Microsoft

Microsoft Research

Summit 2021

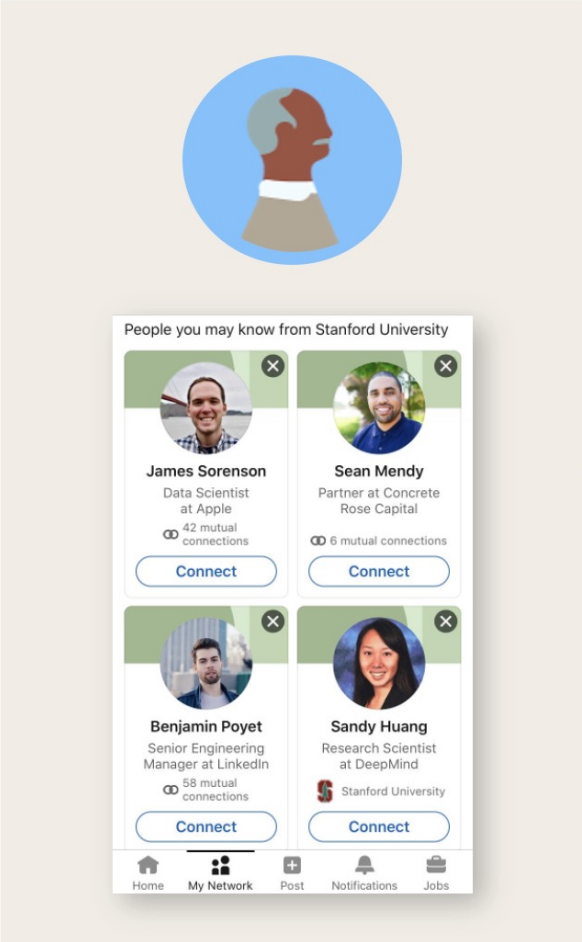
Fairness via Post-Processing for Web-Scale Recommender Systems

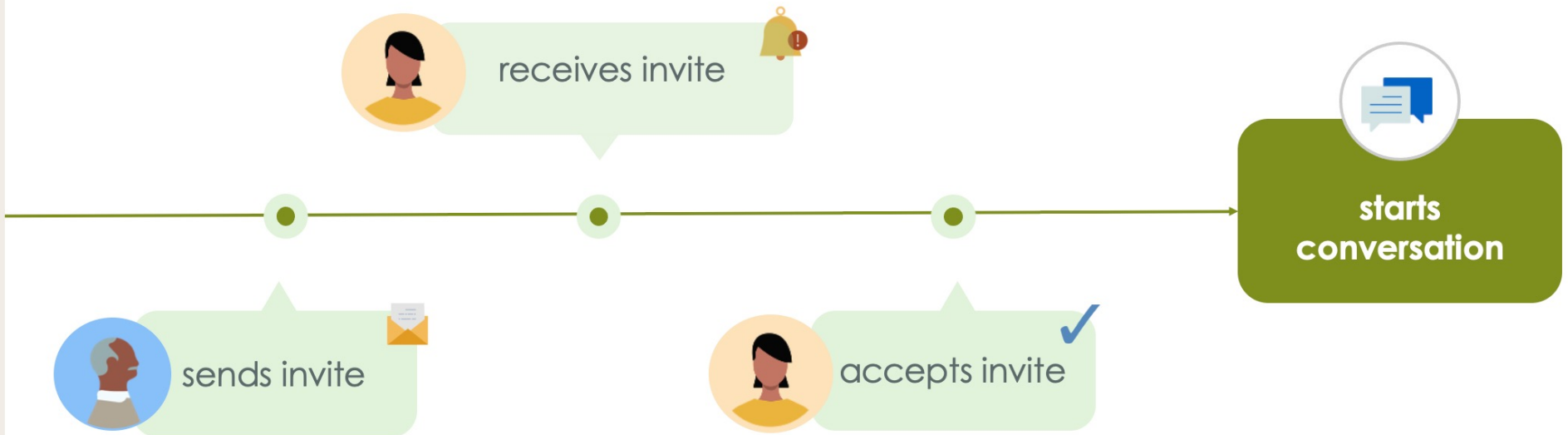
Kinjal Basu

Tech Lead for Responsible AI, LinkedIn

Joint work with Preetam Nandy, Cyrus Diciccio, Heloise Logan, Divya Venugopalan and Noureddine El Karoui

Connection Recommendation in LinkedIn





Fairness Criteria for score-based Ranking

Notations: Binary Response Y, Predictors X, Prediction Score s(X), Protected Attribute C

Demographic Parity

$$s(X) \perp\!\!\!\perp C$$

Predictive Rate Parity

$$Y \perp\!\!\!\perp C \mid s(X)$$

Equality of Opportunity

$$s(X) \perp\!\!\!\perp C \mid Y = 1$$

Equalized Odds

$$s(X) \perp\!\!\!\perp C \mid Y \in \{0, 1\}$$

Full Paper

Fairness Criteria for score-based Ranking

Notations: Binary Response Y, Predictors X, Prediction Score s(X), Protected Attribute C

Demographic Parity

$$s(X) \perp\!\!\!\perp C$$

Predictive Rate Parity

$$Y \perp\!\!\!\perp C \mid s(X)$$

Equality of Opportunity

$$s(X) \perp\!\!\!\perp C \mid Y = 1$$

Equalized Odds

$$s(X) \perp\!\!\!\perp C \mid Y \in \{0, 1\}$$

Full Paper

Mitigation Strategies

Post-Processing Solutions:

- Apply a transformation $s(X) \rightarrow F(s(X))$ such that the transformed scores satisfy some metric definition.
- Among all feasible $F(\cdot)$, we want to use the one that optimizes model performance.
- Careful consideration for position bias.

Equality of Opportunity

- $F(\cdot)$ = CDF of the score corresponding to $C = c$
- Monotonic Transformation ➡ Unchanged ranking for each group

Equalized Odds

- $F(\cdot)$ = Minimizer of $|s(X) - F(s(X))|$ w.r.t. Equalized Odds constraints
- Can be solved as a Linear Program

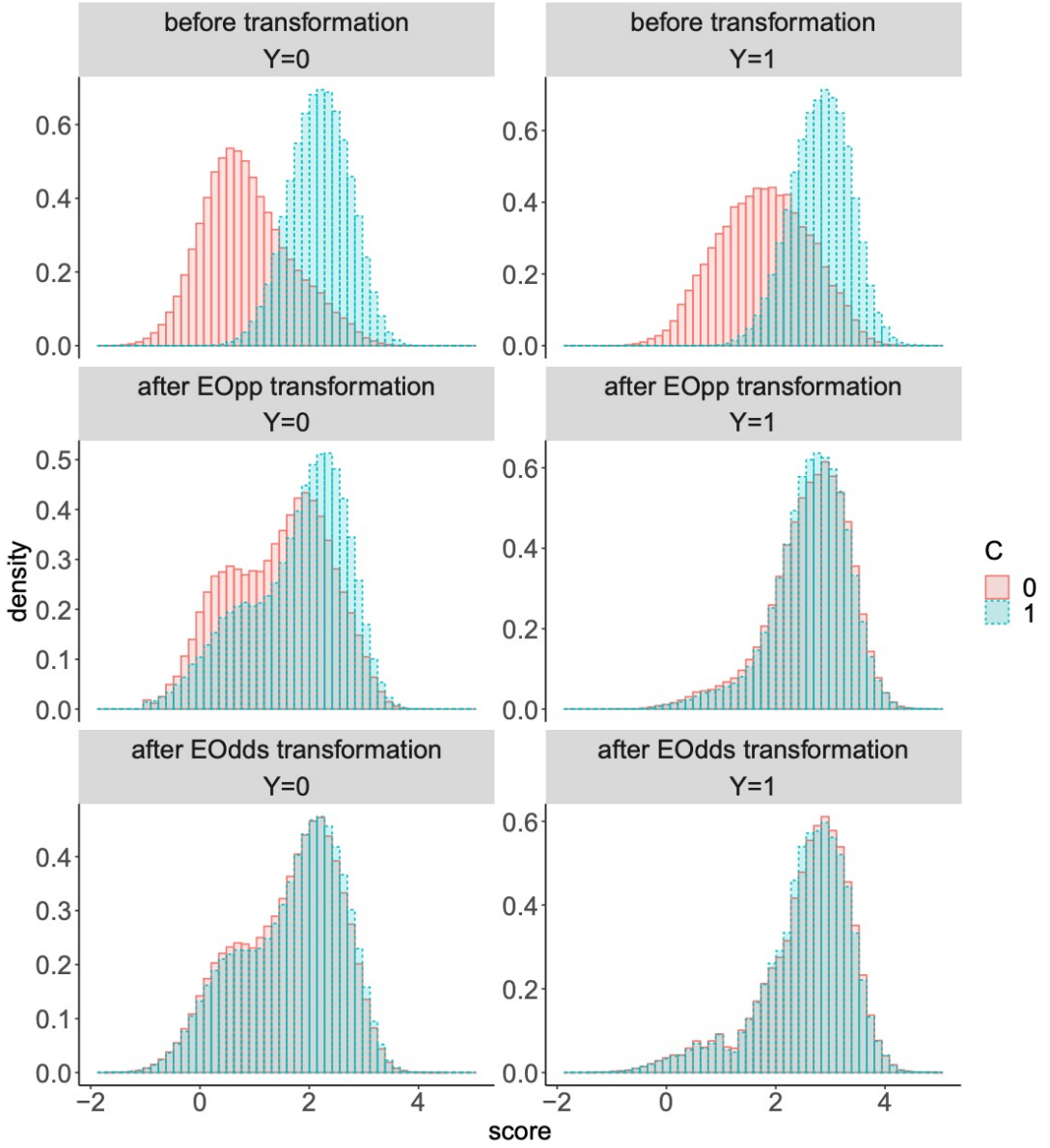
Full Paper

Results of Online Experiment

- **Score** $s(X)$**:** PYMK score for top 100 positions
- **Label** Y**:** Invitation Sent
- **Attribute** C : Infrequent and Frequent members

Invitation Metrics	*EOpp*		*EOdds*	
	IM	FM	IM	FM
Sent	+5.72%	Neutral	+2.77%	Neutral
Accepted	+ 4.85%	Neutral	+ 2.26%	Neutral

Table 1. A/B Experimentation results for the two fairness re-rankers. In both setups, we observed improved metrics of invitations sent and accepted by IMs without any statistically significant impact to the same metrics corresponding to FMs.



Full Paper


Thank You!