Adapting a Kidney Exchange Algorithm to Align with Human Values

Rachel Freedman, Jana Schaich Borg, Walter Sinnott-Armstrong, John P. Dickerson, Vincent Conitzer

Duke UNIVERSITY

COMPUTER SCIENCE
UNIVERSITY OF MARYLAND

Why let AI make moral decisions?

1. Speed

   Autonomous Vehicles

2. Magnitude

   Advertising

3. Computational Complexity

   Kidney Exchanges...

Kidney Exchanges

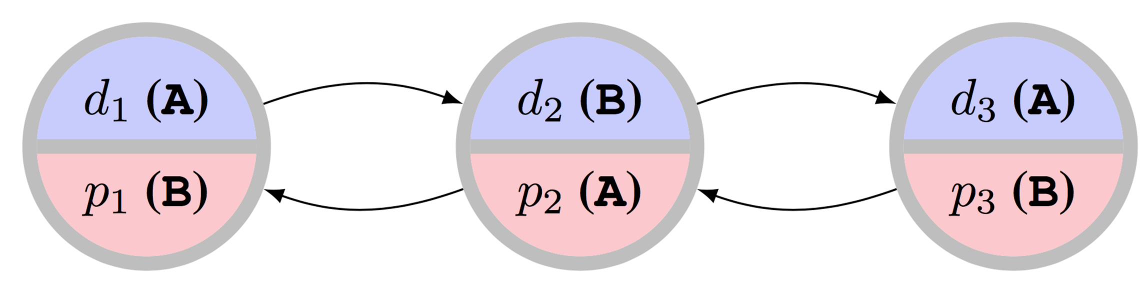


Figure 1: A compatibility graph with three patient-donor pairs and two possible 2-cycles. Donor and patient blood types are given in parantheses.

- There are **100,000 patients** on the US kidney transplant waiting list.
- Kidney exchanges allow patients with willing but incompatible live donors to **swap donors**.
- Algorithms developed in the AI community are used to find **optimal matchings**.

How to break ties between patients?

Patient A is **70 years old**, has **1 alcoholic drink per month**, and has **no other major health problems**.

Patient B is **30 years old**, has **5 alcoholic drinks per day**, and has **skin cancer in remission**.

...who should get the kidney?

Mturker Preferences

Profile	Age	Drinking	Cancer	Preferred
1 (YRH)	30	rare	healthy	94.0%
3 (YRC)	30	rare	cancer	76.8%
2 (YFH)	30	frequently	healthy	63.2%
5 (ORH)	70	rare	healthy	56.1%
4 (YFC)	30	frequently	cancer	43.5%
7 (ORC)	70	rare	cancer	36.3%
6 (OFH)	70	frequently	healthy	23.6%
8 (OFC)	70	frequently	cancer	6.4%

Table 2: Profile ranking according to Kidney Allocation Survey responses. The "Preferred" column describes the percentage of time the indicated profile was chosen among all the times it appeared in a comparison.

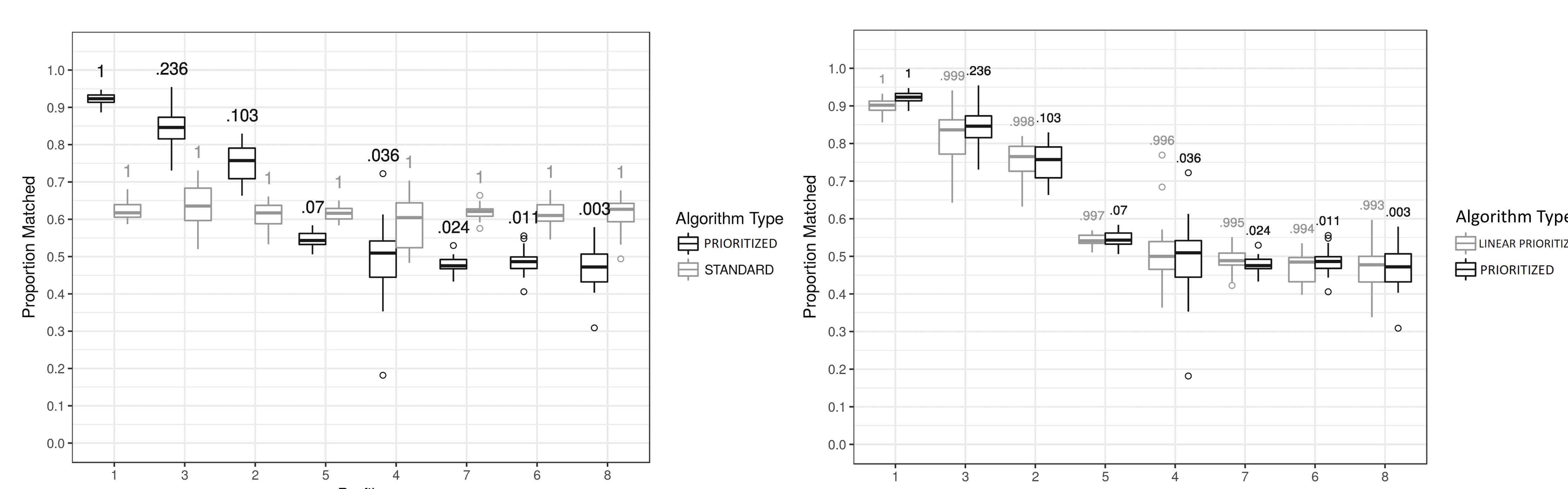
Incorporating these values

1. Aggregate into "scores" using the **Bradley-Terry model**:

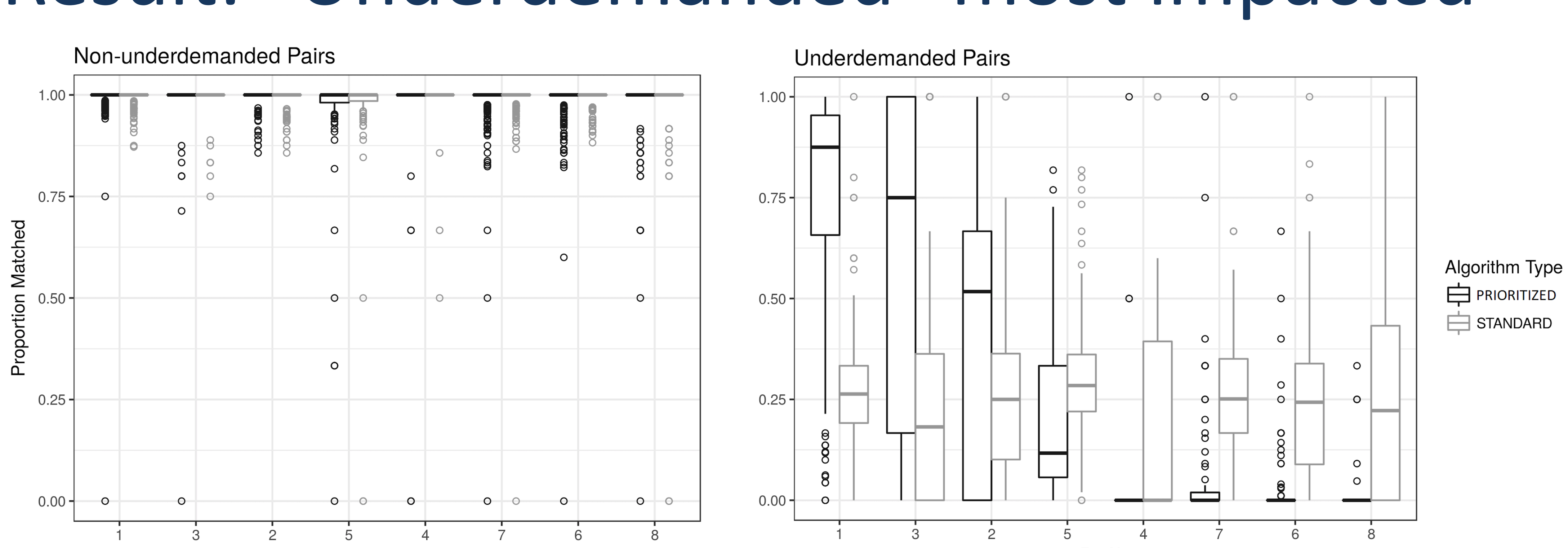
$$P(i > j) = \frac{p_i}{p_i + p_j}$$

2. Use these scores to break ties between otherwise optimal matchings

Result: Order matters; differences don't



Result: "Underdemanded" most impacted



Underdemanded *patient-donor pairs have blood types that make them difficult to match.*

This paper will be presented at the AAAI main conference at **10am, Feb 7th**. It received an honorable mention for outstanding student paper.

This work is partially supported by the project "How to Build Ethics into Robust Artificial Intelligence" funded by the Future of Life Institute, and by NSF IIS-1527434. We thank Lirong Xia, Zhibing Zhao, and Kyle Burris, and members of our Moral AI group at Duke, including Yuan Deng, Kenzie Doyle, Jeremy Fox, Max Kramer, and Eitan Sapiro-Gheiler, for feedback on this work.