as well as state-of-the-art kidney-exchange-specific models (Dickerson, Manlove, Plaut, Sandholm, and Trimble 2016)—and the expected lowest $\alpha\%$ total weight. Our model reduces the number of cases with extremely low realized edge weights while improving average performance over standard baselines.

We propose two data-driven methods for solving the kidney exchange problem under the mean-risk model. The first approach is based on the sample average approximation (SAA), which directly uses an empirical distribution. The second approach considers the mismatch between the empirical distribution and the true distribution by drawing on the distributionally robust optimization (DRO) literature (Esfahani and Kuhn 2018). The DRO approach is especially useful for realistic scenarios in which the sample data are not representative of the true underlying distribution. DRO-based methods aim to find an *optimal robust* solution, which achieves the highest matching score under all possible deviations between the empirical distribution and the true distribution. For both approaches, we propose tractable mixed-integer linear/conic programs that can be efficiently solved by standard solvers.

In the experiments, we show on realistic data that both models improve worst-case performance when compared to standard baselines. For these experiments we use 32 randomly-generated exchange graphs resembling the structure of real exchanges, using anonymized data from the United Network for Organ Sharing (UNOS) US-wide kidney exchange and to reflect edge weight uncertainty, we define an edge weight distribution inspired by real sources of uncertainty, described in the next section.

2 PRELIMINARY

We begin by briefly overviewing the standard model for kidney exchange; we then give a brief background on risk assessment, sample average approximation, and distributionally robust optimization.

2.1 Kidney Exchange

Kidney exchanges are often modeled as a directed graph $G = (V, E)$, where vertices $v \in V$ represent both patient-donor pairs P and non-directed donors N' who enter the exchange without a paired patient (i.e., $V = P \cup N'$) (Roth, Sönmez, and Ünver 2004; Roth, Sönmez, et al. 2005; Abraham, Blum, and Sandholm 2007). Directed edges $e \in E$ represent potential transplants from the donor of one vertex to the patient of another. Each edge $e \in E$ has a *weight* w_e which is set by the exchange (often by written policy); weights typically reflect some combination of the medical and moral utility of a transplant. Matches take place in two types of structures: cycles and chains. A *cycle* in a kidney exchange represents a potential cyclic trade between patient-donor pairs in P, where the donor at one pair gives her kidney to the patient at the subsequent pair. A *chain* in a kidney exchange begins when a non-directed donor in N' donates his kidney to the patient of a pair in P, whose paired donor donates her kidney to another patient-donor pair, and so on until a logistical problem arises and the chain concludes.

The kidney exchange *clearing problem* is to select a *feasible* set of cycles and chains in G that maximize overall edge weight. Cycles and chains must be *vertex disjoint* because no donor can give more than one kidney. Additionally, cycles are typically limited in length to 3 or 4 due to logistical constraints (UNOS 2015). We denote by $\mathscr{M}$ the set of all feasible *matchings* (i.e., cycle and chain packings) in G. The most general formulation of the clearing problem is $\text{argmax}_{\mathbf{x}\in\mathscr{M}}\, \mathbf{x}\cdot\mathbf{w}$, where binary decision variables $\mathbf{x}$ represent edges, or cycles and chains. This problem is both NP- and APX-hard (Abraham, Blum, and Sandholm 2007; Biró, Manlove, and Rizzi 2009), and is equivalent to the traditional (weighted) cycle packing problem where cycles may have some length constraint. Note that we can treat chains as cycles with a zero-weight edge connecting the final patient-donor pair in the chain to the initiating non-directed donor.

Uncertainty in kidney exchange. There are several sources of edge weight uncertainty in kidney exchange. We now discuss one primary cause of uncertainty. It is common to choose edge weights that are reflective of the *survival time* of the transplanted organ (Massie, Leanza, Fahmy, Chow, Desai, Luo, King, Bowring, and Segev 2016; UNOS 2015). This is the approach taken by (Li, Lieberman, Macke, Carrillo, Ho, Wellen, and Das 2019), who use the Living Donor Kidney Profile Index (LKDPI)—a score based on the estimated survival time of a transplanted kidney in the donor. These scores are estimated from past transplants (e.g., using Cox survival analysis), and are thus inherently uncertain. For a real LKDPI calculator, we refer readers to (LKDPI-Calculator 2015). In line with the assumptions of the LDKPI, we assume that each edge weight is drawn from an exponential distribution that depends on patient and donor characteristics. We use this distributional assumption to define an edge weight distribution for our experiments in Section 4. Past work has addressed a complementary form of uncertainty over the *existence* of an edge, rather than the quality of the edge. Approaches to that problem variously use stochastic programming (Dickerson, Procaccia, and Sandholm 2019; Anderson, Ashlagi, Gamarnik, and Roth 2015) as well as a (non-distributionally-)robust optimization (Carvalho, Klimentova, Glorie, Viana, and Constantino 2020; McElfresh, Bidkhori, and Dickerson 2019) and, most recently, mixed-integer optimization (Zheng, Shen, and Shi 2015; Bidkhori, Dickerson, McElfresh, and Ren 2020).

2.2 Risk Measures & Assessment

In this paper, we incorporate Conditional Value-at-Risk, also known as expected shortfall, into the cardinality-constrained cycle and chain packing problem, with application to organ exchange. CVaR is one of the most common risk measures with applications in support vector machines (SVM), reinforcement learning, financial management, and many other areas. In addition, optimizing CVaR is often mathematically tractable (Rockafellar, Uryasev, et al. 2000). We propose two complementary approaches to solving our model: one based on the sample average approximation, and the other on distributionally robust optimization.

Stochastic programming via SAA. Our first approach uses sample average approximation (SAA), which was first studied by (Shapiro and Homem-de Mello 1998) and (Kleywegt, Shapiro, and Homem-de Mello 2002). SAA approximates the unknown distribution through empirical distributions. If the data samples are drawn from the *true* unknown distribution, the optimal solution of SAA converges to the true optimal with probability 1; however, if only a small number of samples are available, the performance of the obtained solution is often poor.

Distributionally robust optimization. Our second approach relies on distributionally robust optimization (DRO), a modeling paradigm for optimization under uncertainty. DRO offers computational scalability and promising out-of-sample performance even with limited data, which bridges the gap between robust optimization (Bertsimas and Sim 2004) and stochastic optimization (Shapiro, Dentcheva, and Ruszczyński 2009). This approach begins by defining an ambiguity set $\mathscr{P}$ to contain possible distributions of the uncertain optimization parameters. Consider the example of edge weight uncertainty: we might design an edge weight ambiguity set $\mathscr{P}$ that contains the possible edge weight distributions $\mathbb{P}$ with high probability.

Then, the DRO optimization problem optimizes the solution quality for the *worst-case distribution* inside the ambiguity set $\mathscr{P}$. For kidney exchange (a maximization problem), this corresponds to a *minimization*

over $\mathscr{P}$; for example, suppose that edge weights $\mathbf{w}$ in a kidney exchange graph are random, with distribution contained in ambiguity set $\mathscr{P}$; Problem 1 is the DRO formulation for kidney exchange in this setting:

$$\max_{\mathbf{x}\in\mathscr{X}} \min_{\mathbb{P}\in\mathscr{P}} \quad \mathbf{x}\cdot\mathbf{w}. \tag{1}$$

The ambiguity set $\mathscr{P}$ plays an important role in DRO (Delage and Ye 2010; Bertsimas, Gupta, and Kallus 2018). In this paper, we draw on intuition from the Wasserstein metric when defining our ambiguity set (Esfahani and Kuhn 2018), discussed in greater depth in the next section.

3 MEAN-RISK KIDNEY EXCHANGE MODEL

Our kidney exchange model balances the expected outcome with a worst-case outcome, using a conditional value-at-risk (CVaR) approach, which we describe in detail later in this section. In contrast, most kidney exchange models focus on the *expected* matching weight (e.g., (Dickerson, Procaccia, and Sandholm 2019; Klimentova, Pedroso, and Viana 2016; Anderson, Ashlagi, Gamarnik, and Roth 2015)) or the *worst-case* outcome (e.g., (McElfresh, Bidkhori, and Dickerson 2019; Carvalho, Klimentova, Glorie, Viana, and Constantino 2020)). To our knowledge, the only prior work that considers a CVaR model of kidney exchange are due to Zheng et al. (Zheng, Shen, and Shi 2015), which does not limit the cycle or chain length (and for this reason, they can use a min-cost flow approach), and Bidkhori et al. (Bidkhori, Dickerson, McElfresh, and Ren 2020), which does not consider uncertainty in edge weights.

To account for edge weight uncertainty, we propose the following objective function for kidney exchange:

$$\mu + \gamma \times \mu_\alpha.$$

At a high level, μ represents the expected edge weight, while μ_α represents $\alpha\%$ worst-case mean. Compared to the standard model, the second term is new and measures the variance over different match weights. The positive parameter γ controls the trade-off between the average performance and the risk of the solution. It is predetermined by the users and represents the desirable conservative. The larger the gamma is, the more conservative the optimal solution is.

Notation. We use G to denote a kidney exchange graph, consisting of edges $e \in E$ and vertices $v \in V = P \cup N'$, including patient-donor pairs P and non-directed donors (NDDs) N'. The set of cycles on exchange graph G is denoted as C. We use L to denote the chain cap (maximum number of edges in a chain); w_e and w_c denote the edge and cycle weights for each edge $e \in E$ and cycle $c \in C$ (defined as $w_c = \sum_{e\in c} w_e$), respectively. $\delta^-(i)$ represents the set of edges into vertex i, and $\delta^+(i)$ represents the set of edges out of vertex i. Note that edges between an NDD $n \in N'$ and a patient-donor vertex $d \in P$ may only take position 1 in a chain, while edges between two patient-donor pairs may take any position $2, \cdots, L$ in a chain. For convenience, we define the function $\mathscr{K}$ for each edge: $\mathscr{K}(e)$ is the set of all possible positions edge e may take in a chain.

$$\mathscr{K}(e) = \begin{cases} \{1\}, & e \text{ begins in } n \in N', \\ \{2, \cdots, L\}, & e \text{ begins in } d \in P. \end{cases}$$

Chains always begin with an NDD, followed by one or more patient-donor pairs. Since NDDs only appear at the start of a chain, $\mathscr{K}(e) = \{1\}$ for all edges from NDDs; all other edges can appear at or after the 2nd position in a chain. Our model primarily uses the following decision variables:

- $z_c \in \{0,1\}$: 1 if cycle c is used in the matching, and 0 otherwise.
- $y_{ek} \in \{0,1\}$: 1 if edge e is used at position k in a chain, and 0 otherwise.

Given an instance of the kidney exchange problem $G = (V,E)$, we define $\mathbf{W}$ as an $|E|$-dimensional vector, where

$$W_e = -\sum_{k\in\mathscr{K}(e)} y_{ek} - \sum_{c\in C} \mathbf{1}(e \in c) z_c, \forall e \in E,$$

indicating if edge e is used (-1) or not (0) in a final matching. Recall from §2.2 that $\mathbf{w}$ represents the random edge weights. So, $\langle \mathbf{w}, \mathbf{W} \rangle$ is the "loss" (negative weight) of the matching. While it is common in the literature to represent kidney exchange as a maximization problem (with a non-negative objective), we instead use a minimization problem (with non-positive objective) to be consistent with the convention of risk-minimization. The average of the $\alpha\%$ worst (highest) losses is known as the conditional value-at-risk (CVaR) (Rockafellar, Uryasev, et al. 2000) at level α. Therefore, we formulate the model as

$$\begin{aligned} \min \quad & \mathbb{E}_{\mathbb{P}}(\langle \mathbf{w}, \mathbf{W} \rangle) + \gamma \text{CVaR}_{\alpha}(\langle \mathbf{w}, \mathbf{W} \rangle) \\ s.t. \quad & \{\mathbf{y}, \mathbf{z}\} \in \mathscr{X}, \end{aligned} \tag{2}$$

where $\mathbb{P}$ is the distribution corresponds to edge weights and $\mathscr{X}$ denotes the set of constraints from the PICEF model; these constraints ensure that decision variables $\mathbf{y}$ and $\mathbf{z}$ correspond to a feasible matching,

$$\mathscr{X} \equiv \begin{cases} \sum\limits_{e \in \delta^-(i)} \sum\limits_{k \in \mathscr{K}(e)} y_{ek} + \sum\limits_{c \in C: i \in c} z_c \leqslant 1, i \in P; \\ \sum\limits_{e \in \delta^-(i) \wedge k \in \mathscr{K}(e)} y_{ek} \geqslant \sum\limits_{e \in \delta^+(i)} y_{e,k+1}, i \in P, k \in \{1, \cdots, L-1\}; \\ \sum\limits_{e \in \delta^+(i)} y_{e1} \leqslant 1, i \in N'; \\ y_{ek} \in \{0,1\}, e \in E, k \in \mathscr{K}(e), z_c \in \{0,1\}, c \in C. \end{cases}$$

The objective function minimizes a weighted sum of the mean and CVaR_{α} of the loss of a matching, where $\alpha \in (0, 100]$ is the confidence level of CVaR_{α}. The parameter $\gamma \geqslant 0$ is set by the user, and defines the trade-off between the mean and CVaR_{α}: $\gamma = 0$ means that this model only optimizes the expected matching weight, and $\gamma > 0$ includes the CVaR objective. We reformulate CVaR_{α} by introducing an auxiliary variable d, as in (Rockafellar, Uryasev, et al. 2000). Thus, (2) is equivalent to

$$\begin{aligned} \min_{\mathbf{y}, \mathbf{z}, d} \quad & \mathbb{E}_{\mathbb{P}}(\langle \mathbf{w}, \mathbf{W} \rangle) + \gamma \left\{ d + \frac{1}{\alpha} \mathbb{E}_{\mathbb{P}}\left[(\langle \mathbf{w}, \mathbf{W} \rangle - d)^+ \right] \right\} \\ s.t. \quad & \{\mathbf{y}, \mathbf{z}\} \in \mathscr{X}. \end{aligned} \tag{3}$$

In the following, we propose two approaches to solve (3) based on different data environments, which are sample average approximation (§3.1) and distributionally robust optimization (§3.2).

3.1 Solve Mean-Risk Model (3) with SAA

In this section, we assume the available edge weight measurements are representative of the *true* edge weight distribution. Our approach then uses these measurements to approximately solve Problem (3). If there is sufficient data to accurately characterize the edge weight distribution, we might instead try to solve Problem (3) exactly. However, even in this case, Problem (3) is not analytically solvable. We propose an SAA-based approach to *approximately* solve (3). The main results are summarized in Proposition 1. Suppose N *measurements* of each edge's weight are available or generated from the edge weight distribution; we denote the n-th measurement as $\hat{\mathbf{w}}_n$, $1 \leqslant n \leqslant N$. Problem (4) can be solved efficiently using general-purpose solvers like CPLEX.

Proposition 1 The SAA of (3) is equivalent to the following mixed-integer program.

$$\begin{aligned} \min_{\mathbf{y}, \mathbf{z}, d} \quad & \frac{1}{N} \sum_{n=1}^{N} \Pi_n + \gamma d \\ s.t. \quad & W_e = - \sum_{k \in \mathscr{K}(e)} y_{ek} - \sum_{c \in C} \mathbf{1}(e \in c) z_c, \forall e, \\ & \Pi_n \geqslant \langle \hat{\mathbf{w}}_n, \mathbf{W} \rangle, \forall n, \\ & \Pi_n \geqslant (1 + \frac{\gamma}{\alpha}) \langle \hat{\mathbf{w}}_n, \mathbf{W} \rangle - \frac{d\gamma}{\alpha}, \quad \forall n, \{\mathbf{y}, \mathbf{z}\} \in \mathscr{X}. \end{aligned} \tag{4}$$

3.2 Solve Mean-Risk Model (3) with DRO

In this section, we consider a setting where the available data are highly noisy and *not representative* of the true distribution. In this section, we refer to two different distributions: the *empirical* distribution (the distribution of the N initial weight measurements), and the *realized* distribution (the "true" underlying distribution). The empirical distribution is often estimated statistically and suffers from possible errors. For example, finite data size, existence of outliers, and dynamic patient and donor pools all lead to errors in the empirical distribution. Therefore, the mismatch between the empirical distribution and the true distribution may deteriorate the performance of the proposed model. To solve this problem, we propose a model based on distributionally robust optimization (DRO) to hedge against the possible errors in the empirical distribution. In the DRO approach, we assume that the true edge weight distribution $\mathbb{P}$ is contained in an ambiguity set $\mathscr{P}$, which is formulated as:

$$\begin{aligned} \min_{\mathbf{y},\mathbf{z},d} \max_{\mathbb{P}\in\mathscr{P}} \quad & \mathbb{E}_{\mathbb{P}}\left\{\max\left[\langle\hat{\mathbf{w}}_n,\mathbf{W}\rangle,(1+\frac{\gamma}{\alpha})\langle\hat{\mathbf{w}}_n,\mathbf{W}\rangle-\frac{d\gamma}{\alpha}\right]\right\}+\gamma d \\ s.t. \quad & \{\mathbf{y},\mathbf{z}\}\in\mathscr{X}. \end{aligned} \tag{5}$$

Ambiguity Set Based on Wasserstein Metric. Recently, the Wasserstein metric has achieved success in various applications. We construct the ambiguity set using a Wasserstein ball centered at the empirical distribution, which captures all the "similar" distributions; their similarities can be adjusted through the radius of the ball. Next we formalize this model, beginning with the definition of ambiguity set $\mathscr{P}$. Suppose $\mathbf{w}\in\mathbb{R}^{|E|}_{\geqslant 0}$, $\Xi\subseteq\mathbb{R}^{|E|}_{\geqslant 0}$, and $M(\Xi)$ contains all the probability distributions Q supported on Ξ with $\int_\Xi\|\mathbf{w}\|Q(d\mathbf{w})<\infty$. $\|\cdot\|$ represents an arbitrary p-norm. Let $\hat{\mathbb{P}}$ be the empirical distribution of the edge weight $\mathbf{w}$ and $\theta\geqslant 0$. Ambiguity set $\mathscr{P}$ is defined as: $\mathscr{P}=\mathscr{B}_\theta(\hat{\mathbb{P}})=\{\mathbb{P}\in M(\Xi): W(\mathbb{P},\hat{\mathbb{P}})\leqslant\theta\}$. We define $W(\mathbb{P},\hat{\mathbb{P}}):=\inf\left\{\int_{\Xi^2}\|\mathbf{w}-\mathbf{w}'\|\Pi(d\mathbf{w},d\mathbf{w}')\right\}$, where Π is a joint distribution of $\mathbf{w}$ and $\mathbf{w}'$ with marginals $\mathbb{P}$ and $\hat{\mathbb{P}}$. Using this definition of $\mathscr{P}$, Proposition 2 shows that (5) has a closed-form reformulation, which is a mixed-integer program if $\|\|_*$ is $\|\|_1$ or $\|\|_\infty$, and a mixed-integer conic program for other p-norms. Thus, it can be solved using solvers like CPLEX or Gurobi.

Proposition 2 Suppose each edge weight w_e satisfies $w_{min}\leqslant w_e\leqslant w_{max}$, $\forall e\in E$. Let N be the number of empirical edge measurements. Problem (5) is equivalent to

$$\begin{aligned} \min_{\mathbf{W},d,\lambda,\mathbf{s},\boldsymbol{\eta},\boldsymbol{\mu}} \quad & \lambda\theta+\frac{1}{N}\sum_{i=1}^{N}s_i+\gamma d \\ s.t. \quad & b_{k'}+\langle\mathbf{a}_{k'},\hat{\mathbf{w}}_i\rangle+\sum_{j=1}^{|E|}\eta_{ik'j}(w_{max}-\hat{w}_{ij})+\sum_{j=1}^{|E|}\mu_{ik'j}(\hat{w}_{ij}-w_{min})\leqslant s_i,\forall i\leqslant N,\forall k'\in\{1,2\}, \\ & \|\boldsymbol{\eta}_{ik'}-\boldsymbol{\mu}_{ik'}-\mathbf{a}_{k'}\|_*\leqslant\lambda,\forall i\leqslant N,\forall k'\in\{1,2\}, \\ & \eta_{i,k',j}\geqslant 0,\quad \forall i\leqslant N,\forall k'\in\{1,2\},\forall j\leqslant|E|, \\ & \mu_{i,k',j}\geqslant 0,\quad \forall i\leqslant N,\forall k'\in\{1,2\},\forall j\leqslant|E|, \\ & W_e=-\sum_{k\in\mathscr{K}(e)}y_{ek}-\sum_{c\in C}\mathbf{1}(e\in c)z_c,\forall e, \\ & \mathbf{a}_1=\mathbf{W},\ b_1=0,\ \mathbf{a}_2=(1+\frac{\gamma}{\alpha})\mathbf{W},\ b_2=-\frac{d\gamma}{\alpha},d,\lambda\in\mathbb{R},\ \mathbf{s}\in\mathbb{R}^N,\{\mathbf{y},\mathbf{z}\}\in\mathscr{X}. \end{aligned} \tag{6}$$

4 EXPERIMENTS

Next we demonstrate the benefits of both the SSA-CVaR method of Section 3.1 and the DRO-CVaR method of Section 3.2. For both of these experiments we use 32 randomly-generated exchange graphs resembling the structure of real exchanges, using anonymized data from the United Network for Organ Sharing (UNOS) US-wide kidney exchange. For additional information regarding the features available in

UNOS data, we refer the reader to (OPTN-Data 2021). To reflect edge weight uncertainty, we define an edge weight distribution inspired by real sources of uncertainty, described in the next section. To highlight the differences between the SAA and DRO approaches, we consider two uncertainty paradigms. First, in Section 4.2 we consider a setting where the empirical distribution is equivalent to the *true* underlying distribution. In Section 4.3 we relax this assumption, and we draw edge weight measurements from a *different* (noisy) distribution, to simulate a setting where the empirical distribution is not representative of the true distribution. This is reflective of real-world settings where data is sparse, or the true distribution is difficult to measure; these settings are the primary motivation for our DRO-CVaR approach.

4.1 Setting

Edge weight distribution: survival rate. We define a simple edge weight distribution based on the Living Donor Kidney Profile Index (LKDPI), a recent metric (Massie, Leanza, Fahmy, Chow, Desai, Luo, King, Bowring, and Segev 2016) that is now used as a decision support tool for kidney transplantation. Using the method of (Li, Lieberman, Macke, Carrillo, Ho, Wellen, and Das 2019), we assume that the survival rate of a donor organ is an exponential random variable, with mean proportional to the LKDPI. We define a simple distribution by assuming that edge weights depend *only* on the donor LKDPI.

Each donor node is randomly assigned to have *low LKDPI* (14.93) or *high LKDPI* (59.37), each with probability $1/2$; this corresponds to the mean LKDPI plus/minus one standard deviation of the distribution estimated by (Saidman, Roth, Sönmez, Ünver, and Delmonico 2006). After assigning an LKDPI for each donor, we draw edge weight (i.e., survival time) from the exponential distribution defined in (Li, Lieberman, Macke, Carrillo, Ho, Wellen, and Das 2019): $w_e \sim 14.78e^{-0.01239 \times \text{LKDPI}}$.

Stochastic and deterministic edges. In practice the variance of edge weight distributions can be very different—some edge weights may be very uncertain, while others may be nearly constant (Massie, Leanza, Fahmy, Chow, Desai, Luo, King, Bowring, and Segev 2016). To simulate this, we randomly assign edges to be either *stochastic* (drawn from its underlying distribution) or *deterministic* (the edge weight is equal to the *first* draw from its distribution). Note that we always define an edge distribution, but deterministic edges only use a single draw from this distribution. In our experiments, half of all edges are deterministic while the other half are probabilistic.

4.2 SAA-CVaR

In this setting, N edge weight measurements are made of each edge, and we assume that these measurements come from the *correct* edge weight distribution. We compare three methods that use these N measurements: our SAA-CVaR approach, a state-of-the-art Robust Optimization approach (McElfresh, Bidkhori, and Dickerson 2019), and a non-robust matching algorithm which maximizes the sample mean edge weight. We show that SAA-CVaR has better $\alpha\%$ worst-case performance than both other methods. Figure 1.

Implementation details. We denote the SAA-CVaR approach of Section 3.1 as `SAA`. For each of the 32 exchange graphs, we simulate the LKDPI edge weight distribution; we simulate 4 different edge weight distributions (randomly) for each graph (in effect, creating 4 different graph instances, each sharing the same underlying structure). For each graph-distribution pair we simulate $N = 200$ draws (i.e., *measurements*) from the edge weight distribution, and use these measurements to find optimal matchings using each of the following methods: `NR`: maximize the sample mean edge weight (where the sample mean is calculated over all N simulated weight measurements), `RO` the (non-distributionally-)robust optimization approach of (McElfresh, Bidkhori, and Dickerson 2019), and `SAA`, our proposed approach. For `RO` we use parameter $\Gamma = 5$, meaning that up to 5 edges may have realized weight of 0. For `SAA` we set α to 50, meaning that this approach aims to maximize realized matching weight over the 50% worst-case outcomes.To vary the balance between the mean matching weight and the worst-case objective we vary parameter γ, with $\gamma \in \{0.1, 1.0, 10.0, 100.0\}$. With $\gamma = 0$, `SAA` only optimizes the expected matching weight; as γ increases, the objective is weighted more toward the $\alpha\%$ worst-case outcomes.

Metrics. To test these methods, we simulate 1000 *realizations* of each edge weight. Our approach is designed to protect against the $\alpha\%$ worst-case outcomes; so, we compare the means of the lowest $\alpha\%$ realized matching weights for each method. We use the following metric to compare each method

$$\text{FRAC-NR} \equiv (\mu_\alpha^{\text{X}} - \mu_\alpha^{\text{NR}})/\mu_\alpha^{\text{NR}},$$

where μ_α^{X} is the mean of the $\alpha\%$ lowest-weight realizations for method `X`. We use the non-robust method `NR` as a baseline to compare the previous robust-optimization-based method `RO` with our proposed method `SAA`. In many cases the robust methods identified the same optimal matching as non-robust; in order to highlight the differences between robust and non-robust, we remove these cases from our results.

Results. Figure 1 shows the simulation results. Recall that parameter γ moderates between the *expected* edge weight and the CVaR objective (to maximize the $\alpha\%$ worst-case edge weight); as demonstrated in Fig. 1, when γ is small the behavior of `SAA` is similar to non-robust (i.e., FRAC-NR is almost always zero). With larger γ, `SAA` prioritizes the CVaR objective, and achieves a higher weight for $\alpha\%$ worst-case weight (i.e., FRAC-NR is positive). Note that the robust optimization approach of (McElfresh, Bidkhori, and Dickerson 2019) (`RO`) is far too conservative, and in fact achieves lower CVaR objective than non-robust.

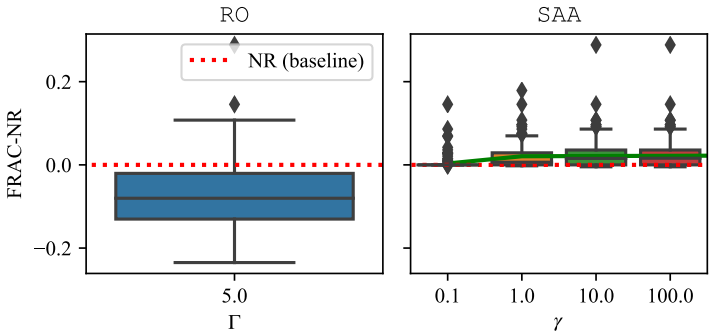

Figure 1: FRAC-NR for both the `RO` method due to McElfresh et al. (McElfresh, Bidkhori, and Dickerson 2019) and our `SAA` method, for $\gamma \in \{0.1, 1.0, 10.0, 100.0\}$.

4.3 DRO-CVaR

In this setting, unlike the previous section, we assume that edge weights may be drawn from an *incorrect* distribution. This setting—where the true distribution is unknown or difficult to measure—is the primary motivation for our distributionally robust approach, DRO-CVaR. We show that our DRO-CVaR approach achieves better *worst-case* performance than both the Robust Optimization approach of (McElfresh, Bidkhori, and Dickerson 2019) as well as our `SAA` approach (which is less conservative than DRO-CVaR). The main results are plotted in Figure 2. In these experiments edge weight "measurements" are drawn from the true edge weight distribution, while edge weight "realizations" have adversarial (negative) noise—to simulate errors in edge weight measurements.

Implementation details. We denote the DRO-CVaR approach of Section 3.2 as `DRO`. For each exchange graph, we simulate the LKDPI edge weight distribution; as before, we create 4 different edge weight distributions. For each graph-distribution pair we simulate $N = 20$ measurements from the edge distribution; these measurements are used by each matching approach. We use far smaller N in these experiments because the MILP formulation for `DRO` grows polynomially in N, and is intractable for large N (recall that the kidney exchange problem is NP-hard, though in practice small problems can be solved easily).

To simulate errors in the empirical weight distribution, we add noise to the *realized* weights—which is not present in the N *measured* weights provided to the matching methods. The `DRO` approach should protect

against *worst case* deviations from the empirical distributions; for this reason we create the adversarial case—where all noise is *negative*. After simulating each edge weight realization, we randomly assign half of the edges to receive negative noise of -10 or -20, each with probability $1/2$. As before, we simulate 1000 noisy realizations, and calculate FRAC-NR as described above. In these experiments we use $\alpha = 40$ (rather than $\alpha = 50$ as in the previous experiments) to make our approach more risk-averse, since edge weights are more "noisy" in these experiments. In these experiment we fix $\gamma = 10$ to strongly prioritize the CVaR objective. For the DRO approach we use $\theta \in \{0.1, 1, 2, 5, 10, 20\}$; the results are shown in Figure 2. As before, we ignore cases where these methods identify the same matching as non-robust.

Results. There is far less variance between methods in these experiments, likely due to the strong adversarial noise applied to each realization. However, it is clear that both SAA and DRO both achieve better (higher) worst-case outcomes than RO. We also see that DRO achieves a better worst-case FRAC-NR than both RO and SAA—which is expected, as DRO should protect against the worst-case outcome. That worst-case gain is substantial compared to the RO method and more marginal compared to the SAA method. However unlike our results in the previous section that were given access to a larger number of samples, all methods in this data-sparse environment appear to perform nearly on par with non-robust under the FRAC-NR measure. Next, we provide statistics about match composition under our methods.

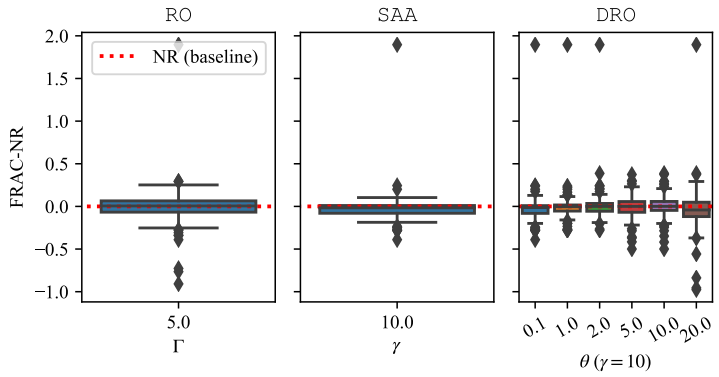

Figure 2: FRAC-NR for the RO method due to McElfresh et al. (McElfresh, Bidkhori, and Dickerson 2019), in addition to our two new methods SAA, and DRO.

4.4 Comparisons of the Structures in Matchings

Here we provide two tables detailing the differences in the matchings produced by each method in the experiments of the last two subsections. In short, these tables demonstrate robust methods tend to avoid *uncertain* edges, even if this results in a lower average matching weight.

Matched edges. First, we present the number of each *type* of edge matched by each method. Our experiments use four different edge types, which are randomly assigned based on each edge's donor, which we summarize as; high-weight deterministic (D-High), high-weight stochastic (S-High), low-weight deterministic (D-Low), and low-weight stochastic (S-Low). Both D-High and S-High have mean weight 12.3, while D-Low and S-Low have mean weight 7.1. However, both D-High and D-Low are exponentially distributed, meaning that the worst-case weight can be quite low (approaching zero) in any particular realization. Our CVaR approach is designed to balance between the mean matching weight, and the worst-case weight in $\alpha\%$ of all possible outcomes. Thus, we expect that robust optimization approaches should favor deterministic edges over probabilistic edges. Table 1 shows the number of each edge matched by each method, compared with non-robust.

Intuitively, most robust methods match more deterministic edges (D-High and D-Low), and fewer stochastic edges (S-High and S-Low) than non-robust. Notably, SAA tends to match *more* deterministic

Table 1: Total number of edges of each type matched by each method. The difference between each method and non-robust is indicated in parentheses.

Method	**D-High**	**S-High**	**D-Low**	**S-Low**
Non-Robust (baseline)	699 (0)	700 (0)	555 (0)	517 (0)
SAA ($\gamma = 0.1$)	706 (+7)	693 (-7)	558 (+3)	514 (-3)
SAA ($\gamma = 1.0$)	735 (+36)	661 (-39)	562 (+7)	514 (-3)
SAA ($\gamma = 10.0$)	747 (+48)	637 (-63)	571 (+16)	514 (-3)
DRO ($\gamma = 10.0$, $\theta = 0.1$)	747 (+48)	637 (-63)	571 (+16)	514 (-3)
DRO ($\gamma = 10.0$, $\theta = 1.0$)	749 (+50)	644 (-56)	555 (0)	505 (-12)
DRO ($\gamma = 10.0$, $\theta = 10.0$)	573 (-126)	769 (+69)	515 (-40)	545 (+28)
RO ($\gamma = 5$)	640 (-59)	683 (-17)	568 (+13)	580 (+63)

edges as γ increases. It is also notable that methods DRO and RO do not strictly follow this trend. As θ increases, DRO matches *fewer* determinisitc edges, and RO tend to match only low-weight edges (D-Low and S-Low). This behavior helps explain Figures 2 and 3: in contrast to SAA, methods DRO and RO are selecting edges with uncertain edge weights.

Cycle and chain structure. Next, we analyze the overall matching structure. In kidney exchange, the solution (matching) is composed of cycles and chains of a fixed length. The composition of each matching can be a useful indicator of the quality or robustness of a solution. Table 2 shows the total number of cycles and chains of each *length* matched by each method, over all exchange graphs. There is little difference between the *types* of structures produced by each method—i.e., the resulting cycle and chain lengths are very similar. As in the case of edge *existence* uncertainty, where risk-averse methods will select smaller structures due to their decreased fragility to edge failure, we see that nearly all robust methods result in more 2-cycles and fewer 3-cycles than non-robust.

Table 2: Total number of cycles and chains of each length matched by each method; difference between each method and non-robust is given in parentheses.

Method	**2-cycles**	**3-cycles**	**1-chains**	**2-chains**	**3-chains**	**4-chains**
Non-Robust (baseline)	198 (0)	433 (0)	48 (0)	50 (0)	52 (0)	118 (0)
SAA ($\gamma = 0.1$)	208 (+10)	428 (-5)	49 (+1)	44 (+1)	58 (+6)	115 (-3)
SAA ($\gamma = 1.0$)	203 (+5)	429 (-4)	44 (-4)	51 (-4)	59 (+7)	114 (-4)
SAA ($\gamma = 10.0$)	206 (+8)	425 (-8)	44 (-4)	47 (-4)	56 (+4)	119 (+1)
DRO ($\gamma = 10.0$, $\theta = 0.1$)	203 (+5)	425 (-8)	45 (-3)	47 (-3)	55 (+3)	121 (+3)
DRO ($\gamma = 10.0$, $\theta = 1.0$)	198 (0)	423 (-10)	45 (-3)	48 (-3)	57 (+5)	119 (+1)
DRO ($\gamma = 10.0$, $\theta = 10.0$)	202 (+4)	412 (-21)	55 (+7)	33 (+7)	55 (+3)	119 (+1)
RO ($\gamma = 5$)	215 (+17)	418 (-15)	44 (-4)	51 (-4)	43 (-9)	128 (+10)

5 CONCLUSIONS & FUTURE RESEARCH

We proposed data-driven methods to solve the mean-risk kidney exchange model (2), which is a real-world instantiation of the (cardinality-constrained) cycle packing problem. On realistic data drawn from a large, fielded US-based kidney exchange, we validated that our methods strike a balance between protecting against worst-case realizations and maintaining strong average-case performance. In many matching applications—including kidney allocation-it is likely that uncertainty is correlated, due to medical and/or non-medical reasons, with sensitive attributes such as race; past work has considered this in the deterministic setting (Klimentova, Viana, Pedroso, and Santos 2020; Farnadi, St-Arnaud, Babaki, and Carvalho 2021), and valuable future work could extend those approaches to our setting.

REFERENCES

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

Anderson, R., I. Ashlagi, D. Gamarnik, and A. E. Roth. 2015. "Finding Long Chains in Kidney Exchange Using the Traveling Salesman Problem". *Proceedings of the National Academy of Sciences* 112(3):663–668.

Bertsimas, D., V. Gupta, and N. Kallus. 2018. "Robust Sample Average pproximation". *Mathematical Programming* 171(1):217–282.

Bertsimas, D., and M. Sim. 2004. "The Price of Robustness". *Operations Research* 52(1):35–53.

Bidkhori, H., J. P. Dickerson, D. C. McElfresh, and K. Ren. 2020. "Kidney Exchange with Inhomogeneous Edge Existence Uncertainty". In *Conference on Uncertainty in Artificial Intelligence (UAI)*.

Biró, P., L. Burnapp, B. Haase, A. Hemke, R. Johnson, J. van de Klundert, and D. Manlove. 2017. "Kidney Exchange Practices in Europe". *First Handbook of the COST Action CA15210: European Network for Collaboration on Kidney Exchange Programmes (ENCKEP)*.

Biró, P., D. F. Manlove, and R. Rizzi. 2009. "Maximum Weight Cycle Packing in Directed Graphs, with Application to Kidney Exchange Programs". *Discrete Mathematics, Algorithms and Applications* 1(04):499–517.

Carvalho, M., X. Klimentova, K. Glorie, A. Viana, and M. Constantino. 2020. "Robust Models for the Kidney Exchange oroblem". *INFORMS Journal on Computing*.

Chen, Z.-Z., R. Tanahashi, and L. Wang. 2008. "An Improved Randomized Approximation Algorithm for Maximum Triangle Packing". In *International Conference on Algorithmic Applications in Management*, 97–108. Springer.

Delage, E., and Y. Ye. 2010. "Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems". *Operations Research* 58(3):595–612.

Dickerson, J. P., D. F. Manlove, B. Plaut, T. Sandholm, and J. Trimble. 2016. "Position-indexed formulations for kidney exchange". In *Proceedings of the 2016 ACM Conference on Economics and Computation*, 25–42.

Dickerson, J. P., A. D. Procaccia, and T. Sandholm. 2019. "Failure-Aware Kidney Exchange". *Management Science* 65(4):1768–1791. Earlier version appeared at EC-13.

Dickerson, J. P., and T. Sandholm. 2017. "Multi-Organ Exchange". *Journal of Artificial Intelligence Research* 60:639–679.

ErdHos, P., and T. Gallai. 1959. "On Maximal Paths and Circuits of Graphs". *Acta Mathematica Academiae Scientiarum Hungarica* 10(3-4):337–356.

Ergin, H., T. Sönmez, and M. U. Unver. 2017. "Multi-Donor Organ Exchange". *Working paper*.

Esfahani, P. M., and D. Kuhn. 2018. "Data-Driven Distributionally Robust Optimization Using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations". *Mathematical Programming* 171(1-2):115–166.

Farnadi, G., W. St-Arnaud, B. Babaki, and M. Carvalho. 2021. "Individual Fairness in Kidney Exchange Programs". In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 35, 11496–11505.

Freedman, R., J. S. Borg, W. Sinnott-Armstrong, J. P. Dickerson, and V. Conitzer. 2020. *Artificial Intelligence* 283:103261.

Kleywegt, A. J., A. Shapiro, and T. Homem-de Mello. 2002. "The Sample Average Approximation Method for Stochastic Discrete Optimization". *SIAM Journal on Optimization* 12(2):479–502.

Klimentova, X., J. P. Pedroso, and A. Viana. 2016. "Maximising Expectation of the Number of Transplants in Kidney Exchange Programmes". *Computers & Operations Research* 73:1–11.

Klimentova, X., A. Viana, J. P. Pedroso, and N. Santos. 2020. "Fairness Models for Multi-Agent Kidney Exchange Programmes". *Omega*:102333.

Li, Z., K. Lieberman, W. Macke, S. Carrillo, C.-J. Ho, J. Wellen, and S. Das. 2019. "Incorporating Compatible Pairs in Kidney Exchange: A Dynamic Weighted Matching Model". In *Proceedings of the 2019 ACM Conference on Economics and Computation*, 349–367. ACM.

LKDPI-Calculator 2015. "Live Donor KDPI Calculator". http://www.transplantmodels.com/lkdpi/.

Lovász, L. 1968. "On Covering of Graphs". In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, 231–236. Academic Press New York.

Manlove, D., and G. O'Malley. 2015. "Paired and Altruistic Kidney Donation in the UK: Algorithms and Experimentation". *ACM Journal of Experimental Algorithmics* 19(1).

Massie, A. B., J. Leanza, L. M. Fahmy, E. K. Chow, N. M. Desai, X. Luo, E. A. King, M. G. Bowring, and D. L. Segev. 2016. "A Risk Index for Living Donor Kidney Transplantation". *American Journal of Transplantation* 16(7):2077–2084.

McElfresh, D., H. Bidkhori, and J. P. Dickerson. 2019. "Scalable Robust Kidney Exchange". In *AAAI Conference on Artificial Intelligence (AAAI)*.

OPTN-Data 2021. "OPTN: Organ Procurement and Transplantation Network". https://optn.transplant.hrsa.gov/data.

Pyber, L. 1996. "Covering the Edges of a Connected Graph by Paths". *Journal of Combinatorial Theory, Series B* 66(1):152–159.

Rapaport, F. T. 1986. "The Case for a Living Emotionally Related International Kidney Donor Exchange Registry". *Transplantation Proceedings* 18:5–9.

Rockafellar, R. T., S. Uryasev et al. 2000. "Optimization of Conditional Value-at-Risk". *Journal of Risk* 2:21–42.
Roth, A. E., T. Sönmez et al. 2005. *American Economic Review* 95(2):376–380.
Roth, A. E., T. Sönmez, and M. U. Ünver. 2004. *The Quarterly Journal of Economics* 119(2):457–488.
Saidman, S. L., A. Roth, T. Sönmez, U. Ünver, and F. Delmonico. 2006. "Increasing the Opportunity of Live Kidney Donation by Matching for Two and Three Way Exchanges". *Transplantation* 81(5):773–782.
Shapiro, A., D. Dentcheva, and A. Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM.
Shapiro, A., and T. Homem-de Mello. 1998. "A Simulation-based Approach to Two-Stage Stochastic Programming with Recourse". *Mathematical Programming* 81(3):301–325.
UNOS 2015. "Revising Kidney Paired Donation Pilot Program Priority Points". OPTN/UNOS Public Comment Proposal.
Zheng, Q. P., S. Shen, and Y. Shi. 2015. "Loss-Constrained Minimum Cost Flow under Arc Failure Uncertainty with Applications in Risk-Aware Kidney Exchange". *IIE Transactions* 47(9):961–977.

AUTHOR BIOGRAPHIES

DUNCAN C. MCELFRESH is a Ph.D. candidate in Applied Mathematics at the University of Maryland. He earned his Bachelor's and Master's degree from the Colorado School of Mines. His research interests include applications of algorithms and AI for resource allocation and decision making. His email address is dmcelfre@umd.edu. His website is https://duncanmcelfresh.github.io/.

KE REN is a Ph.D. candidate at the Department of Industrial Engineering at the University of Pittsburgh. He earned his Master's degree in Electrical and Electronics Engineering from the University of Rochester. His research interests include machine learning, optimization, artificial intelligence, and supply chain management. His email address is KER102@pitt.edu.

JOHN P. DICKERSON is an assistant professor in the Department of Computer Science at the University of Maryland, with affiliate appointments in Applied Mathematics & Scientific Computing, as well as in the Human-Computer Interaction Lab. He is a 2019 NSF CAREER Awardee, as well as a 2019 Google Faculty Research Awardee, one of IEEE *Intelligent Systems* 2020 AI's 10 to Watch, and a 2021 Google Research Scholar Awardee. He received a PhD in Computer Science from Carnegie Mellon University, where he was a Facebook fellow, NDSEG fellow, and Siebel scholar. His email address is john@cs.umd.edu. His website is http://jpdickerson.com/.

HODA BIDKHORI is an assistant professor at the Department of Industrial Engineering at the University of Pittsburgh. She earned her Ph.D. in Applied Mathematics from the Massachusetts Institute of Technology (MIT), where she subsequently spent several years as a postdoctoral researcher and lecturer in Operations Research and Statistics. Her current research focuses on the theory and applications of data-driven decision-making. Her email address is bidkhori@pitt.edu. Her website is https://sites.google.com/view/hodabidkhori/.