Balancing Lexicographic Fairness and a Utilitarian Objective with Application to Kidney Exchange

Duncan C. McElfresh[†,‡]
[†]Department of Mathematics
University of Maryland
dmcelfre@math.umd.edu

John P. Dickerson[†,‡]
[‡]Department of Computer Science
University of Maryland
john@cs.umd.edu

Abstract

Balancing fairness and efficiency in resource allocation is a classical economic and computational problem. The price of fairness measures the worst-case loss of economic efficiency when using an inefficient but fair allocation rule; for indivisible goods in many settings, this price is unacceptably high. One such setting is kidney exchange, where needy patients swap willing but incompatible kidney donors. In this work, we close an open problem regarding the theoretical price of fairness in modern kidney exchanges. We then propose a general hybrid fairness rule that balances a strict lexicographic preference ordering over classes of agents, and a utilitarian objective that maximizes economic efficiency. We develop a utility function for this rule that favors disadvantaged groups lexicographically; but if cost to overall efficiency becomes too high, it switches to a utilitarian objective. This rule has only one parameter which is proportional to a bound on the price of fairness, and can be adjusted by policymakers. We apply this rule to real data from a large kidney exchange and show that our hybrid rule produces more reliable outcomes than other fairness rules.

1 Introduction

Chronic kidney disease is a worldwide problem whose societal burden is likened to that of diabetes (Neuen *et al.* 2013). Left untreated, it leads to end-stage renal failure and the need for a donor kidney—for which demand far outstrips supply. In the United States alone, the kidney transplant waiting list grew from $58{,}000$ people in 2004 to over $100{,}000$ needy patients (Hart *et al.* 2016).[1]

To alleviate some of this supply-demand mismatch, *kidney exchanges* (Rapaport 1986; Roth *et al.* 2004) allow patients with willing *living* donors to trade donors for access to compatible or higher-quality organs. In addition to these patient-donor pairs, modern exchanges include *non-directed donors*, who enter the exchange without a patient in need of a kidney. Exchanges occur in cycle- or chain-like structures, and now account for 10% of living transplants in the United States. Yet, access to a kidney exchange does not guarantee equal access to kidneys themselves; for example, certain classes of patients may be particularly disadvantaged based on health characteristics or other logistical factors. Thus, *fairness* considerations are an active topic of theoretical and practical research in kidney exchange and the matching market community in general.

Intuitively, any enforcement of a fairness constraint or consideration may have a negative effect on overall economic efficiency. A quantification of this tradeoff is known as the *price of fairness* (Bertsimas *et al.* 2011). Recent work by Dickerson *et al.* (2014) adapted this concept to the kidney exchange case, and presented two fair allocation rules that strike a balance between fairness and efficiency. Yet, as we show in this paper, those rules can "fail" unpredictably, yielding an arbitrarily high price of fairness.

With this as motivation, we adapt to the kidney exchange case a recent technique for trading off a form of fairness and utilitarianism in a principled manner. This technique is parameterized by a bound on the price of fairness, as opposed to a set of parameters that may result in hard-to-predict final matching behavior, as in past work. We implement our rule in a realistic mathematical programming framework and–on real data from a large, multi-center, fielded kidney exchange–show that our rule effectively balances fairness and efficiency without unwanted outlier behavior.

1.1 Related Work

We briefly overview related work in balancing efficiency and fairness in resource allocation problem. Bertsimas *et al.* (2011) define the price of fairness; that is, the relative loss in system efficiency under a fair allocation rule. Hooker and Williams (2012) give a formal method for combining utilitarianism and equity. We direct the reader to those two papers for a greater overview of research in fairness in general resource allocation problems.

Fairness in the context of kidney exchange was first studied by Roth *et al.* (2005b); they explore concepts like Lorenz dominance in a stylized model, and show that preferring fair allocations can come at great cost. Li *et al.* (2014) extend this model and present an algorithm to solve for a Lorenz dominant matching. Stability in kidney exchange, a concept intimately related to fairness, was explored by Liu *et al.* (2014). The use of randomized allocation machanisms to promote fairness in stylized models is theoretically promising (Fang *et al.* 2015; Aziz *et al.* 2016; Mattei *et al.* 2017). Recent work discusses fairness in stylized random graph models of dynamic kidney exchange (Ashlagi *et al.* 2013; Anderson *et al.* 2015). None of these papers provide practi-

[1] https://optn.transplant.hrsa.gov/converge/data/

cal models that could be implemented in a fully-realistic and fielded kidney exchange.

Practically speaking, Yılmaz (2011) explores in simulation equity issues from combining living and deceased donor allocation; that paper is limited to only short length-two kidney swaps, while real exchanges all use longer cycles and chains. Dickerson *et al.* (2014) introduced two fairness rules explicitly in the context of kidney exchange, and proved bounds on the price of fairness under those rules in a random graph model; we build on that work in this paper, and describe it in greater detail later. That work has been incorporated into a framework for learning to balance efficiency, fairness, and dynamism in matching markets (Dickerson and Sandholm 2015); we note that the fairness rule we present in this paper could be used in that framework as well.

1.2 Our Contributions

Dickerson *et al.* (2014) finds that the theoretical price of fairness in kidney exchange is small when *only* patient-donor pairs participate in the exchange. They did not include non-directed donors (NDDs). However, in modern kidney exchanges, non-directed donors (NDDs) provide many more matches than patient-donor pairs; furthermore, NDDs create more opportunities to expand the fair matching, potentially increasing the price of fairness. Here, we prove that adding NDDs to the theoretical model actually *decreases* the price of fairness, and that—with enough NDDs—the price of fairness is zero.

Real kidney exchanges are less dense and more uncertain than the (standard) theoretical model in which we prove our results. Previous approaches to incorporating fairness into kidney exchange have neglected this fact: they have been either ad-hoc—e.g., "priority points" decided on by committee (Kidney Paired Donation Work Group 2013)—or brittle (Roth *et al.* 2005b; Dickerson *et al.* 2014), resulting in an unacceptably high price of fairness. This paper provides the first approach to incorporating fairness into kidney exchange in a way that both prioritizes disadvantaged participants, but also comes with acceptable worst-case guarantees on the price of fairness. Our method is easily applied as an objective in the mathematical-programming-based clearing methods used in today's fielded exchanges; indeed, using real data we show that this method guarantees a limit on efficiency loss.

Section 1.3 introduces the kidney exchange problem. Section 2 extends work by Ashlagi and Roth (2014) and Dickerson *et al.* (2014), showing that the price of fairness is small on the canonical random graph model even with NDDs. Section 3 shows that two recent fair allocation rules from the kidney exchange literature (Dickerson *et al.* 2014) can perform unacceptably poorly in the worst case. Then, Section 4 presents a new allocation rule that allows policymakers to set a limit on efficiency loss, while also favoring disadvantaged patients. Section 5 shows on real data from a large fielded kidney exchange that our method limits efficiency loss while still favoring disadvantaged patients when possible.

1.3 Preliminaries

A kidney exchange can be represented as a directed *compatibility graph* $G = (V, E)$, with vertices $V = P \cup N$ including both patient donor pairs $p \in P$ and non-directed-donors $n \in N$ (Roth *et al.* 2004; 2005a; 2005b; Abraham *et al.* 2007). A directed edge e is drawn from vertex v_i to v_j if the donor at v_i can give to the patient at v_j. Fielded kidney exchanges consist mainly of directed cycles in G, where each patient vertex in the cycle receives the donor kidney of the previous vertex. Modern exchanges also include non-cyclic structures called chains (Montgomery *et al.* 2006; Rees *et al.* 2009). Here, an NDD donates her kidney to a patient, whose paired donor donates her kidney to another patient, and so on.

In practice, cycles are limited in size, or "capped," to some small constant L, while chains are limited in size to a much larger constant R—or not limited at all. This is because all transplants in a cycle must execute *simultaneously*; if a donor whose paired patient had already received a kidney backed out of the donation, then some participant in the market would be strictly worse off than before. However, chains need not be executed simultaneously; if a donor backs out after her paired patient receives a kidney, then the chain breaks but no participant is strictly worse off. We will discuss how these caps affect fairness and efficiency in the coming sections.

The goal of kidney exchange programs is to find a *matching* M—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 (although ongoing work explores multi-donor kidney exchange (Ergin *et al.* 2017; Farina *et al.* 2017)). The *clearing problem* in kidney exchange is to find a matching M^* that maximizes some utility function $u : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M}$ is the set of all legal matchings. Real kidney exchanges typically optimize for 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 disadvantaging others. This is formalized in the following section.

1.4 The Price of Fairness

As an example for this paper, we focus on *highly-sensitized* patients, who have a very low probability of their blood passing a feasibility test with a random donor organ; thus, finding a kidney is often quite hard, and their median waiting time for an organ jumps by a factor of three over less sensitized patients.[2] Utilitarian objectives will, in general, marginalize these patients. Sensitization is determined using the Calculated Panel Reactive Antibody (CPRA) level of each patient, which reflects the likelihood that a patient will find a matching donor.

Formally the sensitization of each patient-donor vertex v be $v_s \in [0, 100]$, the CPRA level of v's patient; NDD vertices are not associated with patients, so they do not have sensitization levels. Each patient-donor vertex $v \in P$ is considered highly sensitized if v_s exceeds threshold $\tau \in$

[2] https://optn.transplant.hrsa.gov/data/

$[0, 100]$, and lowly-sensitized otherwise. These vertex sets V_H and V_L are defined as:

- Lowly sensitized: $V_L = \{v \mid v \in P : v_s < \tau\}$
- Highly sensitized: $V_H = \{v \mid v \in P : v_s \geq \tau\}$.

By definition, highly-sensitized patients are harder to match than lowly-sensitized patients. Naturally, efficient matching algorithms prioritize easy-to-match vertices in V_L, marginalizing V_H. Let $u_f : \mathcal{M} \to \mathbb{R}$ be a *fair* utility function. Formally, a utility function is fair when its corresponding optimal match M_f^* is viewed as fair, where M_f^* is defined as:

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

Bertsimas *et al.* (2011) defined the *price of fairness* to be the "relative system efficiency loss under a fair allocation assuming that a fully efficient allocation is one that maximizes the sum of [participant] utilities." Caragiannis *et al.* (2009) defined an essentially identical concept in parallel. Formally, given a fair utility function u_f and the utilitarian utility function u, the price of fairness is:

$$\mathsf{POF}(\mathcal{M}, u_f) = \frac{u\left(M^*\right) - u\left(M_f^*\right)}{u\left(M^*\right)} \tag{1}$$

The price of fairness $\mathsf{POF}(\mathcal{M}, u_f)$ is the relative loss in (utilitarian) efficiency caused by choosing the fair outcome M_f^* rather than the most efficient outcome.

In the next section we show that the theoretical price of fairness in kidney exchange is small, even when both cycles *and chains* are used—thus generalizing an earlier result due to Dickerson *et al.* (2014) to modern kidney exchanges.

2 The Theoretical Price of Fairness with Chains is Low (or Zero)

In this section we use the random graph model for kidney exchange introduced by Ashlagi and Roth (2014) to show that the theoretical price of fairness is always small, especially when NDDs are included. A complete description of this model can be found in Appendix A.1. Dickerson *et al.* (2014) finds that without NDDs, the maximum price of fairness is $2/33$. Adding NDDs to this model creates more opportunities to match highly sensitized patients, which could potentially lead to a higher price of fairness. However we find that including chains in this model only *decreases* the price of fairness; furthermore, when the ratio of NDDs to patient-donor pairs is high enough, the price of fairness is zero.

2.1 Price of Fairness

Ashlagi and Roth (2014) characterize efficient matchings in a random graph model without chains, and Dickerson *et al.* (2014) build on this to show that the price of fairness without chains is bounded above by $2/33$. Dickerson *et al.* (2012) extend the efficient matching of Ashlagi and Roth (2014) to include chains, but do not calculate the price of fairness. In this work, we close the remaining gap in theory regarding the price of fairness with chains.

Given $|P|$ patient-donor pairs, we parameterize the number of NDDs $|N|$ with $\beta \geq 0$ such that $|N| = \beta|P|$. Theorems 1 and 2 state our two main results: adding chains to the random graph model does not increase the price of fairness, and when the fraction of NDDs is high enough ($\beta > 1/8$), the price of fairness is zero. The proofs of the following theorems are given in Appendix A.

Theorem 1. *Adding NDDs to the random graph model ($\beta > 0$) does not increase the upper bound on the price of fairness found by Dickerson* et al. *(2014).*

Proof Sketch: We explore every possible efficient matching on the random graph model with chains; only four of these matchings have nonzero price of fairness. For each case, we compare the price of fairness to that of the efficient matching without chains found in Dickerson *et al.* (2014), and find that the upper bound does not increase.

Theorem 2. *The price of fairness is zero when $\beta > 1/8$.*

Proof sketch: For each matching with nonzero price of fairness, $\beta \leq 1/8$. When $\beta > 1/8$, a different matching occurs, and the price of fairness is zero.

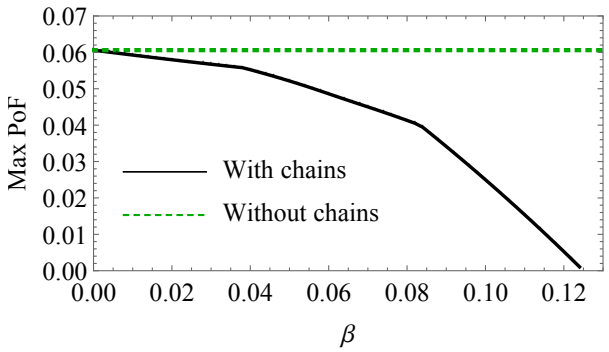


Figure 1: Price of fairness with chains. (The horizontal dotted line at $2/33$ is the price of fairness without chains.)

To illustrate these results, we compute the price of fairness when $\beta \in [0, 1/8]$. These calculations confirm our theoretical results, as shown in Figure 2.1: the price of fairness decreases as β increases, and is zero when $\beta > 1/8$.

The worst-case price of fairness is small in the random graph model, with or without NDDs. However, real exchange graphs are typically much sparser and less uniform—in reality the price of fairness can be high. In the next section, we discuss two notions of fairness in kidney exchange and determine their worst-case price of fairness.

3 The Price of Fairness in State-of-the-Art Fair Rules can be Arbitrarily Bad

The price of fairness depends on how fairness is defined. This is especially true in real exchanges where the price of fairness can be unacceptably high.

In this section, we discuss two kidney-exchange-specific fairness rules introduced by Dickerson *et al.* (2014): lexicographic fairness and weighted fairness. These rules favor the disadvantaged class, or classes, without considering overall

loss in efficiency; we will show in the worst case these rules allow the the price of fairness to approach 1 (i.e., total efficiency loss). Proofs of these theorems are in Appendix B.

3.1 Lexicographic Fairness

As proposed by Dickerson *et al.* (2014), α-lexicographic fairness assigns nonzero utility only to matchings that award at least a fraction α of the maximum possible fair utility. Letting $u_H(M)$ and $u_L(M)$ be the utility assigned to only vertices in V_H and V_L, respectively, the utility function for α-lexicographic fairness is given in Equation (2).

$$u_\alpha(M) = \begin{cases} u_L(M) + u_H(M) \\ \qquad \text{if } u_H(M) \geq \alpha \max\limits_{M' \in \mathcal{M}} u_H(M') \\ 0 \qquad \text{otherwise.} \end{cases} \tag{2}$$

Theorems 3 and 4 state that strict lexicographic fairness ($\alpha = 1$) allows the price of fairness to approach 1.

Theorem 3. *For any cycle cap L there exists a graph G such that the price of fairness of G under α-lexicographic fairness with $0 < \alpha \leq 1$ is bounded by $\textsf{POF}(\mathcal{M}, u_\alpha) \geq \frac{L-2}{L}$.*

Theorem 4. *For any chain cap R there exists a graph G such that the price of fairness of G under the α-lexicographic fairness rule with $0 < \alpha \leq 1$ is bounded by $\textsf{POF}(\mathcal{M}, u_\alpha) \geq \frac{R-1}{R}$.*

Thus, α-lexicographic fairness allows for a price of fairness that approaches 1 as the cycle and chain cap increase.

3.2 Weighted Fairness

The weighted fairness rule (Dickerson *et al.* 2014) defines a utility function by first modifying the original edge weights w_e by a multiplicative factor $\gamma \in \mathbb{R}$ such that

$$w'_e = \begin{cases} (1+\gamma)w_e & \text{if } e \text{ ends in } V_H \\ w_e & \text{otherwise.} \end{cases}$$

Then the weighted fairness rule u_{WF} is

$$u_{WF}(M) = \sum_{c \in M} u'(c),$$

where $u'(c)$ is the utility of a chain or cycle c with modified edge weights.

The modified edge weights prompt the matching algorithm to include more highly-sensitized patients; as in the lexicographic case, we now show that the price of fairness approaches 1 under weighted fairness.

Theorem 5. *For any cycle cap L and $\gamma \geq L-1$, there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\textsf{POF}(\mathcal{M}, u_{WF}) \geq \frac{L-2}{L}$.*

Theorem 6. *For any chain cap R and $\gamma \geq R-1$, there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\textsf{POF}(\mathcal{M}, u_{WF}) \geq \frac{R-1}{R}$.*

In the worst case, weighted fairness allows a price of fairness that approaches 1 as the cycle and chain caps increase. The price of fairness also approaches 1 as γ increases.

Theorem 7. *With no chain cap, there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\textsf{POF}(\mathcal{M}, u_{WF}) \geq \frac{\gamma}{\gamma+1}$.*

A similar result exists with cycles rather than chains.

Theorem 8. *With no cycle cap there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\textsf{POF}(\mathcal{M}, u_{WF}) \geq \frac{\gamma}{\gamma+1}$.*

These bounds show that weighted fairness allows for a price of fairness that approaches 1, i.e., arbitrarily bad, as the cycle cap, chain cap, or γ increase.

We have shown that the worst-case prices of fairness approach 1 under both the lexicographic and weighted fairness rules of Dickerson *et al.* (2014). Next, we propose a rule that favors disadvantaged groups, but also strictly *limits* the price of fairness using a parameter set by policymakers.

4 Hybrid Fairness Rule

In this section, we present a hybrid fair utility function that balances lexicographic fairness and a utilitarian objective. We generalize the hybrid utility function proposed by Hooker and Williams (2012), which chooses between a Rawlsian (or maximin) objective and a utilitarian objective for multiple classes of agents.

4.1 Utilitarian and Rawlsian Fairness

Consider two classes of agents that receive utilities $u_1(X)$ and $u_2(X)$, respectively, for outcome X. The fairness rule introduced by Hooker and Williams (2012) maximizes the utility of the worst-off class, unless this requires taking too many resources from other classes. When the inequality exceeds a threshold Δ (i.e., $|u_1(X) - u_2(X)| > \Delta$) they switch to a utilitarian objective that maximizes $u_1(X) + u_2(X)$. The utility function for this rule is

$$u_\Delta(X) = \begin{cases} 2\min(u_1(X), u_2(X)) + \Delta \\ \qquad \text{if } |u_1(X) - u_2(X)| \leq \Delta \\ \\ u_1(X) + u_2(X) \\ \qquad \text{otherwise.} \end{cases}$$

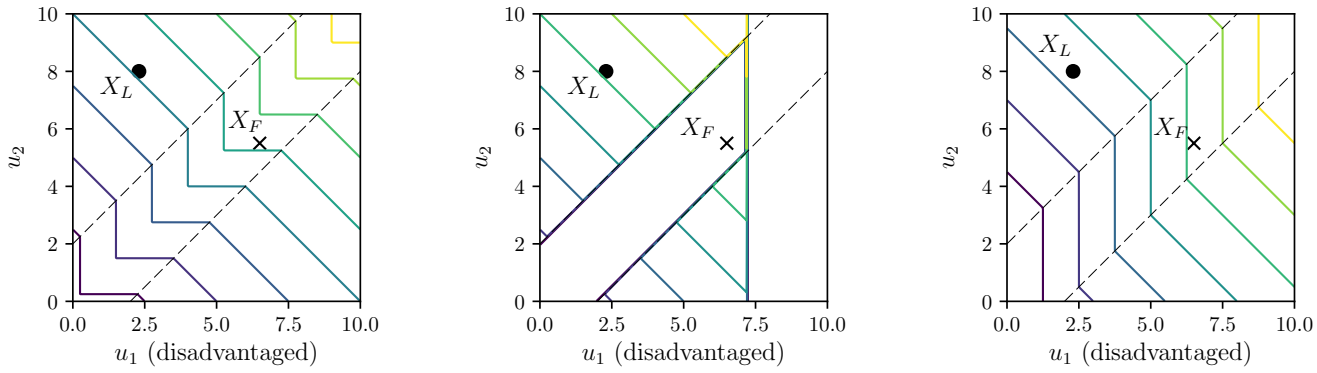
The parameter Δ is problem-specific, and should be chosen by policymakers. Figure 2(a) shows the level sets of this utility function, with $\Delta = 2$. This utility function can be generalized by switching to a different fairness rule in the *fair region* (i.e. when $|u_1(X) - u_2(X)| \leq \Delta$). The next section generalizes this rule using lexicographic fairness.

4.2 Hybrid-Lexicographic Rule

When it is desirable to favor one class of agents g_1 over class g_2, lexicographic fairness favors g_1. We propose a rule that implements lexicographic fairness only when inequality between groups does not exceed Δ. This rule uses two steps: 1) determine whether inequality is small enough to use lexicographic fairness 2) choose the optimal outcome. These steps are outlined below, and formalized in Algorithm 1.

Step 1: Find all outcomes that maximize a hybrid utility function, and determine whether lexicographic fairness is appropriate.

We use a utility function to identify outcomes that satisfy either a lexicographic or utilitarian objective. Equation (3) shows one option for such a utility function, which assigns



(a) Hybrid Rawlsian-Utilitarian (b) Δ-Lexicographic ($u_{\Delta 1}$) (c) Relaxed Δ-Lexicographic (u_{Δ})

Figure 2: Level sets for hybrid fair utility functions with $\Delta = 2$, with example outcomes X_L and X_F.

strict lexicographic utility ($\alpha = 1$) according to Equation (2) in the fair region, and utilitarian utility otherwise.

$$u_{\Delta 1}(X) = \begin{cases} u_1(X) + u_2(X) & \text{if } |u_1(X) - u_2(X)| \leq \Delta \\ & \text{and } u_1(X) = \max_{X' \in \mathcal{X}} (u_1(X')) \\ u_1(X) + u_2(X) & \text{if } |u_1(X) - u_2(X)| > \Delta \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

where $\mathcal{X}$ is the set of all possible outcomes. Figure 2(b) shows the contours $u_{\Delta 1}$. This utility function is clearly too harsh—it assigns zero utility to outcomes in the fair region that do not maximize u_1, and its optimal outcomes are not always Pareto efficient. Consider outcomes X_F and X_L in Figure 2(b). X_F is in the fair region but does not maximize u_1, so $u_{\Delta 1}(X_F) = 0$; X_L is in the utilitarian region but is less efficient, so $u_{\Delta 1}(X_L) = u(X_L)$. Under utility function $u_{\Delta 1}$, the less-efficient outcome X_L is chosen over X_F.

To address this problem we introduce u_Δ in Equation (4), which relaxes $u_{\Delta 1}$. For outcomes in the fair region (that is, with $|u_1 - u_2| \leq \Delta$), utility is assigned proportional to u_1. As shown in Figure 2(c), the contours of u_Δ are continuous.

$$u_\Delta(X) = \begin{cases} u_1(X) + u_2(X) - \Delta & \text{if } u_2(X) - u_1(X) > \Delta \\ 2u_1(X) & \text{if } |u_1(X) - u_2(X)| \leq \Delta \\ u_1(X) + u_2(X) + \Delta & \text{if } u_1(X) - u_2(X) > \Delta \end{cases} \tag{4}$$

Let X_{OPT} be the set of outcomes that maximize u_Δ. If any outcomes in X_{OPT} are in the utilitarian region , then any utilitarian-optimal outcome is selected. However, if any outcomes in X_{OPT} are in the fair region, then Step 2 must be used. This process is described below, and formalized in Algorithm 1.

Step 2: If any solution in X_{OPT} is in the fair region, select the lexicographic-optimal solution in the fair region.

The utility function u_Δ assigns the same utility to all solutions in the fair region with the same $u_1(X)$, no matter the value of $u_2(X)$. However, if there exist two outcomes X_A and X_B such that $u_1(X_A) = u_1(X_B)$ and $u_2(X_A) > u_2(X_B)$, then X_A is lexicographically preferred to X_B.

Algorithm 1 FairMatching

Input: Threshold Δ, matchings $\mathcal{M}$
Output: Fair matching M^*

```
M_OPT ← arg max_{M∈M} u_Δ(M)
if |M_OPT| > 1 then
    Select a matching M ∈ M_OPT
    if M is in the utilitarian region then
        M* ← M
    else
        M_1 ← {M' ∈ M_OPT | u_1(M') = u_1(M)}
        M* ← arg max_{M'∈M_1} u_2(M')
else
    M* ← M_OPT
```

4.3 Hybrid Rule for Several Classes

We now generalize the hybrid-lexicographic fairness rule to more than two classes. Consider a set $\mathcal{P}$ of classes g_i, $i = 1, \ldots, |\mathcal{P}|$. Let there be an ordering $\succ$ over g_i, where $g_a \succ g_b$ indicates that g_a should receive higher priority over g_b. WLOG, let the preference ordering over g_i be $g_1 \succ g_2 \succ \cdots \succ g_P$. Let $u_i(X)$ be the utility received by group i under outcome X. As in the previous section, we 1) use a utility function to determine whether lexicographic fairness is appropriate, then 2) select either a lexicographic- or utilitarian-optimal outcome.

Step 1: To define a utility function, we observe that in Equation (4), in the utilitarian region a positive offset Δ is added if $u_1(X) > u_2(X)$, and a negative offset is added otherwise. With $|\mathcal{P}|$ classes, each solution in the utilitarian region receives a utility offset of $+\Delta$ if $u_1(X) > u_i(X)$, and $-\Delta$ otherwise, for each class $i = 2, 3, \ldots, |\mathcal{P}|$. As in the previous section, these offsets ensure continuity in the utility function, and ensure that at least one of the maximizing outcomes will be Pareto optimal.

$$u_\Delta(X) = \begin{cases} |\mathcal{P}| \cdot u_1(X) \\ \qquad \text{if } \max_i(u_i(X)) - \min_i(u_i(X)) \leq \Delta, \\ u_1(X) + \sum_{i=2}^{|\mathcal{P}|}(u_i(X) + \text{sgn}(u_1(X) - u_i(X))\Delta) \\ \qquad \text{otherwise} \end{cases} \tag{5}$$

Step 2: Let X_{OPT} be the set of solutions that maximize u_Δ. If all optimal solutions are in the utilitarian region, any utilitarian-optimal solution is selected. If any optimal solution is in the fair region, then the lexicographic-optimal solution in the fair region must be selected, subject to the preference ordering $g_1 \succ g_2 \succ \cdots \succ g_{|\mathcal{P}|}$.

Algorithm 2 FairMatching for $|\mathcal{P}| \geq 2$ classes

Input: Threshold Δ, matchings $\mathcal{M}$
Output: Fair matching M^*

```
M_OPT ← arg max_{M∈M} u_Δ(M)
if |M_OPT| > 1 then
    Select a matching M ∈ M_OPT
    if M in utilitarian region then
        M* ← M
    else
        M_1 ← {M' ∈ M_OPT | u_1(M') = u_1(M)}
        for i = 2, ..., |P| do
            M_i ← arg max_{M'∈M_{i-1}} u_i(M')
        M* ← any matching in M_{|P|}
else
    M* ← M_OPT
```

4.4 Price of Fairness for the Hybrid-Lexicographic Rule

Theorem 9 gives a bound on the price of fairness for the hybrid-lexicographic rule; its proof is given in Appendix B.

Theorem 9. *Assume the optimal utilitarian outcome X_E receives utility $u(X_E) = u_E$, with most prioritized class $g_1 \in \mathcal{P}$ receiving utility u_1, and Z other classes $g_i \in \mathcal{P}$ such that $u_1(X_E) > u_i(X_E)$. Then, $\mathsf{POF}(\mathcal{M}, u_\Delta) \leq \frac{2((|\mathcal{P}|-1)-Z)\Delta}{u_E}$.*

4.5 Hybrid Fairness in Kidney Exchange

The hybrid-lexicographic fairness rule in Equation (4) is easily applied to kidney exchange, with u_H and u_L the total utility received by highly-sensitized and lowly-sensitized patients, respectively,

$$u_\Delta(M) = \begin{cases} u_L(M) + u_H(M) - \Delta & \text{if } u_L(M) - u_H(M) > \Delta \\ 2u_H(M) & \text{if } |u_L(M) - u_H(M)| \leq \Delta \\ u_L(M) + u_H(M) + \Delta & \text{if } u_H(M) - u_L(M) > \Delta \end{cases} \tag{6}$$

In the following section, we demonstrate the practical effectiveness of the hybrid-lexicographic rule by testing it on real kidney exchange data.

5 Experiments

In this section, we compare the behavior of α-lexicographic, weighted, and hybrid-lexicographic fairness. All code for these experiemnts are available on GitHub.[3] We use each rule to find the optimal fair outcomes for 314 real kidney exchanges from the United Network for Organ Sharing (UNOS), collected between 2010 and 2016. To solve the kidney exchange clearing problem (KEP) we use the PICEF formulation introduced by Dickerson *et al.* (2016), with cycle cap 3 and various chain caps. In real exchanges, not all recommended edges in a matching result in successful transplants. To reflect this uncertainty, we use the concept of failure-aware kidney exchange introduced in (Dickerson *et al.* 2013): all edges in the exchange can fail with probability $(1 - p)$; the matching algorithm maximizes *expected* matching weight, considering edge success probability p.

5.1 Procedure

For each UNOS exchange graph G, we use the following procedure to implement each fairness rule. We repeat the following procedure for chain caps 0, 3, 10, and 20, and for edge success probabilities $p = 0.1n$, with $n = 1, 2, \ldots, 10$.

1. Find the efficient matching M_E by solving the to optimality the NP-hard kidney exchange problem (KEP) on G.
2. Find the fair matching M_F by solving the KEP on $G' = (V, E')$, where each edge $e \in E'$ has weight 1 if e ends in V_H and 0 otherwise.
3. **Weighted Fairness:** Find the γ-fair matching M_γ by solving the KEP on $G^\gamma = (V, E^\gamma)$, where each edge $e \in E^\gamma$ has weight $1 + \gamma$ if e ends in V_H and 1 otherwise. After finding M_γ, the reported utilities are calculated using edge weights of E and not E'. We use weight parameters $\gamma = 2n$, with $n = 0, 1, 2, \ldots, 10$.
4. **α-Lexicographic Fairness:** Find the α-fair matching M_α by solving the KEP on G, with the additional constraint $u_H(M_\alpha) \geq \alpha u_H(M_E)$. We use parameters $\alpha = 0.1n$, with $n = 0, 1, 2, \ldots, 10$.
5. **Hybrid-Lexicographic Fairness:** Find the Δ-fair matching M_Δ using the α-fair matchings M_α, and Algorithm 1. That is, $M_\Delta = \text{FairMatching}(\Delta, M_\alpha)$. We use parameters $\Delta = 0.1n \cdot u(M_E)$, with $n = 0, 1, 2, \ldots, 10$.

Throughout this procedure, we calculate the utility of the efficient matching (u_E) and the fair matching (u_F) for each UNOS graph, and for each fairness rule—with parameters $\alpha \in [0, 1]$, $\gamma \in [0, 20]$, and $\Delta \in [0, u(M_E)]$.

There are two important outcomes of each fairness rule: Price of Fairness (PoF), and fraction of the fair score ($\%F$). To calculate PoF we use the definition in Equation (1), using u_E and u_F. We define $\%F$ as the fraction of the maximum highly sensitized utility, achieved by $M_{\{\alpha,\gamma,\Delta\}}$, defined as

$$\%F(M_{\{\alpha,\gamma,\Delta\}}, M_F) = u_H(M_{\{\alpha,\gamma,\Delta\}})/u_H(M_F).$$

PoF and $\%F$ indicate the efficiency loss and the fairness of each rule, respectively.

[3] https://github.com/duncanmcelfresh/FairKidneyExchange

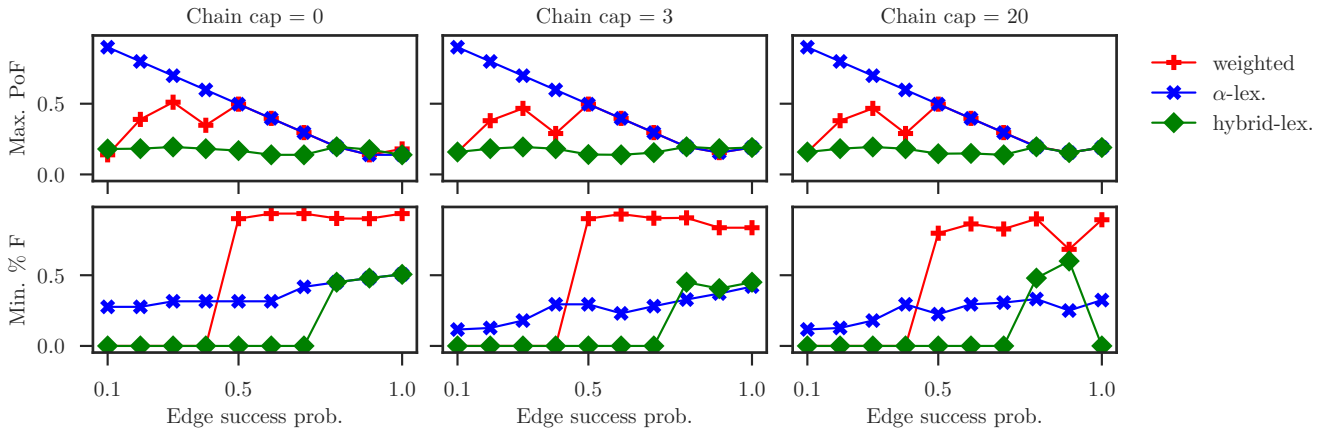


Figure 3: Worst-case price of fairness and $\%F$ for various edge success probabilities, and fairness parameters $\alpha = 0.1$, $\gamma = 0.1$, $\Delta = 0.1u(M_E)$.

5.2 Results and Discussion

Each fairness rule offers a parameter that balances efficiency and fairness. Two of these rules guarantee a certain outcome: α-lexicographic guarantees fairness, but allows high efficiency loss, while hybrid-lexicographic bounds overall efficiency loss. Weighted fairness makes no guarantees.

The price of fairness can be high in real exchanges, especially when edge success probability p is small. In failure-aware kidney exchange, cycles and chains of length k receive utility proportional to p^k. Fair matchings often use longer cycles and chains than the efficient matching, in order to reach highly sensitized patients; this leads to a high price of fairness when p is small.

Even when α and γ are small, there are cases when both α-lexicographic and weighted fairness allow for a high PoF. This becomes worse with lower edge probability. Figure 3 shows the worst-case PoF and $\%F$ for each rule, for the smallest parameters tested, for a range of edge success probabilities. Appendix C contains results for all parameter values tested.

Hybrid-lexicographic fairness limits PoF within the guaranteed bound of 0.2; this comes at the cost of a low $\%F$—when edge success probability is small, hybrid-lexicographic fairness awards zero fair utility in the worst case. α-lexicographic fairness produces the opposite behavior: $\%F$ is always larger than the guaranteed bound of 0.1, but the worst-case price of fairness grows steadily as edge probability decreases.

Theory suggests that the price of fairness is small on denser random graphs (see Section 2). We empirically confirm this theoretical finding by calculating the worst-case price of fairness and $\%F$ for random graphs of various sizes generated from real data; these results are given in Section C. In this case—when the price of fairness is small—α-lexicographic fairness may be appropriate, as overall efficiency loss is not severe.

Both α-lexicographic and hybrid-lexicographic fairness are useful, depending on the desired outcome. Policymakers may choose between these rules, and set the parameters α and Δ to guarantee either a minimum $\%F$ or a maximum price of fairness.

6 Conclusion

We addressed the classical problem of balancing fairness and efficiency in resource allocation, with a specific focus on the kidney exchange application area. Extending work by Ashlagi and Roth (2014) and Dickerson *et al.* (2014), we show that the theoretical price of fairness is small on a random graph model of kidney exchange, when both cycles and chains are used. However this model is too optimistic—real kidney exchanges are less certain and more sparse, and in reality the price of fairness can be unacceptably high.

Drawing on work by Hooker and Williams (2012), which is not applicable to kidney exchange, we provided the first approach to incorporating fairness into kidney exchange in a way that prioritizes marginalized participants, but also comes with acceptable worst-case guarantees on overall efficiency loss. Furthermore, our method is easily applied as an objective in the mathematical-programming-based clearing methods used in today's fielded exchanges. Using data from a large fielded kidney exchange, we showed that our method bounds efficiency loss while also prioritizing marginalized participants when possible.

Moving forward, it would be of theoretical and practical interest to address fairness in a realistic *dynamic* model of a matching market like kidney exchange (Anshelevich *et al.* 2013; Akbarpour *et al.* 2014; Anderson *et al.* 2015; Dickerson and Sandholm 2015). For example, how does prioritizing a class of patients in the present affect their, or other groups', long-term welfare? Similarly, exploring the effect on long-term efficiency of the single-shot Δ we use in this paper would be of practical importance; to start, Δ can be viewed as a hyperparameter to be tuned (Thornton *et al.* 2013).

References

David Abraham, Avrim Blum, and Tuomas Sandholm. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 295–304, 2007.

Mohammad Akbarpour, Shengwu Li, and Shayan Oveis Gharan. Dynamic matching market design. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, page 355, 2014.

Ross Anderson, Itai Ashlagi, David Gamarnik, and Yash Kanoria. A dynamic model of barter exchange. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1925–1933, 2015.

Elliot Anshelevich, Meenal Chhabra, Sanmay Das, and Matthew Gerrior. On the social welfare of mechanisms for repeated batch matching. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 60–66, 2013.

Itai Ashlagi and Alvin E Roth. Free riding and participation in large scale, multi-hospital kidney exchange. *Theoretical Economics*, 9(3):817–863, 2014.

Itai Ashlagi, Patrick Jaillet, and Vahideh H. Manshadi. Kidney exchange in dynamic sparse heterogenous pools. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 25–26, 2013.

Haris Aziz, Aris Filos-Ratsikas, Jiashu Chen, Simon Mackenzie, and Nicholas Mattei. Egalitarianism of random assignment mechanisms. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2016.

Dimitris Bertsimas, Vivek F Farias, and Nikolaos Trichakis. The price of fairness. *Operations Research*, 59(1):17–31, 2011.

Ioannis Caragiannis, Christos Kaklamanis, Panagiotis Kanellopoulos, and Maria Kyropoulou. The efficiency of fair division. International Workshop on Internet and Network Economics (WINE), 2009.

John P. Dickerson and Tuomas Sandholm. FutureMatch: Combining human value judgments and machine learning to match in dynamic environments. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 622–628, 2015.

John P. Dickerson, Ariel D. Procaccia, and Tuomas Sandholm. Optimizing kidney exchange with transplant chains: Theory and reality. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 711–718, 2012.

John P. Dickerson, Ariel D. Procaccia, and Tuomas Sandholm. Failure-aware kidney exchange. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, pages 323–340, 2013.

John P. Dickerson, Ariel D. Procaccia, and Tuomas Sandholm. Price of fairness in kidney exchange. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1013–1020, 2014.

John P. Dickerson, David Manlove, Benjamin Plaut, Tuomas Sandholm, and James Trimble. Position-indexed formulations for kidney exchange. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2016.

Haluk Ergin, Tayfun Sönmez, and M Utku Ünver. Multi-donor organ exchange, 2017. Working paper.

Wenyi Fang, Aris Filos-Ratsikas, Søren Kristoffer Stiil Frederiksen, Pingzhong Tang, and Song Zuo. Randomized assignments for barter exchanges: Fairness vs. efficiency. In *International Conference on Algorithmic Decision Theory (ADT)*, 2015.

Gabriele Farina, John P. Dickerson, and Tuomas Sandholm. Operation frames and clubs in kidney exchange. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

A. Hart, J. M. Smith, M. A. Skeans, S. K. Gustafson, D. E. Stewart, W. S. Cherikh, J. L. Wainright, G. Boyle, J. J. Snyder, B. L. Kasiske, and A. K. Israni. Kidney. *American Journal of Transplantation (Special Issue: OPTN/SRTR Annual Data Report 2014)*, 16, Issue Supplement S2:11–46, 2016.

John N Hooker and H Paul Williams. Combining equity and utilitarianism in a mathematical programming model. *Management Science*, 58(9):1682–1693, 2012.

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

Jian Li, Yicheng Liu, Lingxiao Huang, and Pingzhong Tang. Egalitarian pairwise kidney exchange: Fast algorithms via linear programming and parametric flow. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 445–452, 2014.

Yicheng Liu, Pingzhong Tang, and Wenyi Fang. Internally stable matchings and exchanges. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 1433–1439, 2014.

Nicholas Mattei, Abdallah Saffidine, and Toby Walsh. Mechanisms for online organ matching. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

Robert Montgomery, Sommer Gentry, William H Marks, Daniel S Warren, Janet Hiller, Julie Houp, Andrea A Zachary, J Keith Melancon, Warren R Maley, Hamid Rabb, Christopher Simpkins, and Dorry L Segev. Domino paired kidney donation: a strategy to make best use of live non-directed donation. *The Lancet*, 368(9533):419–421, 2006.

Brendon L Neuen, Georgina E Taylor, Alessandro R Demaio, and Vlado Perkovic. Global kidney disease. *The Lancet*, 382(9900):1243, 2013.

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

Michael Rees, Jonathan Kopke, Ronald Pelletier, Dorry Segev, Matthew Rutter, Alfredo Fabrega, Jeffrey Rogers, Oleh Pankewycz, Janet Hiller, Alvin Roth, Tuomas Sandholm, Utku Ünver, and Robert Montgomery. A nonsimultaneous, extended, altruistic-donor chain. *New England Journal of Medicine*, 360(11):1096–1101, 2009.

Alvin Roth, Tayfun Sönmez, and Utku Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119(2):457–488, 2004.

Alvin Roth, Tayfun Sönmez, and Utku Ünver. A kidney exchange clearinghouse in New England. *American Economic Review*, 95(2):376–380, 2005.

Alvin Roth, Tayfun Sönmez, and Utku Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125(2):151–188, 2005.

Chris Thornton, Frank Hutter, Holger H Hoos, and Kevin Leyton-Brown. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. In *International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 847–855. ACM, 2013.

Özgür Yılmaz. Kidney exchange: An egalitarian mechanism. *Journal of Economic Theory*, 146(2):592–618, 2011.

A Price of Fairness in the Random Graph Model

Ashlagi and Roth (2014) characterize efficient matchings in a random graph model without chains, and Dickerson *et al.* (2014) build on this to show that the price of fairness without chains is bounded above by $2/33$. Dickerson *et al.* (2012) extend the efficient matching of Ashlagi and Roth (2014) to include chains, but do not calculate the price of fairness. We close the remaining theory gap regarding the price of fairness with chains. Appendix A.1 describes the random graph model, and Appendix A.2 presents the theoretical price of fairness with chains.

A.1 Random Graph Model

Let all patient-donor pairs P be partitioned into subsets $V^{X\text{-}Y}$ for each patient blood type X and donor blood type Y. These subsets will be further partitioned into lowly- and highly sensitized pairs $V_L^{X\text{-}Y}$ and $V_H^{X\text{-}Y}$. Let μ_X be the fraction of both patients and donors of each blood type X.

Let N^X be the set of NDDs of blood type X. Let $\beta|P|$ be the total number of NDDs, with the same blood type distribution as patients. That is, $|N^X| = \beta\mu_X|P|$, with $X \in \{A, B, AB, O\}$.

Patient-donor vertices may be blood-type compatible, but will not be connected by a directed edge due to tissue-type incompatibility. Let $\bar{p}$ be the fraction of patient-donor pairs that are blood-type-compatible, but tissue-type-incompatible.

We refer to certain blood-type vertex subsets of as follows:

1. $V^{\text{A-B}}$ and $V^{\text{B-A}}$: reciprocal pairs
2. $V^{X\text{-}X}$: self-demanded pairs
3. $V^{\text{AB-B}}, V^{\text{AB-A}}, V^{\text{AB-O}}, V^{\text{A-O}}, V^{\text{B-O}}$: over-demanded pairs
4. $V^{\text{A-AB}}, V^{\text{B-AB}}, V^{\text{O-A}}, V^{\text{O-B}}, V^{\text{O-AB}}$: under-demanded pairs

To reflect real-world exchanges, assume $\bar{p} > 1 - \lambda$, $\mu_{\text{O}} > \mu_{\text{A}} > \mu_{\text{B}} > \mu_{\text{AB}}$, and $\bar{p} < 2/5$. WLOG, let $|V^{\text{A-B}}| > |V^{\text{B-A}}|$, and assume that the absolute difference between these pools grows sublinearly with the size of the exchange, that is $|V^{\text{A-B}}| - |V^{\text{B-A}}| = o(n)$.

A.2 The Price of Fairness With Chains

We calculate the price of fairness in this model by exploring all of the possible ways that the efficient matching can proceed, which depends on β. We state without proof that there are only four possible matchings with nonzero price of fairness, and several matchings with zero price of fairness. It is tedious, but straightforward, to confirm this statement, using the assumptions made while constructing these matchings. Figure 4 shows each possible matching on this model, and some of the impossible matchings.

Propositions 2, 3, 4, and 5 give the price of fairness for each of the four matchings with nonzero price of fairness; for each of these cases, $\beta < \mu_{\text{AB}}(1-\bar{p})$. Proposition 1 states that the price of fairness is zero when $\beta > \mu_{\text{AB}}(1-\bar{p})$.

In all of these matchings, the price of fairness is bounded above by the price of fairness without NDDs, found by Dickerson *et al.* (2014); Theorem 1 states this finding, which uses by Lemmas 2 and 3.

Theorem 2 states that the price of fairness is zero when $\beta > 1/8$, and Lemmas 4, 5, 6, and 7 give bounds on β for each matching with nonzero price of fairness.

We start with the efficient matching proposed in (Dickerson *et al.* 2012) using cycles and chains up to length 3. This matching may proceed in many different ways, depending on β. However, most outcomes are impossible based on the canonical assumptions for the random graph model. Figure 4 shows all possible ways that the matching can proceed.

Lemma 1 states that even without chains, all highly-sensitized patients except for those in $V^{\text{O-AB}}$ are matched in the efficient matching, only using cycles; this Lemma will be used in all following propositions.

Lemma 1. *Denote by $\mathcal{M}$ the set of matchings in $G(n)$ using cycles and chains up to length 3. As $n \to \infty$, a.s. all highly sensitized pairs can be matched with no efficiency loss under the lexicographic fairness rule, except for those of type O-AB.*

(This Lemma uses the same efficient matching introduced by Dickerson (Dickerson et al. *2012).)*

sketch. Begin with the efficient matching M^* using only cycles up to length 3, proposed by Dickerson in (Dickerson *et al.* 2014). M^* matches all over-demanded and self-demanded vertices with high probability, but leaves some under-demanded vertices unmatched. We proceed through the initial steps of matching M^* to show that *all* vertices in $V_H^{\text{O-A}}$, $V_H^{\text{O-B}}$, $V_H^{\text{A-AB}}$, and $V_H^{\text{B-AB}}$ are matched.

1. Match all vertices in $V^{\text{B-A}}$ in 2-cycles with $V^{\text{A-B}}$, exhausting $V^{\text{B-A}}$ and leaving $|V^{\text{A-B}}| \propto o(n)$.
2. Match all remaining vertices in $V^{\text{A-B}}$ in 3-cycles with $V^{\text{B-O}}$ and $V^{\text{O-A}}$. There are only $|V^{\text{A-B}}| \propto o(n)$ of these cycles, which will become negligible to the price of fairness as $n \to \infty$.
3. Match all remaining vertices in $V^{\text{A-O}}$ in 2-cycles with $V^{\text{O-A}}$. Note that $|V^{\text{A-O}}| \propto \bar{p}\mu_A\mu_O$ and $|V^{\text{O-A}}| \propto \mu_A\mu_O$. The $V^{\text{A-O}}$ vertices are exhausted first if $|V^{\text{A-O}}| < |V^{\text{O-A}}|$, which holds almost surely because $\bar{p}\mu_A\mu_O < \mu_A\mu_O$ due to the assumption $\bar{p} < 2/5$. All highly sensitized vertices $V_H^{\text{O-A}}$ are matched because $(1-\lambda)\mu_A\mu_O < \bar{p}\mu_A\mu_O$ holds under the assumption $1 - \lambda < \bar{p}$. Thus both $V^{\text{A-O}}$ and $V_H^{\text{O-A}}$ are exhausted, and $|V^{\text{O-A}}| \propto \mu_A\mu_O(1-\bar{p})$.
4. Match all remaining vertices in $V^{\text{B-O}}$ in 2-cycles with $V^{\text{O-B}}$. Note that $|V^{\text{B-O}}| \propto \bar{p}\mu_B\mu_O$ and $|V^{\text{O-B}}| \propto \mu_B\mu_O$. As before, the a.s. $|V^{\text{O-B}}| > |V^{\text{B-O}}|$. All highly sensitized vertices $V_H^{\text{O-B}}$ are matched a.s., because $\bar{p}\mu_B\mu_O > (1-\lambda)\mu_B\mu_O$ holds under the assumption $\bar{p} > 1-\lambda$. Thus both $V^{\text{B-O}}$ and $V_H^{\text{O-B}}$ are exhausted, and $|V^{\text{O-B}}| \propto \mu_B\mu_O(1-\bar{p})$.
5. Match all vertices in $V^{\text{AB-A}}$ in 2-cycles with $V^{\text{A-AB}}$. Note that, $|V^{\text{AB-A}}| \propto \bar{p}\mu_A\mu_{\text{AB}}$ and $|V^{\text{A-AB}}| \propto \mu_A\mu_{\text{AB}}$. As before, a.s. $|V^{\text{A-AB}}| > |V^{\text{A-AB}}|$. All highly sensitized vertices $V_H^{\text{A-AB}}$ are matched, because $\bar{p}\mu_A\mu_{\text{AB}} >$

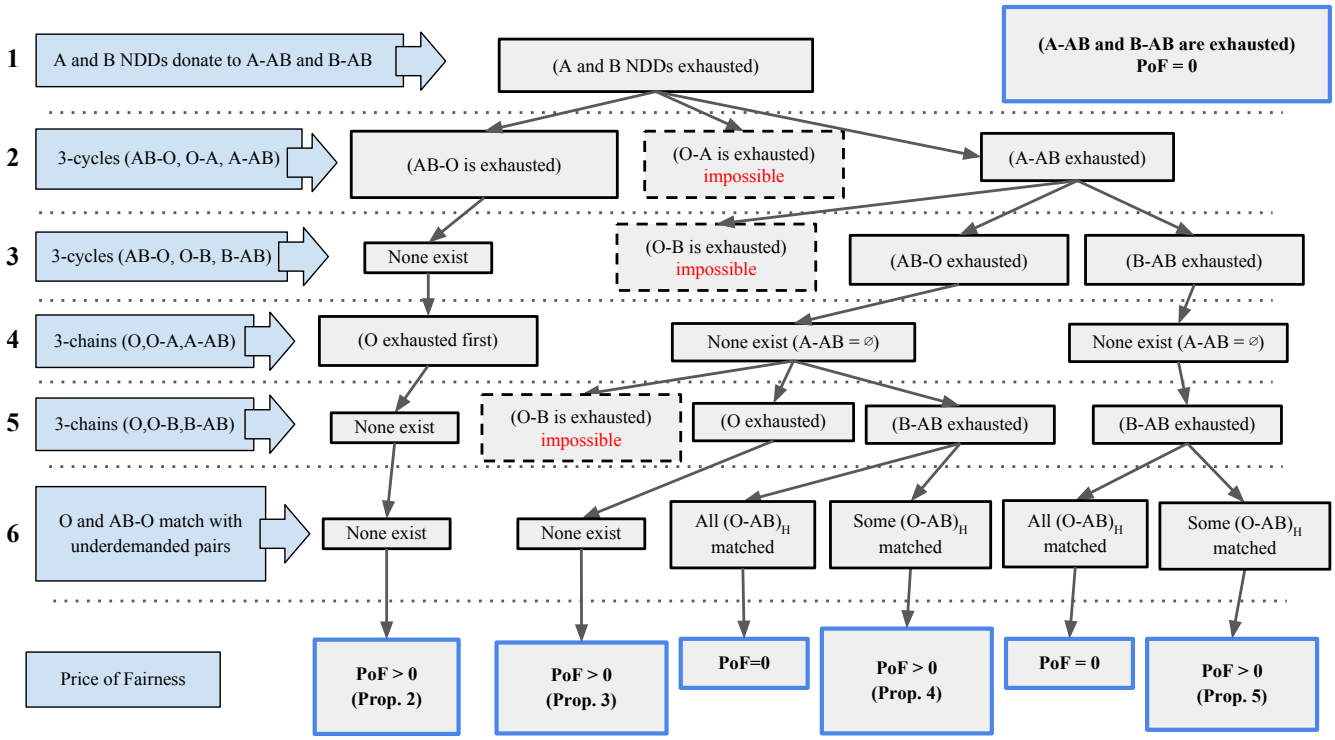


Figure 4: All possible matchings on the random graph model. Boxes with blue borders represent the matching outcomes, and boxes with black borders represent intermediate steps in each matching. Some of the impossible matchings are shown as boxes with dashed black borders.

$(1-\lambda)\mu_A\mu_{\text{AB}}$ under the assumption $\bar{p} > 1-\lambda$. Thus both $V^{\text{AB-A}}$ and $V_H^{\text{A-AB}}$ are exhausted, and $|V^{\text{A-AB}}| \propto \mu_A\mu_O(1-\bar{p})$.

6. Match all vertices in $V^{\text{AB-B}}$ in 2-cycles with $V^{\text{B-AB}}$. Note that, $|V^{\text{AB-B}}| \propto \bar{p}\mu_B\mu_{\text{AB}}$ and $|V^{\text{B-AB}}| \propto \mu_B\mu_{\text{AB}}$, and a.s. $|V^{\text{B-AB}}| > |V^{\text{AB-B}}|$. All highly sensitized vertices $V_H^{\text{B-AB}}$ are matched, because $\bar{p}\mu_B\mu_{\text{AB}} > (1-\lambda)\mu_B\mu_{\text{AB}}$ under the assumption $\bar{p} > 1-\lambda$. Thus both $V^{\text{AB-B}}$ and $V_H^{\text{B-AB}}$ are exhausted, and $|V^{\text{B-AB}}| \propto \mu_B\mu_O(1-\bar{p})$.

Thus, these initial steps of matching M^* exhaust all highly sensitized pairs in $V_H^{\text{O-A}}$, $V_H^{\text{O-B}}$, $V_H^{\text{A-AB}}$, and $V_H^{\text{B-AB}}$. □

With uniform edge weights, lexicographic fairness requires that we match the maximum possible number of highly sensitized vertices. Lemma 1 states that the efficient matching M^* includes all highly sensitized patients, except for those in $V^{\text{O-AB}}$. Therefore all efficiency loss–and price of fairness–is caused by matching vertices in $V_H^{\text{O-AB}}$.

Using both chains and cycles increases overall efficiency. In the dense graph model used in this Appendix, adding chains can only decrease the price of fairness.

Proposition 1 in (Dickerson *et al.* 2014) states that with only cycles up to length 3, and assuming $\bar{p} > 1-\lambda$, and $\mu_{\text{O}} < 3\mu_{\text{A}}/2$, and $\mu_{\text{O}} > \mu_{\text{A}} > \mu_{\text{B}} > \mu_{\text{AB}}$, the price of fairness is at most $\frac{2}{33}$. In the dense graph model used here, adding chains tightens this upper bound.

The following propositions tighten the upper bound on the price of fairness, for every possible value of β.

Proposition 1. *Assume*

1 $\beta > (1-\bar{p})\mu_{AB}$.

Denote by $\mathcal{M}$ the set of matchings in $G(n)$ using cycles and chains up to length 3. As $n \to \infty$, almost surely $\mathit{POF}(\mathcal{M}, u_{LEX}) = 0$.

sketch. We begin by executing the initial steps of matching M^* as done in the proof of Lemma 1, matching all highly sensitized vertices except for those in $V_H^{\text{O-AB}}$. The following steps continue the matching M^* from Lemma 1.

7. A- and B-type NDDs donate to $V^{\text{A-AB}}$ and $V^{\text{B-AB}}$, respectively. Note that $|N^A| \propto \beta\mu_A$ and $|V^{\text{A-AB}}| \propto (1-\bar{p})\mu_A\mu_{\text{AB}}$. Assuming $\beta > (1-\bar{p})\mu_{\text{AB}}$, the inequality $\beta\mu_{\text{A}} > \mu_A\mu_{\text{AB}}(1-\bar{p})$ holds and a.s. $|N^{\text{A}}| > |V^{\text{A-AB}}|$. By the same argument, a.s. $|N^{\text{B}}| > |V^{\text{B-AB}}|$. Thus, both $V^{\text{A-AB}}$ and $V^{\text{B-AB}}$ are exhausted, and $|N^{\text{B}}| \propto \mu_{\text{B}}(\beta-(1-\bar{p})\mu_{\text{AB}})$ and $|N^{\text{A}}| \propto \mu_{\text{A}}(\beta-(1-\bar{p})\mu_{\text{AB}})$.
8. Create cycles of the form (AB-O, O-X, X-AB), with X $\in \{\text{A},\text{B}\}$. None of these cycles occur because both $V^{\text{A-AB}}$ and $V^{\text{B-AB}}$ have been exhausted in previous steps.
9. Create chains of the form (O,O-X,X-AB), with X $\in \{\text{A},\text{B}\}$. None of these cycles occur, because both $V^{\text{A-AB}}$ and $V^{\text{B-AB}}$ have been exhausted in previous steps.

10. Remaining O-type NDDs donate to remaining under-demanded vertices, beginning with $V^{\text{O-AB}}$. Note that no O-type NDDs have been used in previous steps, so $|N^{\text{O}}| \propto \beta\mu_{\text{O}}$.
11. 2-cycles are created with $V^{\text{AB-O}}$ and remaining under-demanded vertices, beginning with $V^{\text{O-AB}}$. Note that no vertices in $V^{\text{AB-O}}$ have been used in previous steps, so $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{O}}\mu_{\text{AB}}$.

The final two steps match up to $|V^{\text{AB-O}}| + |N^{\text{O}}| \propto \beta\mu_{\text{O}} + \bar{p}\mu_{\text{O}}\mu_{\text{AB}}$ vertices in $V^{\text{O-AB}}$. The only remaining highly-sensitized vertices are in $V_H^{|}OAB \propto (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}$. Assuming that $\bar{p} > 1-\lambda$, the inequality $\beta\mu_{\text{O}} + \bar{p}\mu_{\text{O}}\mu_{\text{AB}} > \bar{p}\mu_{\text{O}}\mu_{\text{AB}} > (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}$ holds, and a.s. $|V^{\text{AB-O}}| + |N^{\text{O}}| > |V_H^{|}OAB|$. This exhausts all vertices in $|V_H^{\text{O-AB}}|$. All other highly-sensitized vertices were matched in steps 1-6 of, as in Lemma 1. Thus, all highly sensitized vertices can be matched with no efficiency loss, and the price of fairness is zero.

□

Proposition1 assumes that β is extremely large, specifically $\beta > 1/4 > (1-\bar{p})\mu_{\text{AB}}$. In practice, $\beta < 0.01$ – that is, the number of NDDs in an exchange is often less than 1% of the size of the exchange. The following Propositions address the price of fairness when $\beta < (1-\bar{p})\mu_{\text{AB}} < 1/4$.

Proposition 2. *Assume*

A.1 $\beta < \mu_A(1-\bar{p}) - \bar{p}\mu_{AB}$

A.2 $\beta < \mu_{AB}(1-\bar{p}) - \bar{p}\mu_{AB}\mu_O/\mu_A$

A.3 $\beta < \mu_{AB}\left(\frac{\mu_A}{\mu_A+\mu_O} - \bar{p}\right)$

These constraints imply $\beta \in [0, 1/8]$.Denote by $\mathcal{M}$ the set of matchings in $G(n)$ using cycles and chains up to length 3. Almost surely as $n \to \infty$, the price of fairness is

$$\textit{POF}(\mathcal{M}, u_{LEX}) = \frac{(1-\lambda)\mu_O\mu_{AB}}{u_E}$$

with

$$\begin{aligned} u_E = \bar{p}\Big[&2\mu_{AB}\mu_B + 2\mu_{AB}\mu_A + 3\mu_{AB}\mu_O \\ &+2\mu_A\mu_O + 2\mu_B\mu_O + \mu_O^2 + \mu_A^2 + \mu_B^2 + \mu_{AB}^2\Big] \\ &+2\mu_A\mu_B + \beta(\mu_A + \mu_B + 2\mu_O) \end{aligned}$$

sketch. We begin with matching M^* as done in the proof of Lemma 1, matching all highly sensitized vertices except for those in $V_H^{\text{AB-O}}$. We now complete the efficient matching using both 3-cycles and 3-chains as in (Dickerson *et al.* 2012).

7. A- and B-type NDDs donate to $V^{\text{A-AB}}$ and $V^{\text{B-AB}}$, respectively. Note that $|N^A| \propto \beta\mu_A$ and $|V^{\text{A-AB}}| \propto (1-\bar{p})\mu_A\mu_{\text{AB}}$. The inequality $\beta\mu_{\text{A}} < \mu_A\mu_{\text{AB}}(1-\bar{p})$ holds due to assumption **A.2**, and a.s. $|N^{\text{A}}| < |V^{\text{A-AB}}|$. By the same argument, a.s. $|N^{\text{B}}| < |V^{\text{B-AB}}|$. Thus, both N^{A} and N^{B} are exhausted, and $|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{A}}$ and $|V^{\text{B-AB}}| \propto \mu_{\text{B}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{B}}$.
8. Create cycles of the form (AB-O, O-A, A-AB). The current size of each vertex group is
   (1) $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$
   (2) $|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{O}}$
   (3) $|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{A}}$

   The inequality $(1) < (2)$ holds due to the model assumptions, so a.s. $|V^{\text{AB-O}}| < |V^{\text{O-A}}|$. Note that the inequality $(1) < (3)$ can be written as

   $$\beta < \mu_{\text{AB}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}/\mu_{\text{A}}$$

   which holds by assumptions **A.2**, and a.s. $|V^{\text{AB-O}}| < |V^{\text{A-AB}}|$. Executing these cycles exhausts $V^{\text{AB-O}}$ and leaves the following vertices remaining
   (1) $|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{O}} - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$.
   (2) $|V^{\text{A-AB}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{AB}} - \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \beta\mu_{\text{A}}$
9. Create cycles of the form (AB-O, O-B, B-AB). The previous step exhausted $V^{\text{AB-O}}$, so none of these cycles occur.
10. Create chains of the form (O,O-A,A-AB). The current size of each vertex group is
    (1) $|N^{\text{O}}| \propto \beta\mu_{\text{O}}$
    (2) $|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{O}} - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$
    (3) $|V^{\text{A-AB}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{AB}} - \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \beta\mu_{\text{A}}$

    The inequality $(1) < (2)$ holds due to assumption **A.1**, so a.s. $|N^{\text{O}}| < |V^{\text{O-A}}|$. Note that inequality $(1) < (3)$ can be written as

    $$\beta < \mu_{\text{AB}}\left(\frac{\mu_{\text{A}}}{\mu_{\text{A}}+\mu_{\text{O}}} - \bar{p}\right)$$

    which holds due to **A.3**. Thus a.s. $|N^{\text{O}}| < |V^{\text{A-AB}}|$, and $|N^{\text{O}}|$ is exhausted. The vertices unmatched by these chains are
    (1) $|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{O}} - \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \beta\mu_{\text{O}}$
    (2) $|V^{\text{A-AB}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{AB}} - \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \beta(\mu_{\text{A}} + \mu_{\text{O}})$
11. Remaining O-type NDDs donate to remaining under-demanded vertices. The previous step exhausted N^{O}, so none of these donations occur.
12. 2-cycles are created with $V^{\text{AB-O}}$ and remaining under-demanded vertices. The previous step exhausted $V^{\text{AB-O}}$, so none of these cycles occur.

In the efficient matching described above, the number of *matched* pairs in each under-demanded group is

$$\begin{aligned} |V^{\text{O-A}}| &\propto \mu_{\text{O}}(\beta + \bar{p}(\mu_{\text{A}} + \mu_{\text{AB}})) \\ |V^{\text{O-B}}| &\propto \bar{p}\mu_{\text{B}}\mu_{\text{O}} \\ |V^{\text{A-AB}}| &\propto (\beta + \bar{p}\mu_{\text{AB}})(\mu_{\text{A}} + \mu_{\text{O}}) \\ |V^{\text{B-AB}}| &\propto (\beta + \bar{p}\mu_{\text{AB}})\mu_{\text{B}} \\ |V^{\text{O-AB}}| &= 0 \end{aligned}$$

Combining these with the over-demanded and self-demanded vertices, the total size of the efficient matching is

$$u_E = \bar{p}\Big[2\mu_{\text{AB}}\mu_{\text{B}} + 2\mu_{\text{AB}}\mu_{\text{A}} + 3\mu_{\text{AB}}\mu_{\text{O}} + 2\mu_{\text{A}}\mu_{\text{O}} + 2\mu_{\text{B}}\mu_{\text{O}} + \mu_{\text{O}}^2 + \mu_{\text{A}}^2 + \mu_{\text{B}}^2 + \mu_{\text{AB}}^2\Big] + 2\mu_{\text{A}}\mu_{\text{B}} + \beta\left(\mu_{\text{A}} + \mu_{\text{B}} + 2\mu_{\text{O}}\right)$$

This efficient matching includes all highly sensitized vertices except for those in $V_H^{\text{O-AB}}$. To calculate the price of fairness we now find the size of the fair matching. We match each vertex in $V_H^{\text{O-AB}}$ by removing a 3-cycle of the form (AB-O, O-A, A-AB) and creating a 2-cycle (AB-O, O-AB). This matching used $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{O}}\mu_{\text{AB}}$ 3-cycles of this form, while $|V_H^{\text{O-AB}}| \propto (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}$. The model assumption $\bar{p} > 1 - \lambda$ ensures that $|V^{\text{AB-O}}| > |V_H^{\text{O-AB}}|$, and all vertices in $V_H^{\text{O-AB}}$ can be matched in this way.

To match each vertex in $V_H^{\text{O-AB}}$, we remove from the matching one vertex from both $V^{\text{O-A}}$ and $V^{\text{A-AB}}$. Thus the total efficiency loss is $|V_H^{\text{O-AB}}| \propto (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}$. The price of fairness is

$$\mathsf{POF}(\mathcal{M}, u_{LEX}) = \frac{(1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}}{u_E}$$

With u_E defined previously. □

Proposition 3. *Assume*

1 $\beta < \mu_{AB}(1-\bar{p}) - \mu_{AB}\mu_O\bar{p}/(\mu_A + \mu_B)$

2 $\beta < \frac{\mu_A\mu_{AB}(1-\bar{p}) + \mu_B\mu_O(1-\bar{p}) - \bar{p}\mu_O\mu_{AB}}{\mu_A + \mu_O}$

3 $\beta > \mu_{AB}(1-\bar{p}) - \mu_{AB}\mu_O\bar{p}/\mu_A$

4 $\beta < \mu_{AB}(1-\bar{p}) - \bar{p}\mu_{AB}\mu_O/\mu_A + (1-\bar{p})\mu_B\mu_O/\mu_A$

5 $\beta < \mu_{AB}(1-\bar{p}) - \mu_O\mu_{AB}/(1-\mu_{AB})$

*Note that as written, constraint **4** is a looser bound than **5**, and can be removed. However it is convenient to leave **4** for clarity. These constraints imply $\beta \in [0, 1/12]$. Denote by $\mathcal{M}$ the set of matchings in $G(n)$ using cycles and chains up to length 3. Almost surely as $n \to \infty$, the price of fairness is*

$$\mathsf{POF}(\mathcal{M}, u_{LEX}) = \frac{(1-\lambda)\mu_O\mu_{AB}}{u_E}$$

with

$$u_E = \bar{p}\Big[2\mu_{AB}\mu_B + 2\mu_{AB}\mu_A + 3\mu_{AB}\mu_O + 2\mu_A\mu_O + 2\mu_B\mu_O + \mu_O^2 + \mu_A^2 + \mu_B^2 + \mu_{AB}^2\Big] + 2\mu_A\mu_B + \beta\left(\mu_A + \mu_B + 2\mu_O\right)$$

sketch. We begin with matching M^* as done in the proof of Lemma 1, matching all highly sensitized vertices except for those in $V_H^{\text{AB-O}}$. We now complete the efficient matching using both 3-cycles and 3-chains as in (Dickerson *et al.* 2012).

7. A- and B-type NDDs donate to $V^{\text{A-AB}}$ and $V^{\text{B-AB}}$, respectively. Note that $|N^A| \propto \beta\mu_A$ and $|V^{\text{A-AB}}| \propto (1-\bar{p})\mu_A\mu_{\text{AB}}$. The inequality $\beta\mu_{\text{A}} < \mu_A\mu_{\text{AB}}(1-\bar{p})$ holds due to assumption **A.1**, and a.s. $|N^{\text{A}}| < |V^{\text{A-AB}}|$. By the same argument, a.s. $|N^{\text{B}}| < |V^{\text{B-AB}}|$. Thus, both N^{A} and N^{B} are exhausted, and $|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{A}}$ and $|V^{\text{B-AB}}| \propto \mu_{\text{B}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{B}}$.
8. Create cycles of the form (AB-O, O-A, A-AB). The current size of each vertex group is
   (1) $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$
   (2) $|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{O}}$
   (3) $|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{A}}$

   Note that the inequality $(3) < (1)$ can be written as

   $$\beta > \mu_{\text{AB}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}/\mu_{\text{A}}$$

   which holds by assumption **3**, and a.s. $|V^{\text{A-AB}}| < |V^{\text{AB-O}}|$. The inequality $(3) < (2)$ can be written as

   $$\beta > (1-\bar{p})(\mu_{\text{AB}} - \mu_{\text{O}})$$

   which holds by model assumptions, and a.s. $|V^{\text{A-AB}}| < |V^{\text{O-A}}|$. Executing these cycles exhausts $V^{\text{A-AB}}$ and leaves the following vertices remaining
   $|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}(\mu_{\text{O}} - \mu_{\text{AB}}) + \mu_{\text{A}}\beta$
   $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + \mu_{\text{A}}\beta$
9. Create cycles of the form (AB-O, O-B, B-AB). The current size of each vertex group is
   (1) $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + \mu_{\text{A}}\beta$
   (2) $|V^{\text{O-B}}| \propto (1-\bar{p})\mu_{\text{B}}\mu_{\text{O}}$
   (3) $|V^{\text{B-AB}}| \propto \mu_{\text{B}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{B}}$

   Inequality $(1) < (2)$ can be written as

   $$\beta < \mu_{\text{AB}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}/\mu_{\text{A}} + (1-\bar{p})\mu_{\text{B}}\mu_{\text{O}}/\mu_A$$

   which holds by assumption **4**. Inequality $(1) < (3)$ can be written as

   $$\beta < \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/(\mu_{\text{A}} + \mu_{\text{B}})$$

   which holds by assumption **1**.
   Executing these cycles exhausts $V^{\text{AB-O}}$ and leaves the following vertices remaining
   $|V^{\text{O-B}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + (\mu_{\text{B}} - \bar{p}(\mu_{\text{AB}} + \mu_{\text{B}}))\mu_{\text{O}} - \beta\mu_{\text{A}}$
   $|V^{\text{B-AB}}| \propto ((1-\bar{p})\mu_{\text{AB}} - \beta)(\mu_{\text{A}} + \mu_{\text{B}}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$
10. Create chains of the form (O,O-A,A-AB). Previous steps exhausted $V^{\text{A-AB}}$ so none of these chains occur.
11. Create chains of the form (O,O-B,B-AB). The current size of each vertex group is
    (1) $|N^{\text{O}}| \propto \beta\mu_{\text{O}}$
    (2) $|V^{\text{O-B}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + (\mu_{\text{B}} - \bar{p}(\mu_{\text{AB}} + \mu_{\text{B}}))\mu_{\text{O}} - \beta\mu_{\text{A}}$
    (3) $|V^{\text{B-AB}}| \propto ((1-\bar{p})\mu_{\text{AB}} - \beta)(\mu_{\text{A}} + \mu_{\text{B}}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$

The inequality $(1) < (2)$ can be written as

$$\beta < \frac{\mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + \mu_{\text{B}}\mu_{\text{O}}(1-\bar{p}) - \bar{p}\mu_{\text{O}}\mu_{\text{AB}}}{\mu_{\text{A}} + \mu_{\text{O}}}$$

which holds by assumption **2**. The inequality $(1) < (3)$ can be written as

$$\beta < \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{O}}\mu_{\text{AB}}/(1-\mu_{\text{AB}})$$

which holds by assumption **5**. Executing these chains exhausts N^{O} and leaves the following vertices remaining

$$|V^{\text{O-B}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + (\mu_{\text{B}} - \bar{p}(\mu_{\text{AB}} + \mu_{\text{B}}))\mu_{\text{O}} - \beta(\mu_{\text{A}} + \mu_{\text{O}})$$

$$|V^{\text{B-AB}}| \propto (1-\bar{p})\mu_{\text{AB}} - \beta)(\mu_{\text{A}} + \mu_{\text{B}}) - (\beta + \bar{p}\mu_{\text{AB}})\mu_{\text{O}}$$

12. Remaining O-type NDDs donate to remaining under-demanded vertices. The previous step exhausted N^{O}, so none of these donations occur.
13. 2-cycles are created with $V^{\text{AB-O}}$ and remaining under-demanded vertices. The previous steps exhausted $V^{\text{AB-O}}$, so none of these cycles occur.

In the efficient matching described above, the number of *matched* pairs in each under-demanded group is

$$|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}(\mu_{\text{O}} - \mu_{\text{AB}}) + \mu_{\text{A}}\beta$$

$$|V^{\text{O-B}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + (\mu_{\text{B}} - \bar{p}(\mu_{\text{AB}} + \mu_{\text{B}}))\mu_{\text{O}} - \beta(\mu_{\text{A}} + \mu_{\text{O}})$$

$$|V^{\text{A-AB}}| = 0$$
$$|V^{\text{B-AB}}| \propto (1-\bar{p})\mu_{\text{AB}} - \beta)(\mu_{\text{A}} + \mu_{\text{B}}) - (\beta + \bar{p}\mu_{\text{AB}})\mu_{\text{O}}$$
$$|V^{\text{O-AB}}| = 0$$

Combining these with the over-demanded and self-demanded vertices, the total size of the efficient matching is

$$u_E = \bar{p}\Big[2\mu_{\text{AB}}\mu_{\text{B}} + 2\mu_{\text{AB}}\mu_{\text{A}} + 3\mu_{\text{AB}}\mu_{\text{O}} + 2\mu_{\text{A}}\mu_{\text{O}} + 2\mu_{\text{B}}\mu_{\text{O}} + \mu_{\text{O}}^2 + \mu_{\text{A}}^2 + \mu_{\text{B}}^2 + \mu_{\text{AB}}^2\Big] + 2\mu_{\text{A}}\mu_{\text{B}} + \beta\left(\mu_{\text{A}} + \mu_{\text{B}} + 2\mu_{\text{O}}\right)$$

This efficient matching includes all highly sensitized vertices except for those in $V_H^{\text{O-AB}}$. To calculate the price of fairness we now find the size of the fair matching. We match each vertex in $V_H^{\text{O-AB}}$ by removing a 3-cycle of the form (AB-O, O-A, A-AB) and creating a 2-cycle (AB-O, O-AB). This matching used $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{O}}\mu_{\text{AB}}$ 3-cycles of this form, while $|V_H^{\text{O-AB}}| \propto (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}$. The model assumptions ensure that $|V^{\text{AB-O}}| > |V_H^{\text{O-AB}}|$, and all vertices in $V_H^{\text{O-AB}}$ can be matched in this way.

To match each vertex in $V_H^{\text{O-AB}}$, we remove from the matching one vertex from both $V^{\text{O-A}}$ and $V^{\text{A-AB}}$. Thus the total efficiency loss is $|V_H^{\text{O-AB}}| \propto (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}$. The price of fairness is

$$\textsf{POF}(\mathcal{M}, u_{LEX}) = \frac{(1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}}{u_E}$$

With u_E defined previously. □

Proposition 4. *Assume*

1 $\beta > \mu_{AB}(1-\bar{p}) - \mu_{AB}\mu_O\bar{p}/\mu_A$
2 $\beta < \mu_{AB}(1-\bar{p}) - \mu_{AB}\mu_O\bar{p}/(\mu_A + \mu_B)$
3 $\beta < \mu_{AB}(1-\bar{p}) - \bar{p}\mu_{AB}\mu_O/\mu_A + (1-\bar{p})\mu_B\mu_O/\mu_A$
4 $\beta > \mu_{AB}\left((1-\bar{p}) - \frac{\mu_O}{1-\mu_{AB}}\right)$
5 $\beta < \mu_{AB}(1-\bar{p}) - \lambda\mu_O\frac{\mu_{AB}}{1-\mu_{AB}}$

These constraints imply $\beta \in [0, 1/8]$. Denote by $\mathcal{M}$ the set of matchings in $G(n)$ using cycles and chains up to length 3. Almost surely as $n \to \infty$, the price of fairness is

$$\textsf{POF}(\mathcal{M}, u_{LEX}) = \frac{(1-\mu_{AB})((1-\bar{p})\mu_{AB} - \beta) - \lambda\mu_{AB}\mu_O}{u_E}$$

with

$$u_E = \mu_{AB}\mu_B + \mu_A\left(\mu_{AB} + 2\mu_B\right) + \beta\mu_O + \bar{p}\big[\mu_A^2 + \mu_A\mu_{AB} + \mu_{AB}^2 + \mu_{AB}\mu_B + \mu_B^2 + 2(\mu_A + \mu_{AB} + \mu_B)\mu_O + \mu_O^2\big]$$

sketch. We begin with matching M^* as done in the proof of Lemma 1, matching all highly sensitized vertices except for those in $V_H^{\text{AB-O}}$. We now complete the efficient matching using both 3-cycles and 3-chains as in (Dickerson *et al.* 2012).

7. A- and B-type NDDs donate to $V^{\text{A-AB}}$ and $V^{\text{B-AB}}$, respectively. Note that $|N^A| \propto \beta\mu_A$ and $|V^{\text{A-AB}}| \propto (1-\bar{p})\mu_A\mu_{\text{AB}}$. The inequality $\beta\mu_{\text{A}} < \mu_A\mu_{\text{AB}}(1-\bar{p})$ holds due to assumption **2**, and a.s. $|N^{\text{A}}| < |V^{\text{A-AB}}|$. By the same argument, a.s. $|N^{\text{B}}| < |V^{\text{B-AB}}|$. Thus, both N^{A} and N^{B} are exhausted, and $|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{A}}$ and $|V^{\text{B-AB}}| \propto \mu_{\text{B}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{B}}$.
8. Create cycles of the form (AB-O, O-A, A-AB). The current size of each vertex group is
   (1) $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$
   (2) $|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{O}}$
   (3) $|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{A}}$

   Note that the inequality $(3) < (1)$ can be written as

   $$\beta > \mu_{\text{AB}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}/\mu_{\text{A}}$$

   which holds by assumption **1**, and a.s. $|V^{\text{A-AB}}| < |V^{\text{AB-O}}|$. The inequality $(3) < (2)$ can be written as

   $$\beta > (1-\bar{p})(\mu_{\text{AB}} - \mu_{\text{O}})$$

   which holds by model assumptions, and a.s. $|V^{\text{A-AB}}| < |V^{\text{O-A}}|$. Executing these cycles exhausts $V^{\text{A-AB}}$ and leaves the following vertices remaining

$$|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}(\mu_{\text{O}} - \mu_{\text{AB}}) + \mu_{\text{A}}\beta$$
$$|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + \mu_{\text{A}}\beta$$

9. Create cycles of the form (AB-O, O-B, B-AB). The current size of each vertex group is
   (1) $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + \mu_{\text{A}}\beta$
   (2) $|V^{\text{O-B}}| \propto (1-\bar{p})\mu_{\text{B}}\mu_{\text{O}}$
   (3) $|V^{\text{B-AB}}| \propto \mu_{\text{B}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{B}}$

   Inequality $(1) < (2)$ can be written as

   $$\beta < \mu_{\text{AB}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}/\mu_{\text{A}} + (1-\bar{p})\mu_{\text{B}}\mu_{\text{O}}/\mu_{A}$$

   which holds by assumption **3**. Inequality $(1) < (3)$ can be written as

   $$\beta < \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/(\mu_{\text{A}} + \mu_{\text{B}})$$

   which holds by assumption **2**.
   Executing these cycles exhausts $V^{\text{AB-O}}$ and leaves the following vertices remaining
   $$|V^{\text{O-B}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + (\mu_{\text{B}} - \bar{p}(\mu_{\text{AB}} + \mu_{\text{B}}))\mu_{\text{O}} - \beta\mu_{\text{A}}$$
   $$|V^{\text{B-AB}}| \propto ((1-\bar{p})\mu_{\text{AB}} - \beta)(\mu_{\text{A}} + \mu_{\text{B}}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$$
10. Create chains of the form (O,O-A,A-AB). Previous steps exhausted $V^{\text{A-AB}}$ so none of these chains occur.
11. Create chains of the form (O,O-B,B-AB). The current size of each vertex group is
   (1) $|N^{\text{O}}| \propto \beta\mu_{\text{O}}$
   (2) $|V^{\text{O-B}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + (\mu_{\text{B}} - \bar{p}(\mu_{\text{AB}} + \mu_{\text{B}}))\mu_{\text{O}} - \beta\mu_{\text{A}}$
   (3) $|V^{\text{B-AB}}| \propto ((1-\bar{p})\mu_{\text{AB}} - \beta)(\mu_{\text{A}} + \mu_{\text{B}}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$

   The inequality $(3) < (1)$ can be written as

   $$\beta > \mu_{\text{AB}}\left((1-\bar{p}) - \frac{\mu_{\text{O}}}{1-\mu_{\text{AB}}}\right)$$

   which holds by assumption **4**. The inequality $(3) < (2)$ can be written as

   $$\beta > (1-\bar{p})(\mu_{\text{AB}} - \mu_{\text{O}})$$

   which holds by the model assumptions. Executing these chains exhausts $V^{\text{B-AB}}$ and leaves the following vertices remaining
   $$|N^{\text{O}}| \propto (\beta + \bar{p}\mu_{\text{AB}})(\mu_{\text{A}} + \mu_{\text{B}} + \mu_{\text{O}}) - \mu_{\text{AB}}(\mu_{\text{A}} + \mu_{\text{B}})$$
   $$|V^{\text{O-B}}| \propto \mu_{\text{B}}((\beta + (1-\bar{p})(\mu_{\text{O}} - \mu_{\text{AB}}))$$
12. Remaining O-type NDDs and $V^{\text{AB-O}}$ vertices match with remaining under-demanded vertices, starting with $V^{\text{O-AB}}$. The remaining size of each vertex group is
   (1) $|N^{\text{O}}| \propto (\beta + \bar{p}\mu_{\text{AB}})(\mu_{\text{A}} + \mu_{\text{B}} + \mu_{\text{O}}) - \mu_{\text{AB}}(\mu_{\text{A}} + \mu_{\text{B}})$
   (2) $|V^{\text{O-AB}}| \propto \mu_{\text{AB}}\mu_{\text{O}}$
   (3) $|V^{\text{AB-O}}| = 0$

   After simplifying, the inequality $(1) < (2)$ can be written as

   $$\beta < \mu_{\text{AB}}(1-\bar{p})$$

   which holds by assumption **2**. Thus O-type NDDs are exhausted first, leaving some vertices remaining in $V^{\text{O-AB}}$, with

   $$|V^{\text{O-AB}}| \propto ((1-\bar{p})\mu_{\text{AB}} - \beta)(1-\mu_{\text{AB}})$$

In the efficient matching described above, the number of *matched* pairs in each under-demanded group is

$$|V^{\text{O-A}}| \propto \mu_{\text{A}}(\mu_{\text{AB}}(1-\bar{p}) + \bar{p}\mu_{\text{O}} - \beta)$$
$$|V^{\text{O-B}}| \propto \mu_{\text{B}}(\mu_{\text{AB}}(1-\bar{p}) + \bar{p}\mu_{\text{O}} - \beta)$$
$$|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}}$$
$$|V^{\text{B-AB}}| \propto \mu_{\text{B}}\mu_{\text{AB}}$$
$$|V^{\text{O-AB}}| \propto (\beta + \mu_{\text{AB}}\bar{p})(1-\mu_{\text{AB}}) - \mu_{\text{AB}}(\mu_{\text{A}} + \mu_{\text{B}})$$

Combining these with the over-demanded and self-demanded vertices, the total size of the efficient matching is

$$\begin{aligned} u_E = {} & \mu_{\text{AB}}\mu_{\text{B}} + \mu_{\text{A}}(\mu_{\text{AB}} + 2\mu_{\text{B}}) + \beta\mu_{\text{O}} \\ & + \bar{p}\big[\mu_{\text{A}}^2 + \mu_{\text{A}}\mu_{\text{AB}} + \mu_{\text{AB}}^2 + \mu_{\text{AB}}\mu_{\text{B}} + \mu_{\text{B}}^2 \\ & + 2(\mu_{\text{A}} + \mu_{\text{AB}} + \mu_{\text{B}})\mu_{\text{O}} + \mu_{\text{O}}^2\big] \end{aligned}$$

To calculate the price of fairness we now find the size of the fair matching. The only unmatched highly sensitized patients are in $V_H^{\text{O-AB}}$, some of which were matched in step 12 above. We now show that the number of matched vertices in $V^{\text{O-AB}}$ is smaller than the initial size of $V_H^{\text{O-AB}}$, so not all vertices in $V_H^{\text{O-AB}}$ can be matched. Let $M^{\text{O-AB}}$ be the number of matched vertices in $V^{\text{O-AB}}$, and $H^{\text{O-AB}}$ be the initial size of $V_H^{\text{O-AB}}$. The inequality $M^{\text{O-AB}} < H^{\text{O-AB}}$ can be written as

$$(\beta + \mu_{\text{AB}}\bar{p})(1-\mu_{\text{AB}}) - \mu_{\text{AB}}(\mu_{\text{A}} + \mu_{\text{B}}) < (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}} \tag{7}$$

$$\beta < \mu_{\text{AB}}(1-\bar{p}) - \lambda\mu_{\text{O}}\frac{\mu_{\text{AB}}}{1-\mu_{\text{AB}}} \tag{8}$$

This inequality holds by assumption **5**, and a.s. there are some unmatched vertices in $V_H^{\text{O-AB}}$. The number of unmatched highly sensitized vertices is

$$H^{\text{O-AB}} - M^{\text{O-AB}} \propto (1-\mu_{\text{AB}})((1-\bar{p})\mu_{\text{AB}} - \beta) - \lambda\mu_{\text{AB}}\mu_{\text{O}}$$

.

We match each of these remaining vertices by removing a 3-cycle of the form (AB-O, O-A, A-AB) and creating a 2-cycle (AB-O, O-AB). This matching used $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{O}}\mu_{\text{AB}}$ 3-cycles of this form, while $|V_H^{\text{O-AB}}| \propto (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}$. The model assumptions ensure that $|V^{\text{AB-O}}| > |V_H^{\text{O-AB}}|$, and all remaining vertices in $V_H^{\text{O-AB}}$ can be matched in this way.

To match each remaining vertex in $V_H^{\text{O-AB}}$, we remove from the matching one vertex from both $V^{\text{O-A}}$ and $V^{\text{A-AB}}$. Thus the total efficiency loss is $H^{\text{O-AB}} - M^{\text{O-AB}}$. The price of fairness is

$$\mathsf{POF}(\mathcal{M}, u_{LEX}) = \frac{(1-\mu_{\text{AB}})((1-\bar{p})\mu_{\text{AB}} - \beta) - \lambda\mu_{\text{AB}}\mu_{\text{O}}}{u_E}$$

With u_E defined previously. □

Proposition 5. *Assume*

1 $\beta > \mu_{AB}(1-\bar{p}) - \mu_{AB}\mu_O\bar{p}/(\mu_A+\mu_B)$

2 $\beta < \mu_{AB}(1-\bar{p}) - \lambda\mu_O\frac{\mu_{AB}}{1-\mu_{AB}}$

These constraints imply $\beta \in [0, 1/10]$. *Denote by* $\mathcal{M}$ *the set of matchings in* $G(n)$ *using cycles and chains up to length 3. Almost surely as* $n \to \infty$*, the price of fairness is*

$$\textit{POF}(\mathcal{M}, u_{LEX}) = \frac{(1-\mu_{AB})((1-\bar{p})\mu_{AB} - \beta) - \lambda\mu_{AB}\mu_O}{u_E}$$

with

$$\begin{aligned} u_E = {} & \mu_{AB}\mu_B + \mu_A(\mu_{AB} + 2\mu_B) + \beta\mu_O \\ & + \bar{p}\big[\mu_A^2 + \mu_A\mu_{AB} + \mu_{AB}^2 + \mu_{AB}\mu_B + \mu_B^2 \\ & + 2(\mu_A + \mu_{AB} + \mu_B)\mu_O + \mu_O^2\big] \end{aligned}$$

sketch. We begin with matching M^* as done in the proof of Lemma 1, matching all highly sensitized vertices except for those in $V_H^{\text{AB-O}}$. We now complete the efficient matching using both 3-cycles and 3-chains as in (Dickerson *et al.* 2012).

7. A- and B-type NDDs donate to $V^{\text{A-AB}}$ and $V^{\text{B-AB}}$, respectively. Note that $|N^A| \propto \beta\mu_A$ and $|V^{\text{A-AB}}| \propto (1-\bar{p})\mu_A\mu_{\text{AB}}$. The inequality $\beta\mu_{\text{A}} < \mu_A\mu_{\text{AB}}(1-\bar{p})$ holds due to assumption **2**, and a.s. $|N^{\text{A}}| < |V^{\text{A-AB}}|$. By the same argument, a.s. $|N^{\text{B}}| < |V^{\text{B-AB}}|$. Thus, both N^{A} and N^{B} are exhausted, and $|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{A}}$ and $|V^{\text{B-AB}}| \propto \mu_{\text{B}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{B}}$.
8. Create cycles of the form (AB-O, O-A, A-AB). The current size of each vertex group is
   (1) $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$
   (2) $|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}\mu_{\text{O}}$
   (3) $|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{A}}$

   Note that the inequality $(3) < (1)$ can be written as
   $$\beta > \mu_{\text{AB}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}/\mu_{\text{A}}$$
   which holds by assumption **1** and a.s. $|V^{\text{A-AB}}| < |V^{\text{AB-O}}|$. The inequality $(3) < (2)$ can be written as
   $$\beta > (1-\bar{p})(\mu_{\text{AB}} - \mu_{\text{O}})$$
   which holds by the model assumptions, and a.s. $|V^{\text{A-AB}}| < |V^{\text{O-A}}|$. Executing these cycles exhausts $V^{\text{A-AB}}$ and leaves the following vertices remaining
   $$|V^{\text{O-A}}| \propto (1-\bar{p})\mu_{\text{A}}(\mu_{\text{O}} - \mu_{\text{AB}}) + \mu_{\text{A}}\beta$$
   $$|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + \mu_{\text{A}}\beta$$
9. Create cycles of the form (AB-O, O-B, B-AB). The current size of each vertex group is
   (1) $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{AB}}\mu_{\text{O}} - \mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p}) + \mu_{\text{A}}\beta$
   (2) $|V^{\text{O-B}}| \propto (1-\bar{p})\mu_{\text{B}}\mu_{\text{O}}$
   (3) $|V^{\text{B-AB}}| \propto \mu_{\text{B}}\mu_{\text{AB}}(1-\bar{p}) - \beta\mu_{\text{B}}$

   Inequality $(3) < (2)$ can be written as
   $$\beta > (1-\bar{p})(\mu_{\text{AB}} - \mu_{\text{O}})$$
   which holds by the model assumptions. Inequality $(3) < (1)$ can be written as
   $$\beta > \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/(\mu_{\text{A}} + \mu_{\text{B}})$$
   which holds by assumption **1**.
   Executing these cycles exhausts $V^{\text{B-AB}}$ and leaves the following vertices remaining
   $$|V^{\text{AB-O}}| \propto (\beta - (1-\bar{p})\mu_{\text{AB}})(\mu_{\text{A}} + \mu_{\text{B}}) + \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$$
   $$|V^{\text{B-AB}}| \propto \mu_{\text{B}}(\beta - (1-\bar{p})(\mu_{\text{AB}} - \mu_{\text{O}}))$$
10. Create chains of the form (O,O-A,A-AB). Previous steps exhausted $V^{\text{A-AB}}$ so none of these chains occur.
11. Create chains of the form (O,O-B,B-AB). Previous steps exhausted $V^{\text{B-AB}}$ so none of these chains occur.
12. O-type NDDs and $V^{\text{AB-O}}$ match with remaining under-demanded vertices, starting with $V^{\text{O-AB}}$. The remaining size of each vertex group is
   (1) $|N^{\text{O}}| \propto \beta\mu_{\text{O}}$
   (2) $|V^{\text{AB-O}}| \propto (\beta - (1-\bar{p})\mu_{\text{AB}})(\mu_{\text{A}} + \mu_{\text{B}}) + \bar{p}\mu_{\text{AB}}\mu_{\text{O}}$
   (3) $|V^{\text{O-AB}}| \propto \mu_{\text{O}}\mu_{\text{AB}}$

   Note that the inequality $(1) + (2) < (3)$ can be written as
   $$\beta < \mu_{\text{AB}}(1-\bar{p})$$
   which holds by assumption **2** Thus O-type NDDs are exhausted first, leaving some vertices remaining in $V^{\text{O-AB}}$, with
   $$|V^{\text{O-AB}}| \propto ((1-\bar{p})\mu_{\text{AB}} - \beta)(1-\mu_{\text{AB}})$$

In the efficient matching described above, the number of *matched* pairs in each under-demanded group is

$$\begin{aligned} &|V^{\text{O-A}}| \propto \mu_{\text{A}}(\mu_{\text{AB}} + \bar{p}(\mu_{\text{O}} - \mu_{\text{AB}}) - \beta) \\ &|V^{\text{O-B}}| \propto \mu_{\text{B}}(\mu_{\text{AB}} + \bar{p}(\mu_{\text{O}} - \mu_{\text{AB}}) - \beta) \\ &|V^{\text{A-AB}}| \propto \mu_{\text{A}}\mu_{\text{AB}} \\ &|V^{\text{B-AB}}| \propto \mu_{\text{B}}\mu_{\text{AB}}. \\ &|V^{\text{O-AB}}| \propto (\beta + \mu_{\text{AB}}\bar{p})(1-\mu_{\text{AB}}) - \mu_{\text{AB}}(\mu_{\text{A}} + \mu_{\text{B}}) \end{aligned}$$

Combining these with the over-demanded and self-demanded vertices, the total size of the efficient matching is

$$\begin{aligned} u_E = {} & \mu_{\text{AB}}\mu_{\text{B}} + \mu_{\text{A}}(\mu_{\text{AB}} + 2\mu_{\text{B}}) + \beta\mu_{\text{O}} \\ & + \bar{p}\big[\mu_{\text{A}}^2 + \mu_{\text{A}}\mu_{\text{AB}} + \mu_{\text{AB}}^2 + \mu_{\text{AB}}\mu_{\text{B}} + \mu_{\text{B}}^2 \\ & + 2(\mu_{\text{A}} + \mu_{\text{AB}} + \mu_{\text{B}})\mu_{\text{O}} + \mu_{\text{O}}^2\big] \end{aligned}$$

To calculate the price of fairness we now find the size of the fair matching. The only unmatched highly sensitized patients are in $V_H^{\text{O-AB}}$, some of which were matched in step 12 above. We now show that the number of matched vertices in $V^{\text{O-AB}}$ is smaller than the initial size of $V_H^{\text{O-AB}}$, so not all vertices in $V_H^{\text{O-AB}}$ can be matched. Let $M^{\text{O-AB}}$ be the number

of matched vertices in $V^{\text{O-AB}}$, and $H^{\text{O-AB}}$ be the initial size of $V_H^{\text{O-AB}}$. The inequality $M^{\text{O-AB}} < H^{\text{O-AB}}$ can be written as

$$(\beta + \mu_{\text{AB}}\bar{p})(1-\mu_{\text{AB}}) - \mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}}) < (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}} \tag{9}$$

$$\beta < \mu_{\text{AB}}(1-\bar{p}) - \lambda\mu_{\text{O}}\frac{\mu_{\text{AB}}}{1-\mu_{\text{AB}}} \tag{10}$$

This inequality holds by assumption **2**, and a.s. there are some unmatched vertices in $V_H^{\text{O-AB}}$. The number of unmatched highly sensitized vertices is

$$H^{\text{O-AB}} - M^{\text{O-AB}} \propto (1-\mu_{\text{AB}})((1-\bar{p})\mu_{\text{AB}} - \beta) - \lambda\mu_{\text{AB}}\mu_{\text{O}}.$$

We match each of these remaining vertices by removing a 3-cycle of the form (AB-O, O-A, A-AB) and creating a 2-cycle (AB-O, O-AB). This matching used $|V^{\text{AB-O}}| \propto \bar{p}\mu_{\text{O}}\mu_{\text{AB}}$ 3-cycles of this form, while $|V_H^{\text{O-AB}}| \propto (1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}$. The model assumptions ensure that $|V^{\text{AB-O}}| > |V_H^{\text{O-AB}}|$, and all remaining vertices in $V_H^{\text{O-AB}}$ can be matched in this way.

To match each remaining vertex in $V_H^{\text{O-AB}}$, we remove from the matching one vertex from both $V^{\text{O-A}}$ and $V^{\text{A-AB}}$. Thus the total efficiency loss is $H^{\text{O-AB}} - M^{\text{O-AB}}$. The price of fairness is

$$\mathsf{POF}(\mathcal{M}, u_{LEX}) = \frac{(1-\mu_{\text{AB}})((1-\bar{p})\mu_{\text{AB}} - \beta) - \lambda\mu_{\text{AB}}\mu_{\text{O}}}{u_E}$$

With u_E defined previously. □

Next we compare the price of fairness in Propositions 2, 3, 4, and 5 to the price of fairness in the efficient matching without NDDs, given in Dickerson *et al.* (2014):

$$\text{POF}_0 = \frac{(1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}}{u_E} \tag{11}$$

$$\begin{aligned} u_E = \bar{p}\Big[&2\mu_{\text{AB}}\mu_{\text{B}} + 2\mu_{\text{AB}}\mu_{\text{A}} + 3\mu_{\text{AB}}\mu_{\text{O}} \\ &+2\mu_{\text{A}}\mu_{\text{O}} + 2\mu_{\text{B}}\mu_{\text{O}} + \mu_{\text{O}}^2 + \mu_{\text{A}}^2 + \mu_{\text{B}}^2 + \mu_{\text{AB}}^2\Big] \\ &+2\mu_{\text{A}}\mu_{\text{B}} \end{aligned}$$

The following Lemmas state that POF_0 is an upper bound on the price of fairness when NDDs are used, for each of the four cases when the price of fairness is nonzero.

Lemma 2. *The price of fairness in Propositions 2 and 3 is bounded above by POF$_0$.*

sketch. The price of fairness in Propositions 2 and 3 is

$$\text{POF}_A = \frac{(1-\lambda)\mu_{\text{O}}\mu_{\text{AB}}}{u_E}$$

$$\begin{aligned} u_E = \bar{p}\Big[&2\mu_{\text{AB}}\mu_{\text{B}} + 2\mu_{\text{AB}}\mu_{\text{A}} + 3\mu_{\text{AB}}\mu_{\text{O}} \\ &+2\mu_{\text{A}}\mu_{\text{O}} + 2\mu_{\text{B}}\mu_{\text{O}} + \mu_{\text{O}}^2 + \mu_{\text{A}}^2 + \mu_{\text{B}}^2 + \mu_{\text{AB}}^2\Big] \\ &+2\mu_{\text{A}}\mu_{\text{B}} + \beta\left(\mu_{\text{A}} + \mu_{\text{B}} + 2\mu_{\text{O}}\right) \end{aligned}$$

Both POF_0 and POF_A have the same numerator, and the denominator of POF_A is equal to the denominator of POF_0, with the additional term $\beta\left(\mu_{\text{A}} + \mu_{\text{B}} + 2\mu_{\text{O}}\right)$. Thus when $\beta = 0$, $\text{POF}_0 = \text{POF}_A$, and when $\beta > 0$, $\text{POF}_0 > \text{POF}_A$, and the price of fairness in Propositions 2 and 3 is bounded above by POF_0. □

Lemma 3. *The price of fairness in Propositions 4 and 5 is bounded above by POF$_0$.*

sketch. The price of fairness in Propositions 4 and 5 is

$$\text{POF}_B = \frac{(1-\mu_{\text{AB}})((1-\bar{p})\mu_{\text{AB}} - \beta) - \lambda\mu_{\text{AB}}\mu_{\text{O}}}{u_E}$$

$$\begin{aligned} u_E = \mu_{\text{AB}}\mu_{\text{B}} + \mu_{\text{A}}\left(\mu_{\text{AB}} + 2\mu_{\text{B}}\right) + \beta\mu_{\text{O}} \\ + \bar{p}\big[\mu_{\text{A}}^2 + \mu_{\text{A}}\mu_{\text{AB}} + \mu_{\text{AB}}^2 + \mu_{\text{AB}}\mu_{\text{B}} + \mu_{\text{B}}^2 \\ + 2(\mu_{\text{A}} + \mu_{\text{AB}} + \mu_{\text{B}})\mu_{\text{O}} + \mu_{\text{O}}^2\big] \end{aligned}$$

To show that $\text{POF}_B < \text{POF}_0$ holds, we first show both (1) the numerator of POF_B is smaller than that of POF_0, and (2) the denominator of POF_B is larger than the denominator of POF_0.

(1) In both POF_0 and POF_B, the numerator is proportional to the number of remaining vertices in $V_H^{\text{O-AB}}$, after constructing the efficient matching. In Proposition 4 and 5 the efficient matching contains some vertices in $V_H^{\text{O-AB}}$; without NDDs, the efficient matching contains no vertices in $V_H^{\text{O-AB}}$. Thus, the numerator of POF_B is strictly smaller the numerator of POF_0.

(2) Let the D_0 be the denominator of POF_0, and D_B be the denominator of POF_B. We now show that the inequality $D_0 < D_B$ holds. First, note that this inequality can be written as

$$\mu_{\text{AB}} - (1-\bar{p})\mu_{\text{AB}}^2 + \beta\mu_{\text{O}} > \mu_{\text{AB}}(\bar{p} + \mu_{\text{O}}).$$

Rearranging, we have

$$\beta > (\mu_{\text{AB}}/\mu_{\text{O}})\big[(1-\bar{p})\mu_{\text{AB}} - (\mu_{\text{A}} + \mu_{\text{B}} + \mu_{\text{AB}} - \bar{p})\big]. \tag{12}$$

We now show that inequality 12 is satisfied by the the following assumption on β, made in Propositions 4 and 5:

$$\mathbf{A}: \beta > \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/\mu_{\text{A}}.$$

Next, we show that assumption **A** implies inequality 12, and thus assumption **A** implies $D_0 < D_B$. Assumption **A** implies 12 if the right-hand side of **A** is larger than the right hand side of 12, that is,

$$\begin{aligned} \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/\mu_{\text{A}} > (\mu_{\text{AB}}/\mu_{\text{O}})(1-\bar{p})\mu_{\text{AB}} \\ - (\mu_{\text{AB}}/\mu_{\text{O}})(\mu_{\text{A}} + \mu_{\text{B}} + \mu_{\text{AB}} - \bar{p}) \end{aligned}$$

rearranging, we have

$$\frac{1-\bar{p}}{\bar{p}} > \frac{\mu_{\text{O}}}{\mu_{\text{A}}}\frac{1-\mu_{\text{AB}}-\mu_{\text{B}}}{1-\mu_{\text{B}}}$$

The random graph model assumes $\bar{p} \leq 2/5$, and $\mu_{\text{O}} \leq (3/2)\mu_{\text{A}}$, thus we have

$$\frac{1-\bar{p}}{\bar{p}} \geq \frac{3}{2} > \frac{3}{2}\frac{1-\mu_{\text{AB}}-\mu_{\text{B}}}{1-\mu_{\text{B}}} \geq \frac{\mu_{\text{O}}}{\mu_{\text{A}}}\frac{1-\mu_{\text{AB}}-\mu_{\text{B}}}{1-\mu_{\text{B}}}.$$

This shows that assumption **A** implies $D_0 < D_B$.

Thus, the numerator of POF_0 is larger than the numerator of POF_B, and the denominator of POF_0 is smaller than the denominator of POF_B, and therefore $\text{POF}_B < \text{POF}_0$. □

Lemmas 2 and 3 show that with $\beta > 0$, the price of fairness has the same upper bound as when $\beta = 0$, given in Dickerson *et al.* (2014). That is, adding NDDs to the random graph model does not increase the price of fairness.

Theorem 1. *Adding NDDs to the random graph model ($\beta > 0$) does not increase the upper bound on the price of fairness found by Dickerson* et al. *(2014).*

Proof. When $\beta > 0$, there are only four possible matchings with nonzero price of fairness, and the price of fairness for each case is given in Propositions 2, 3, 4, and 5. Lemmas 2 and 3 state that in each of these four cases, the matching with NDDs has a tighter bound on the price of fairness than the matching without NDDs, given in Dickerson *et al.* (2014). □

Next we show that the price of fairness is zero when $\beta > 1/8$, by finding the maximum possible β for each of the four cases with nonzero price of fairness.

Lemma 4. *In the matching described by Proposition 2, $\beta < 1/8$.*

Proof. Proposition 2 makes the following assumptions on β:

1 $\beta < \mu_{\text{A}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}$
2 $\beta < \mu_{\text{AB}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}/\mu_{\text{A}}$
3 $\beta < \mu_{\text{AB}}\left(\frac{\mu_{\text{A}}}{\mu_{\text{A}}+\mu_{\text{O}}} - \bar{p}\right)$

To determine an upper bound on β, we maximize the right hand side of constraint **3**. Note that the model assumes $\mu_{\text{AB}} < 1/4$, $\mu_{\text{A}} < 1/2$, and $\mu_{\text{A}} + \mu_{\text{O}} < 1$. Using these bounds, and $\bar{p} \to 0$, constraint **3** is bounded by

$$\beta < \mu_{\text{AB}}\left(\frac{\mu_{\text{A}}}{\mu_{\text{A}}+\mu_{\text{O}}} - \bar{p}\right) < (1/4)\frac{(1/2)}{1} = 1/8$$
$$\beta < 1/8$$

Constraints **1** and **2** are looser than constraint **3**: with the values $\bar{p} \to 0$, $\mu_{\text{A}} \to 1/4$, and $\mu_{\text{AB}} \to 1/4$, both constraints reduce to $\beta < 1/4$.

□

Lemma 5. *In the matching described by Proposition 3, $\beta < 1/12$.*

Proof. Proposition 3 makes the following assumptions

1 $\beta < \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/(\mu_{\text{A}}+\mu_{\text{B}})$
2 $\beta < \frac{\mu_{\text{A}}\mu_{\text{AB}}(1-\bar{p})+\mu_{\text{B}}\mu_{\text{O}}(1-\bar{p})-\bar{p}\mu_{\text{O}}\mu_{\text{AB}}}{\mu_{\text{A}}+\mu_{\text{O}}}$
3 $\beta > \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/\mu_{\text{A}}$
4 $\beta < \mu_{\text{AB}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}/\mu_{\text{A}} + (1-\bar{p})\mu_{\text{B}}\mu_{\text{O}}/\mu_{A}$
5 $\beta < \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{O}}\mu_{\text{AB}}/(1-\mu_{\text{AB}})$

Combining **3** and **5**, we have

$$\mu_{\text{O}}\mu_{\text{AB}}/(1-\mu_{\text{AB}}) < \mu_{\text{AB}}(1-\bar{p}) - \beta < \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/\mu_{\text{A}}$$
$$\mu_{\text{O}}\mu_{\text{AB}}/(1-\mu_{\text{AB}}) < \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/\mu_{\text{A}}$$
$$\mathbf{A} : \mu_{\text{A}}/(1-\mu_{\text{AB}}) < \bar{p}$$

Combining constraint **A** with **5** gives a new upper bound on β,

$$\begin{aligned}\beta &< \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{O}}\mu_{\text{AB}}/(1-\mu_{\text{AB}}) \\ &< \mu_{\text{AB}}(1-\mu_{\text{A}}/(1-\mu_{\text{AB}})) - \mu_{\text{O}}\mu_{\text{AB}}/(1-\mu_{\text{AB}})\end{aligned}$$

$$\beta < \mu_{\text{AB}}\left(1 - \frac{\mu_{\text{A}}+\mu_{\text{O}}}{1-\mu_{\text{AB}}}\right)$$

This bound is maximized when when μ_{AB} is maximal, and $(\mu_{\text{A}}+\mu_{\text{O}})$ is minimal. In the random graph model, these values are $\mu_{\text{AB}} \to 1/4$ and $(\mu_{\text{A}}+\mu_{\text{O}}) \to 1/2$, and the numerical bound is

$$\beta < (1/4)\left(1 - \frac{(1/2)}{1-1/4}\right) = 1/12$$
$$\beta < 1/12$$

□

Lemma 6. *In the matching described by Proposition 4, $\beta < 1/8$.*

Proof. Proposition 4 makes the following assumptions on β

1 $\beta > \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/\mu_{\text{A}}$
2 $\beta < \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/(\mu_{\text{A}}+\mu_{\text{B}})$
3 $\beta < \mu_{\text{AB}}(1-\bar{p}) - \bar{p}\mu_{\text{AB}}\mu_{\text{O}}/\mu_{\text{A}} + (1-\bar{p})\mu_{\text{B}}\mu_{\text{O}}/\mu_{A}$
4 $\beta > \mu_{\text{AB}}\left((1-\bar{p}) - \frac{\mu_{\text{O}}}{1-\mu_{\text{AB}}}\right)$
5 $\beta < \mu_{\text{AB}}(1-\bar{p}) - \lambda\mu_{\text{O}}\frac{\mu_{\text{AB}}}{1-\mu_{\text{AB}}}$

Combining **1** and **5** results in the following constraint, which is consistent with the above assumptions:

$$\mathbf{A} : \lambda\frac{\mu_{\text{A}}}{1-\mu_{\text{AB}}} < \bar{p}$$

Note that **5** is maximized when λ is minimized; this occurs when $\lambda + \bar{p} \to 1$, and $\lambda \to 1 - \bar{p}$. In this case, **5** can be relaxed as

$$\begin{aligned}\beta &< \mu_{\text{AB}}(1-\bar{p}) - \lambda\mu_{\text{O}}\frac{\mu_{\text{AB}}}{1-\mu_{\text{AB}}} \\ &< \mu_{\text{AB}}(1-\bar{p}) - (1-\bar{p})\mu_{\text{AB}}\frac{\mu_{\text{O}}}{1-\mu_{\text{AB}}}\end{aligned}$$

$$\begin{aligned}\beta &< \mu_{\text{AB}}(1-\bar{p}) - (1-\bar{p})\mu_{\text{AB}}\frac{\mu_{\text{O}}}{1-\mu_{\text{AB}}} \\ &= (1-\bar{p})\frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}}\end{aligned}$$

Finally, we have

$$\beta < (1-\bar{p})\frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}}$$

The right hand side of this constraint is maximal when $\bar{p}$ is minimal; constraint $\mathbf{A}$ determines the lower bound for $\bar{p}$, with $\lambda \to 1-\bar{p}$:

$$\begin{aligned}(1-\bar{p})\frac{\mu_{\text{A}}}{1-\mu_{\text{AB}}} &< \bar{p} \\ \frac{\mu_{\text{A}}}{1-\mu_{\text{AB}}} &< \bar{p}\left(1+\frac{\mu_{\text{A}}}{1-\mu_{\text{AB}}}\right) \\ \frac{\mu_{\text{A}}}{1-\mu_{\text{AB}}+\mu_{\text{A}}} &< \bar{p}\end{aligned}$$

Using this lower bound on $\bar{p}$, we can further relax **5**

$$\begin{aligned}\beta &< (1-\bar{p})\frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}} \\ &< (1-\frac{\mu_{\text{A}}}{1-\mu_{\text{AB}}+\mu_{\text{A}}})\frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}} \\ &= \frac{1-\mu_{\text{AB}}}{1-\mu_{\text{AB}}+\mu_{\text{A}}}\frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}} \\ &= \frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}+\mu_{\text{A}}}\end{aligned}$$

$$\beta < \frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}+\mu_{\text{A}}}$$

The right hand side is maximal when μ_{AB} is maximal, and $\mu_{\text{AB}}, \mu_{\text{A}}, \mu_{\text{B}}, \mu_{\text{O}} \to 1/4$. This gives the final bound on β,

$$\begin{aligned}\beta &< \frac{(1/4)(1/2)}{1} = 1/8 \\ \beta &< 1/8\end{aligned}$$

□

Lemma 7. *In the matching described by Proposition 5,* $\beta < 1/10$.

Proof. Proposition 5 makes the following assumptions on β

1 $\beta > \mu_{\text{AB}}(1-\bar{p}) - \mu_{\text{AB}}\mu_{\text{O}}\bar{p}/(\mu_{\text{A}}+\mu_{\text{B}})$
2 $\beta < \mu_{\text{AB}}(1-\bar{p}) - \lambda\mu_{\text{O}}\frac{\mu_{\text{AB}}}{1-\mu_{\text{AB}}}$

Combining these assumptions results in the following constraint:

$$\mathbf{A}: \lambda\frac{\mu_{\text{A}}+\mu_{\text{B}}}{1-\mu_{\text{AB}}} < \bar{p}$$

Note that assumption **2** is identical to assumption **5** in Lemma 6. Following the same procedure used in the proof of Lemma 6, **2** can be relaxed as

$$\beta < (1-\bar{p})\frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}}$$

The right hand side of this constraint is maximal when $\bar{p}$ is minimal; constraint $\mathbf{A}$ determines the lower bound for $\bar{p}$, with $\lambda \to 1-\bar{p}$:

$$\begin{aligned}(1-\bar{p})\frac{\mu_{\text{A}}+\mu_{\text{B}}}{1-\mu_{\text{AB}}} &< \bar{p} \\ \frac{\mu_{\text{A}}+\mu_{\text{B}}}{1-\mu_{\text{AB}}} &< \bar{p}\left(1+\frac{\mu_{\text{A}}+\mu_{\text{B}}}{1-\mu_{\text{AB}}}\right) \\ \frac{\mu_{\text{A}}+\mu_{\text{B}}}{2\mu_{\text{A}}+2\mu_{\text{B}}+\mu_{\text{O}}} &< \bar{p}\end{aligned}$$

Using this lower bound on $\bar{p}$, we can further relax **2**

$$\begin{aligned}\beta &< (1-\bar{p})\frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}} \\ &< (1-\frac{\mu_{\text{A}}+\mu_{\text{B}}}{2\mu_{\text{A}}+2\mu_{\text{B}}+\mu_{\text{O}}})\frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}} \\ &= \frac{1-\mu_{\text{AB}}}{2\mu_{\text{A}}+2\mu_{\text{B}}+\mu_{\text{O}}}\frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{1-\mu_{\text{AB}}} \\ &= \frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{2\mu_{\text{A}}+2\mu_{\text{B}}+\mu_{\text{O}}}\end{aligned}$$

$$\beta < \frac{\mu_{\text{AB}}(\mu_{\text{A}}+\mu_{\text{B}})}{2\mu_{\text{A}}+2\mu_{\text{B}}+\mu_{\text{O}}}$$

The right hand side is maximal when μ_{AB} is maximal, and $\mu_{\text{AB}}, \mu_{\text{A}}, \mu_{\text{B}}, \mu_{\text{O}} \to 1/4$. This gives the final bound on β,

$$\begin{aligned}\beta &< \frac{(1/4)(1/2)}{5/4} = 1/10 \\ \beta &< 1/10\end{aligned}$$

□

Combining Lemmas 4, 5, 6, and 7, we find that the price of fairness is zero when $\beta > 1/8$.

Theorem 2. *The price of fairness is zero when* $\beta > 1/8$.

Proof. There are only four matchings with nonzero price of fairness and $\beta > 0$, which are described in Propositions 2, 3, 4, and 5. Lemmas 4, 5, 6, and 7 state that the maximum β for any of these matchings is $1/8$. When $\beta > 1/8$, the matching is not one of these four cases, and the price of fairness is zero. □

Theorems 1 and 2 are the two main theoretical results of this paper: adding NDDs to the random graph model does not increase the upper bound on the price of fairness, and when the proportion of NDDs is high enough ($\beta > 1/8$), the price of fairness is zero. We show this by addressing each of the four efficient matchings on the random graph model with nonzero price of fairness. In each case, and $\beta < 1/8$, and the matching with NDDs has a smaller price of fairness than the matching without NDDs given in Dickerson *et al.* (2014).

To further explore these results, we numerically find the maximum price of fairness for the matchings given in Propositions 2, 3, 4, and 5. For each matching, we find the maximum price of fairness for a range of β, within the defined constraints, using the "NMaximize" function in Mathematica with the nonlinear interior point method.

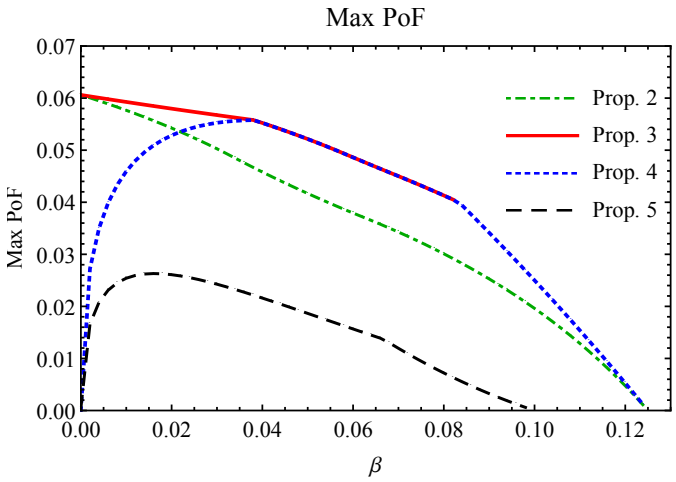


Figure 5: Maixmum price of fairness for each of the four matchings addressed in Propositions 2, 3, 4, and 5.

Figure A.2 confirms both of our main theoretical results: adding NDDs to the efficient matching decreases the upper bound on the price of fairness, and when $\beta > 1/8$ the price of fairness is zero.

B Price of Fairness for α-Lexicographic-, Weighted-, and Hybrid-Lexicographic Fairness

This section presents Theorems and Proofs regarding the price of fairness for the lexicographic, weighted, and hybrid-lexicographic fairness rules.

B.1 Lexicographic Fairness

Theorem 3. *For any cycle cap L there exists a graph G such that the price of fairness of G under the α-lexicographic fairness rule with $0 < \alpha \leq 1$ is bounded by $\mathsf{POF}(\mathcal{M}, u_\alpha) \geq \frac{L-2}{L}$.*

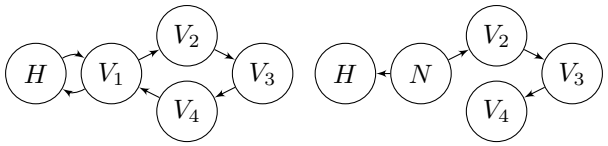


Figure 6: Supporting graphs for Theorems 3 (left) and 4 (right), with cycle cap 4 and chain cap 3, respectively.

Proof. Consider a kidney exchange graph consisting of one highly-sensitized patient H and L non-highly-sensitized patients V_i that form a directed cycle of length L. A 2-cycle connects H with one V_i, as shown in Figure 6. With a cycle cap of L, the optimal utilitarian matching has utility L, while the optimal lexicographic matching has utility $u_\alpha = 2$, for any $0 < \alpha \leq 1$. The price of fairness in this graph is $\mathsf{POF}(\mathcal{M}, u_\alpha) = (L-2)/L$. □

Theorem 4. *For any chain cap R there exists a graph G such that the price of fairness of G under the α-lexicographic fairness rule with $0 < \alpha \leq 1$ is bounded by $\mathsf{POF}(\mathcal{M}, u_\alpha) \geq \frac{R-1}{R}$.*

Proof. Consider the graph used in the proof of Theorem 3, with vertex V_2 as an NDD rather than a pair. With a chain cap of R, the optimal utilitarian matching has utility R, while the optimal α-lexicographic matching has utility $u_\alpha = 1$ for any $0 < \alpha \leq 1$. The price of fairness in this graph is $\mathsf{POF}(\mathcal{M}, u_\alpha) = (R-1)/R$. □

B.2 Weighted Fairness

Theorem 5. *For any cycle cap L and $\gamma \geq L-1$, there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\mathsf{POF}(\mathcal{M}, u_{WF}) \geq \frac{L-2}{L}$.*

Proof. Consider the graph used in the proof of Theorem 3, with all edge weights equal to 1. Weighted fairness increases the weight of the edge ending in H to $(1+\gamma)$. The weighted utility of the 2-cycle is $2+\gamma$, while the weighted utility of the L-cycle is L. If γ is chosen such that $\gamma \geq L-2$, then the 2-cycle will be chosen over the L-cycle, resulting in the price of fairness $\mathsf{POF}(\mathcal{M}, u_{WF}) = (L-2)/L$. □

Theorem 6. *For any chain cap R and $\gamma \geq R-1$, there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\mathsf{POF}(\mathcal{M}, u_{WF}) \geq \frac{R-1}{R}$.*

Proof. Consider the graph used in the proof of Theorem 4, with all weights equal to 1. The weighted utility of the 1-chain is $1+\gamma$, while the weight of the R-chain is R. If γ is chosen such that $\gamma \geq R-1$, then the 1-chain will be chosen over the R-chain, resulting in the price of fairness $\mathsf{POF}(\mathcal{M}, u_{WF}) = (R-1)/R$. □

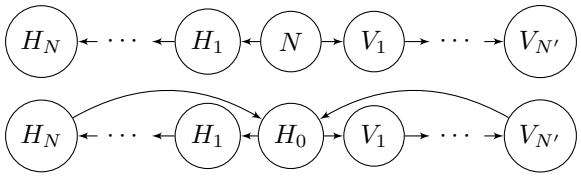


Figure 7: Graphs for Theorems 7 (top) and 8 (bottom).

Theorem 7. *With no chain cap, there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\mathsf{POF}(\mathcal{M}, u_{WF}) \geq \frac{\gamma}{\gamma+1}$.*

Proof. Consider a graph with a single NDD connected to a chain with highly-sensitized patients H_i of length N, and a chain with non-highly sensitized patients V_i of length $N' = \lfloor(\gamma+1)N\rfloor - 1$. Under weighted fairness, the V_i chain receives utility $u_L = \lfloor(\gamma+1)N\rfloor - 1$ while the H_i chain receives utility $u_H = (\gamma+1)N$, so $u_H > u_L$. The price of fairness for this graph is

$$\mathsf{POF}(\mathcal{M}, u_{WF}) = \frac{\lfloor\gamma N_H\rfloor - 1}{\lfloor\gamma N\rfloor + N - 1} \geq \frac{\gamma N - 2}{(\gamma+1)N - 1}.$$

Taking the limit as $N \to \infty$ yields

$$\lim_{N\to\infty} \frac{\gamma N - 2}{(\gamma+1)N - 1} = \frac{\gamma}{\gamma+1},$$

which implies $\mathsf{POF}(\mathcal{M}, u_{WF}) \geq \frac{\gamma}{\gamma+1}$. □

Theorem 8. *With no cycle cap there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\mathsf{POF}(\mathcal{M}, u_{WF}) \geq \frac{\gamma}{\gamma+1}$.*

Proof. Consider the graph used in the proof of Theorem 7, where the NDD N is instead a highly-sensitized pair H_0, and the end vertices of both chains both have edges ending in H_0. Under weighted fairness, the V_i cycle receives utility $u_L = \lfloor(\gamma+1)N\rfloor$, while the H_i chain receives utility $u_H = (\gamma+1)N + 1$, so $u_H > u_L$. The price of fairness for this graph is

$$\mathsf{POF}(\mathcal{M}, u_{WF}) = \frac{\lfloor\gamma\rfloor N - 1}{\lfloor\gamma N\rfloor + N} \geq \frac{\gamma N - 1}{(\gamma+1)N + 1}.$$

Taking the limit as $N \to \infty$ yields

$$\lim_{N\to\infty} \frac{\gamma N - 1}{(\gamma+1)N + 1} = \frac{\gamma}{\gamma+1},$$

□

B.3 Hybrid-Lexicographic

Theorem 9. *Assume the optimal utilitarian outcome X_E receives utility $u(X_E) = u_E$, with one disadvantaged class that receives utility u_1, and Z non-disadvantaged classes such that $u_1(X_E) > u_i(X_E)$. For $|\mathcal{P}|$ classes, $\mathsf{POF}(\mathcal{M}, u_\Delta) \leq \frac{2((|\mathcal{P}|-1)-Z)\Delta}{u_E}$.*

Proof. Consider two outcomes, one in the fair regime (X_F), one in the utilitarian regime (X_E). Let $u_\Delta(X_F) > u_\Delta(X_E)$, such that X_E receives nearly the same utility as X_F; that is, $u_\Delta(X_E) = u_\Delta(X_F) - \epsilon$ for some $0 < \epsilon \ll 1$. WLOG, let there be Z classes i such that $u_1(X_E) > u_i(X_E)$, and

$$\begin{aligned} u_\Delta(X_E) &= u_\Delta(X_F) - \epsilon \\ &\leq \sum_{i=1}^{|\mathcal{P}|} u_i(X_F) + (|\mathcal{P}| - 1)\Delta - \epsilon \end{aligned} \tag{13}$$

Using the definition of utilitarian utility $u_E = \sum_{i=1}^{|\mathcal{P}|} u_i$,

$$u_E(X_E) - u_E(X_F) \leq (2(|\mathcal{P}| - 1) - 2Z)\Delta - \epsilon$$

and the price of fairness is

$$\mathsf{POF}(\mathcal{M}, u_\Delta) \leq \frac{2((\mathcal{P} - 1) - Z)\Delta}{u_E(X_E)}.$$

□

C Experimental Results

This section contains worst-case price of fairness (PoF) and worst-case fairness ($\%F$) for real UNOS graphs, and for simulated graphs; these results were produced using the method described in Section 5.

C.1 UNOS Graphs

Figure C.1 shows the worst-case (maximum) PoF of each fairness rule on the 314 UNOS graphs; Figure C.1 shows worst-case (minimum) $\%F$.

Real exchange graphs are relatively sparse, and have very few feasible matchings. Each fairness rule effectively chooses one of these matchings, based on a fairness criteria. Especially with sparse graphs, fairness is often achieved by using longer cycles or cycles to match highly sensitized vertices. When edge success probability p is high, fairness has little effect on overall utility, and PoF is often below 0.3. With lower edge success probability, using longer cycles and chains causes a huge loss in efficiency: the expected utility of n-cycles and chains is proportional to p^n, which incurs a huge penalty for long cycles and chains when p is small. Thus as p decreases, very small α and β values result in a high PoF. Our results show that for $p \leq 0.8$, even the smallest parameters for α-lexicographic and weighted fairness ($\alpha = 0.1$ and $\beta = 2$) achieve the worst-case PoF. As expected, hybrid-lexicographic fairness limits PoF according to Theorem 9. With two classes of patients (highly- and lowly-sensitized), the theoretical price of fairness is bounded by $\mathsf{POF}(\mathcal{M}, u_\Delta) \leq 2\Delta/u(M_E)$; in the Figures, Δ is scaled by $u(M_E)$, so the upper bound on the price of fairness has a slope of two.

To illustrate the other side of the fairness-efficiency tradeoff, we consider worst case $\%F$. Figure C.1 shows the minimum (worst case) $\%F$ over all UNOS graphs for each fairness rule, and for various edge success probabilities and chain caps.

As expected, α-lexicographic fairness guarantees at $\%F \geq \alpha$; weighted and hybrid-lexicogrpahic fairness do

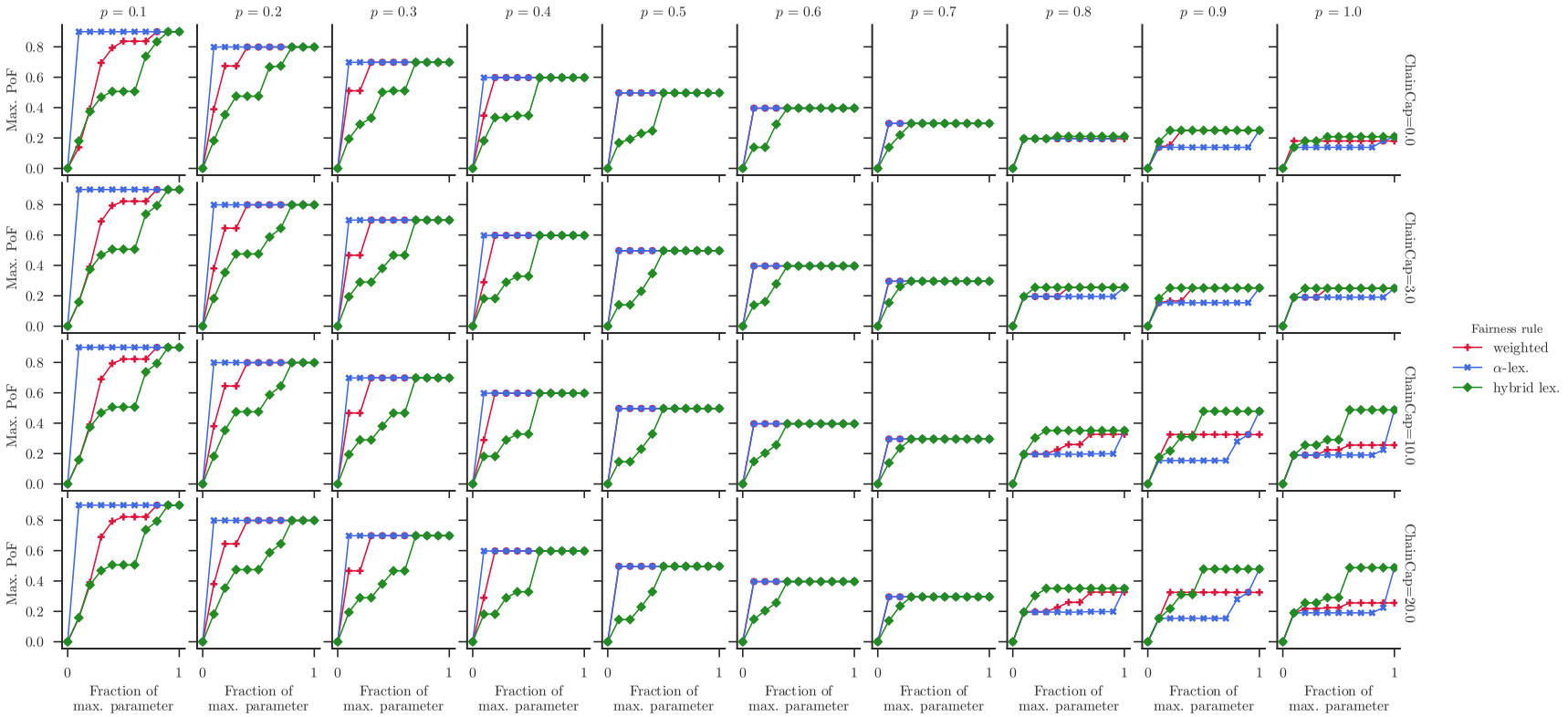


Figure 8: Maximum PoF for each fairness rule. Parameters for each rule are $\alpha \in [0, 1]$, $\beta \in [0, 20]$, and $\Delta \in [0, u(M_E)]$. Rows correspond to edge success probabilities from 0.1 to 1.0; columns correspond to different chain caps: 0, 3, and 20.

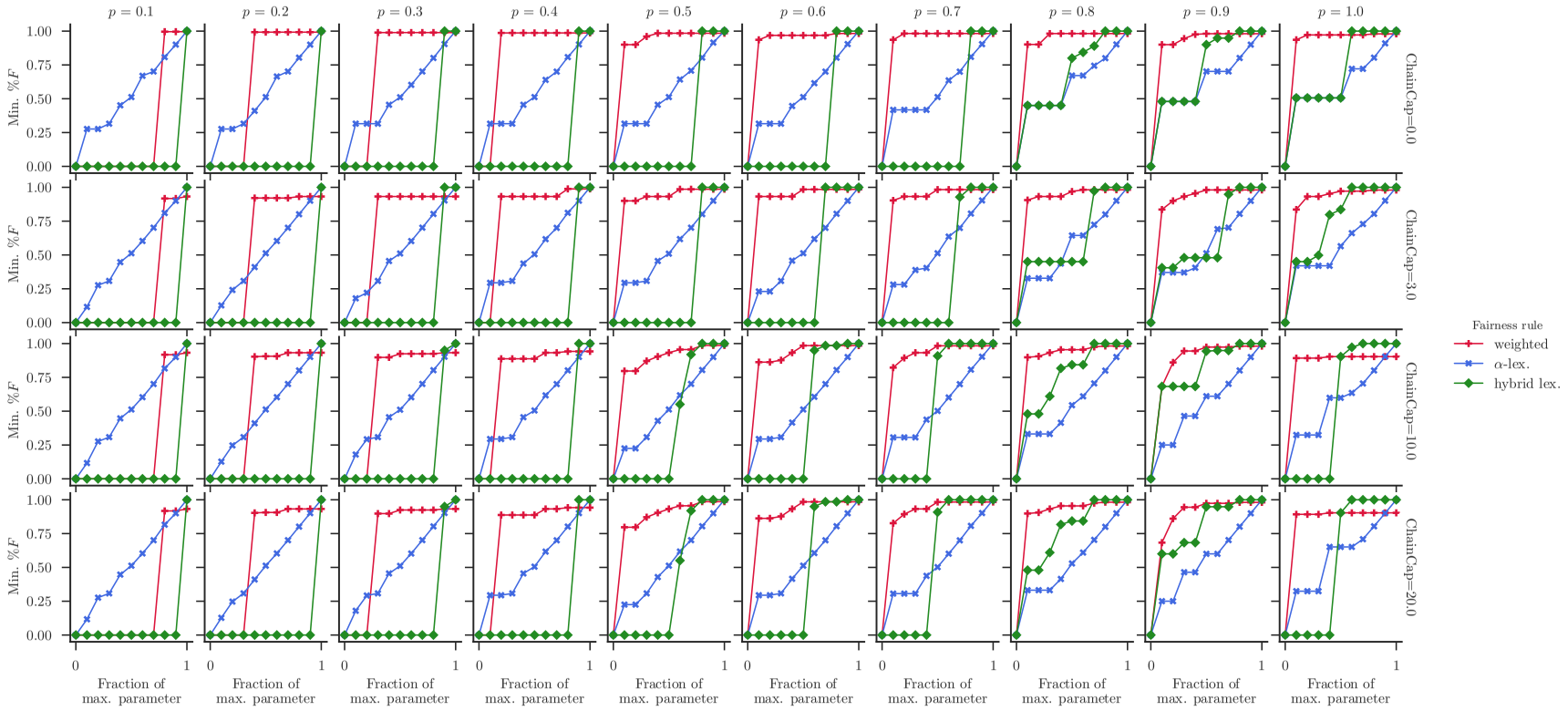


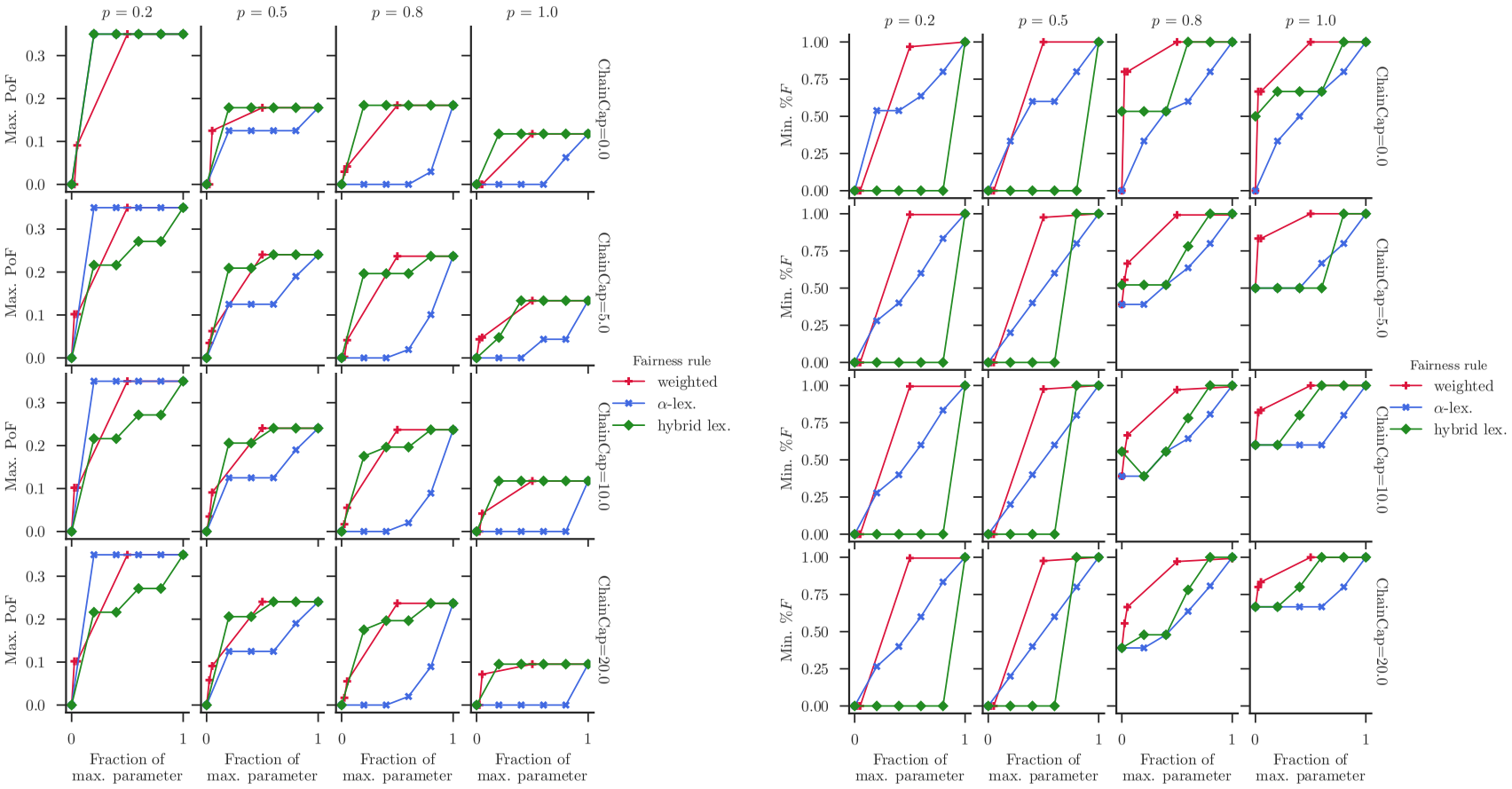
Figure 9: Minimum fraction of the fair score for each fairness rule. Parameters for each rule are $\alpha \in [0, 1]$, $\beta \in [0, 20]$, and $\Delta \in [0, u(M_E)]$. Rows correspond to edge success probabilities from 0.1 to 1.0; columns correspond to different chain caps: 0, 3, and 20.

not make this guarantee. Small edge success probabilities make it impossible to match highly sensitized patients without large efficiency loss; when p becomes small hybrid-lexicographic fairness matches no highly sensitized patients in the worst case.

These results demonstrate the balance between fairness and efficiency offered by both α-lexicographic and hybrid-lexicographic fairness. If fairness is more important than efficiency, then the α-lexicographic rule can be used to guarantee that the resulting matching achieves at least fraction α of the maximum possible fair score. Alternatively, if efficiency is more important than fairness, hybrid-lexicographic fairness can bound the price of fairness using parameter Δ.

C.2 Simulated Exchange Graphs

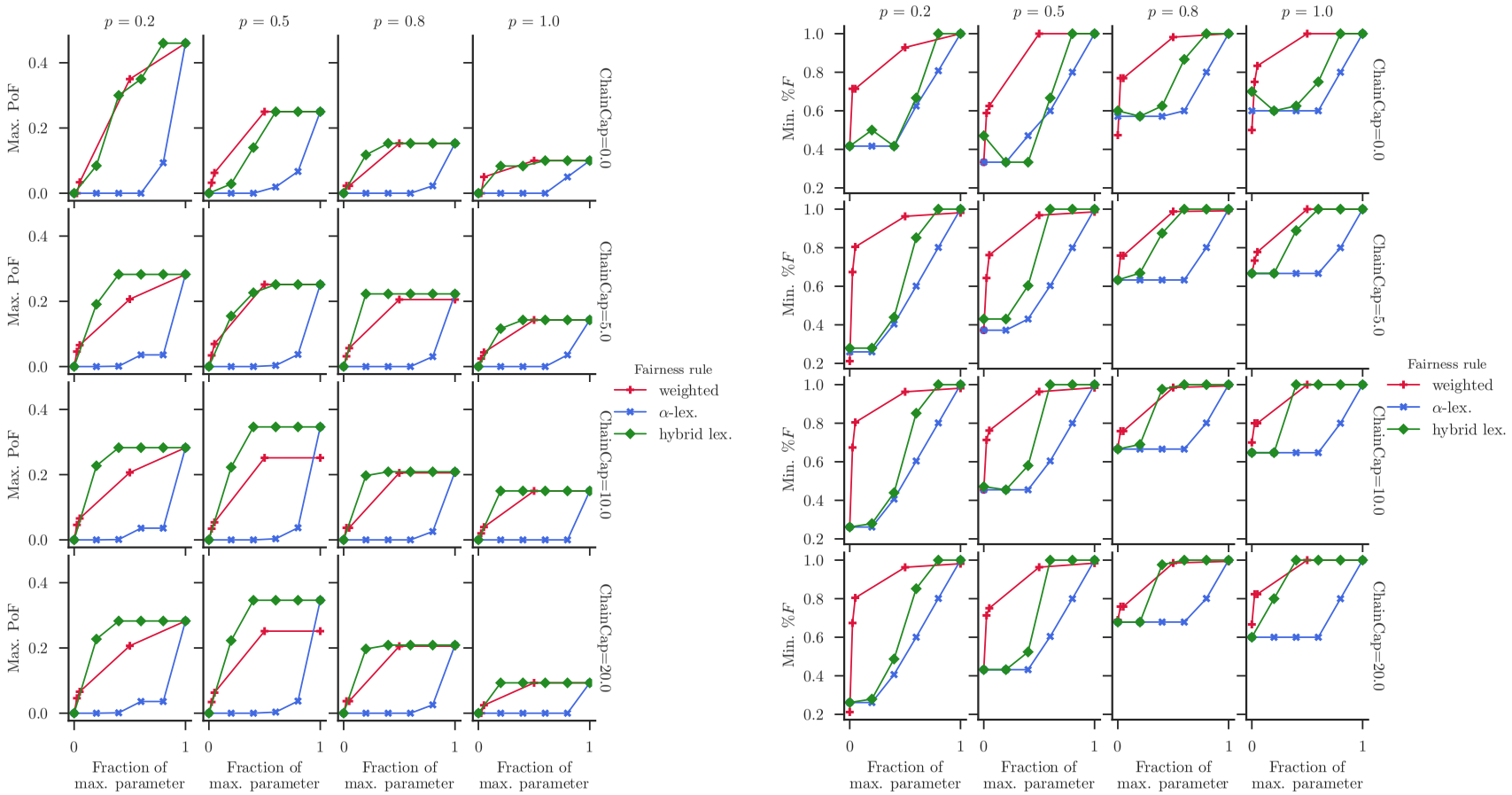
Simulated exchange graphs were drawn from previous UNOS exchanges, using the same method as Dickerson *et al.* (2013). These graphs are typically denser than real graphs, and have a much lower price of fairness. Figures 10 and 11 show the worst-case PoF and $\%F$ on 32 simulated exchanges of size 64 and 128.



(a) Maximum PoF.

(b) Minimum $\%F$.

Figure 10: Worst-case PoF and $\%F$ for 32 64-vertex random graphs. Parameters for each rule are $\alpha \in [0, 1]$, $\beta \in [0, 20]$, and $\Delta \in [0, u(M_E)]$. Rows correspond to edge success probabilities from 0.1 to 1.0; columns correspond to different chain caps: 0, 3, and 20.



(a) Maximum PoF.

(b) Minimum $\%F$.

Figure 11: Worst-case PoF and $\%F$ for 32 128-vertex random graphs. Parameters for each rule are $\alpha \in [0, 1]$, $\beta \in [0, 20]$, and $\Delta \in [0, u(M_E)]$. Rows correspond to edge success probabilities from 0.1 to 1.0; columns correspond to different chain caps: 0, 3, and 20.