Counterfactual Reciprocal Recommender Systems for User-to-User Matching

Kazuki Kawamura
Sony Group Corporation
Japan
Kazuki.Kawamura@sony.com

Takuma Udagawa
Sony Group Corporation
Japan
Takuma.Udagawa@sony.com

Kei Tateno
Sony Group Corporation
Japan
Kei.Tateno@sony.com

arXiv:2508.01867v1 [cs.IR] 3 Aug 2025

Abstract

Reciprocal recommender systems (RRS) in dating, gaming, and talent platforms require mutual acceptance for a match. Logged data, however, over-represents popular profiles due to past exposure policies, creating feedback loops that skew learning and fairness. We introduce Counterfactual Reciprocal Recommender Systems (CFRR), a causal framework to mitigate this bias. CFRR uses inverse propensity scored, self-normalized objectives. Experiments show CFRR improves NDCG@10 by up to 3.5% (e.g., from 0.459 to 0.475 on DBLP, from 0.299 to 0.307 on Synthetic), increases long-tail user coverage by up to 51% (from 0.504 to 0.763 on Synthetic), and reduces Gini exposure inequality by up to 24% (from 0.708 to 0.535 on Synthetic). CFRR offers a promising approach for more accurate and fair user-to-user matching.

CCS Concepts

• **Information systems** → **Recommender systems**; • **Theory of computation** → *Causal reasoning and inference*; • **General and reference** → Metrics.

Keywords

Reciprocal recommendation, selection bias, causal inference, inverse propensity scoring, doubly robust, fairness

ACM Reference Format:

Kazuki Kawamura, Takuma Udagawa, and Kei Tateno. 2025. Counterfactual Reciprocal Recommender Systems for User-to-User Matching. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining Workshop on Two-sided Marketplace Optimization: Search, Discovery, Matching, Pricing & Growth (TSMO '25), August 4, 2025, Toronto, Canada.* ACM, New York, NY, USA, 9 pages.

1 Introduction

Digital platforms increasingly rely on *user-to-user* recommendation. For instance, online dating services have become highly prevalent, and similar matching is pervasive in multiplayer gaming, talent marketplaces, social audio, and peer-learning communities [27]. In these settings, a recommendation is successful only when *both* parties accept it, making reciprocal recommender systems (RRS) fundamentally different from classic item recommenders [23]. The bilateral nature magnifies ethical stakes: a low-exposure user never even appears as a candidate partner, while popular or photogenic users absorb an outsized share of attention [4, 5].

The root cause is **exposure bias**. What the system shows today constrains the data available tomorrow, so learning on *policy-filtered* logs reinforces the same skew—a self-perpetuating *feedback loop* that widens inequality over time [16, 24, 33]. Longitudinal studies confirm that popularity bias inflates recommendation accuracy metrics offline while harming user welfare online [1, 6, 25]. Consequently, offline evaluation and model selection become unreliable, and fairness objectives such as two-sided envy-freeness are systematically violated [19].

Causal inference provides a principled remedy. Inverse Propensity Scoring (IPS) treats every historical recommendation as an intervention and re-weights each logged interaction by the inverse of its display probability, yielding a consistent estimator of the true risk [29]. Unfortunately, real propensities can be tiny, producing extreme weights and high variance; self-normalized IPS (SNIPS) [31] reduces variance but introduces a small finite-sample bias. Stabilized doubly robust (StableDR) estimators [18] further aim to control variance and improve robustness when propensities are small, typically for *item* recommendation. Moreover, existing RRS work often overlooks challenges like bilateral exposure, pair-level propensities, and critical implementation details. The latter include consistently normalizing objectives over observed pairs (which SNIPS inherently addresses [31]) and effectively managing variance in such objectives.

We fill this gap with *Counterfactual Reciprocal Recommender Systems* (CFRR). CFRR first estimates pair-level propensities to correct exposure bias. Then, it optimizes a self-normalized IPS (SNIPS) objective that provides stable training even with extreme weights. The framework naturally supports variance-control techniques—weight truncation and doubly robust augmentation—making it robust enough for industrial deployments where traffic is highly imbalanced. Extensive synthetic experiments demonstrate that, relative to strong RRS baselines, CFRR improves NDCG@10 by up to approximately 3.5% (e.g., on DBLP, from 0.459 to 0.475), widens user coverage by up to approximately 51% (on Synthetic, from 0.504 to 0.763), and trims Gini exposure inequality by up to approximately 24% (on Synthetic, from 0.708 to 0.535), even when propensities are noisy.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

TSMO '25, Toronto, Canada

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Our contributions are three-fold:

- **Bias-consistent and Stable Objective.** User-to-user matching as counterfactual estimation with a consistent IPS loss (correct propensities assumed). Our SNIPS-based CFRR objective provides a stable training framework that effectively handles extreme weights common in reciprocal matching scenarios.
- **Variance-Aware Learning for RRS.** CFRR integrates self-normalization across the entire objective and incorporates established variance reduction techniques like weight truncation and a carefully formulated doubly robust augmentation, adapted for the reciprocal matching context to mitigate variance from extreme weights.
- **Empirical Validation.** Controlled simulations and real-world dataset experiments confirm that CFRR can simultaneously boost ranking quality, expand long-tail exposure, and counter the feedback loop that otherwise amplifies popularity bias.

Despite its promising performance, deploying CFRR in a real system raises several additional considerations. We discuss in later sections how to handle extremely large user bases (positivity in practice), the impact of propensity model misspecification, and how CFRR can be extended to other two-sided platforms and fairness objectives.

2 Related Work

This section reviews existing user-to-user matching approaches, highlighting limitations of methods ignoring selection bias. We then discuss fairness, two-sided constraints crucial in RRS, and situate our work concerning selection bias correction.

2.1 Optimization-Based Methods

Stable Matching. The classic Gale–Shapley algorithm [10] for stable matching, initially for the stable marriage problem, guarantees stability (no unmatched pair mutually prefers each other). It is broadly applied in markets like college admissions [28] and medical residencies. Yet, traditional algorithms assume complete preference orderings, often impractical for large-scale systems. They may also favor one market side, and direct application to dynamic platforms is challenging. Recent work explores scalability for large-scale RRS via parallel/mini-batch computations [20].

Maximum Weight Matching. Other optimization methods model user pairs (u, v) with a score, seeking to maximize total matched pair scores. This is often formulated as maximum weight matching (MWM) [9] or minimum-cost flow [22]. They find global optima under capacity constraints (e.g., one match per user). However, MWM doesn't guarantee bilateral satisfaction or address score estimation bias from data, which can yield suboptimal/unfair outcomes.

2.2 Machine Learning-Based Methods

Collaborative Filtering. Techniques like collaborative filtering, originally for item recommendation, can be adapted to user-user matching by treating users as items [26, 34]. They predict user-user preferences from past interactions or profile similarities. However, sparse user-user data and unaddressed interaction biases in standard CF can yield poor generalization and reinforce popularity bias.

Reciprocal Recommenders. Reciprocal recommender systems (RRS) specifically model mutual preferences. They often learn distinct user representations (initiator/receiver) or use architectures for two-way dynamics [21, 34]. Recent deep learning, like GNNs, better models complex RRS relations [7]. Though expressive for reciprocity, many assume observed interactions represent true preferences. This is often violated by selection/exposure biases from past policies [29]. Recent works offer new causal RRS formulations and metrics, noting unique domain challenges [35].

2.3 Fairness and Two-Sided Constraints

Reciprocal fairness is multifaceted, beyond relevance to equitable exposure/opportunity. Early work explored Rawlsian fairness [14] and Nash Social Welfare balancing [12]. Other research targeted equitable group exposure [30]. However, many such methods presume unbiased preference/compatibility estimates. Fairness/bias correction interplay is critical; fairness on biased data may not yield truly fair outcomes. Recent literature investigates new fairness definitions/algorithms for reciprocal settings (e.g., fair representation, mitigating discrimination [32]). Effectively integrating robust bias correction with fairness objectives remains a challenge.

2.4 Selection Bias in User-to-User Matching

Selection bias plagues most RRS: training data isn't random, heavily influenced by past exposure policies. This creates feedback loops skewing data toward popular users [24]. Causal inference offers a principled remedy. IPS and related estimators reweight/correct data for a balanced distribution (surveyed by [11]). Propensity score estimation is challenging; errors impact performance [8]. Applying these to RRS adds complexities: pair-level propensities (depending on both users) and bilateral outcomes. While some works apply causal reasoning to RRS (e.g., bilateral interventions [35], reciprocal link debiasing [35]), a comprehensive framework for robust pair-level propensities, self-normalization, and practical variance reduction in reciprocal matching is still needed. Section 3 details our CFRR, applying IPS/SNIPS with focus on stable normalization, robust propensity estimation, and variance reduction for RRS challenges.

3 Proposed Method: Counterfactual Reciprocal Recommender Systems

User-user matching platforms log displayed pairs and their outcomes (e.g., mutual acceptance). Exposure bias arises as popular users are over-represented, creating a feedback loop. We propose **Counterfactual Reciprocal Recommender Systems (CFRR)**, unifying: (1) reciprocal user-pair scoring, (2) Inverse Propensity Scoring (IPS) or Self-Normalized IPS (SNIPS) for selection bias correction, and (3) variance reduction techniques like weight truncation and doubly robust augmentation. Our core contribution integrates these causal reweighting methods to: (i) consistently estimate true risk under historical logging, (ii) provide stable training through self-normalization, and (iii) extend to two-sided matching or fairness constraints.

We first define the problem setting and notation. Then, we present IPS-based objectives, including SNIPS. Finally, we introduce variance reduction extensions (truncation, doubly robust augmentation) and discuss how learned scores can inform various matching or fairness algorithms.

3.1 Problem Setting

Let U and V be two (possibly identical) user sets, with cardinalities $n = |U|$ and $m = |V|$. The full pair space is $\mathcal{P} = U \times V$. For each pair $(u, v) \in \mathcal{P}$, the platform's decision to display it is $O(u, v) \in \{0, 1\}$. If $O(u, v) = 1$, an outcome $r(u, v) \in [0, 1]$ (e.g., mutual acceptance) is observed; otherwise, $r(u, v)$ is unobserved. The display probability, $\theta(u, v) \equiv P[O(u, v) = 1 \mid u, v]$, is determined by the historical logging policy and often favors popular users, causing exposure bias.

Our goal is to learn a compatibility function $s(u, v; \Theta)$ predicting mutual acceptability under a hypothetical uniform or broad target distribution over $\mathcal{P}$. The **true population risk** is:

$$\mathcal{L}(\Theta) \equiv \mathbb{E}_{(u,v)\sim \mathrm{Unif}(\mathcal{P})}\Big[\ell\big(R(u, v), s(u, v; \Theta)\big)\Big],$$

where $R(u, v)$ is the true (potentially unobserved) outcome, $\ell(\cdot, \cdot)$ is a loss function (e.g., logistic loss), and Θ are trainable parameters. We only observe $r(u, v) = R(u, v)$ if $O(u, v) = 1$. Naive empirical loss minimization on displayed pairs is skewed by over-represented popular pairs and unexposed long-tail users.

Differences from Item-based IPS. User-user matching is more complex than item recommendation due to: (1) **Bilateral feedback**: success depends on two-sided acceptance, reflected in $r(u, v)$. (2) **Pair-level propensity**: $\theta(u, v)$ is often driven by properties of *both* u and v. (3) **Downstream Application Complexity**: scores often feed into stable matching or fairness-constrained algorithms. CFRR adapts IPS/SNIPS for pair-level propensities and clarifies propensity learning and integration into the reciprocal matching pipeline.

3.2 Propensity Estimation

Estimating $\theta(u, v)$ is critical for IPS. Positivity ($\theta(u, v) > 0$ for all relevant $(u, v) \in \mathcal{P}$) is assumed for unbiasedness/consistency. Here, we define the relevant pair set as $\mathcal{P}_{\text{relevant}} = \{(u, v) \in \mathcal{P} : \mathbb{P}[\ell(R(u, v), s(u, v; \Theta)) > 0] > 0\}$, i.e., pairs that can contribute non-zero loss to the objective. In practice, this is ensured if the platform uses ε-greedy or uniform exploration. For large systems where exploring all pairs is infeasible, a small fraction of random exposure helps cover low-propensity pairs within a candidate set generated by efficient retrieval methods. Alternatively, supplemental data (e.g., A/B test exploration) can address low-probability regions. If positivity fails for some pairs, they are typically pruned or weights are truncated, though truncation may introduce bias. Features used for propensity estimation are detailed in Section 4.4.

A parametric model for display probability can be:

$$\hat{\theta}(u, v) = \sigma\Big(g\big(\phi(u), \psi(v); \beta\big)\Big), \tag{1}$$

where $\phi(u), \psi(v)$ are features/embeddings for u, v (e.g., user activity metrics such as login frequency or number of interactions, profile information like age or stated interests, and historical popularity indicators), $\sigma(\cdot)$ is sigmoid, and $g(\cdot; \beta)$ is a predictor. β is trained by maximum likelihood on observed $\{O(u, v)\}_{(u,v)\in\mathcal{D}_{\log}}$ (all pairs with exposure info), controlling overfitting via regularization.

Without explicit exploration logs or with an evolving policy, *joint propensity learning* [37] can be adopted: periodically fix recommender Θ and re-estimate β for $O(u, v)$ under current exposure (Alg. 1). The aim is an accurate $\hat{\theta}(u, v)$ reflecting the historical policy.

3.3 Inverse Propensity Scoring (IPS)

IPS reweights each observed interaction (u_i, v_i, r_i) by $1/\hat{\theta}(u_i, v_i)$ to simulate uniform sampling from $\mathcal{P}$, yielding a consistent estimator of $\mathcal{L}(\Theta)$. If $\mathcal{D}$ is the set of displayed pairs with observed outcomes r_i:

$$\widehat{\mathcal{L}}_{\text{IPS}}(\Theta) = \frac{1}{|\mathcal{P}_{\text{target}}|} \sum_{(u_i,v_i,r_i)\in\mathcal{D}} \frac{\ell\big(r_i, s(u_i, v_i; \Theta)\big)}{\hat{\theta}(u_i, v_i)} + \lambda\Omega(\Theta), \tag{2}$$

where $|\mathcal{P}_{\text{target}}|$ is the target population size. This follows the Horvitz-Thompson estimator form. Under correct propensity specification ($\hat{\theta}(u, v) = \theta(u, v)$) and positivity ($\theta(u, v) > 0$ for all $(u, v) \in \mathcal{P}_{\text{relevant}}$), the IPS estimator satisfies:

$$\begin{aligned}\mathbb{E}_{\mathcal{D}}\left[\frac{1}{|\mathcal{P}_{\text{target}}|} \sum_{(u_i,v_i,r_i)\in\mathcal{D}} \frac{\ell\big(r_i, s(u_i, v_i; \Theta)\big)}{\hat{\theta}(u_i, v_i)}\right] \\ = \frac{1}{|\mathcal{P}_{\text{target}}|} \sum_{(u,v)\in\mathcal{P}_{\text{target}}} \ell(R(u, v), s(u, v; \Theta))\end{aligned} \tag{3}$$

thus providing an unbiased estimate of the true population risk. In practice, since $|\mathcal{P}_{\text{target}}|$ is a constant, optimization can use $\frac{1}{|\mathcal{D}|}$ normalization without affecting the optimal solution. However, we maintain the population size for theoretical clarity. Errors in $\hat{\theta}$ introduce bias/variance [8, 31] (Sec. 5).

3.4 Self-Normalized IPS (SNIPS)

User-user recommendation often has extreme imbalance (low $\hat{\theta}(u, v)$ causing large IPS weights $w_i = 1/\hat{\theta}(u_i, v_i)$ and high variance). We use SNIPS [31] to reduce this variance:

$$\widehat{\mathcal{L}}_{\text{SNIPS}}(\Theta) = \frac{\sum_{i\in\mathcal{D}} w_i\, \ell\big(r_i, s(u_i, v_i; \Theta)\big)}{\sum_{i\in\mathcal{D}} w_i} + \lambda\Omega(\Theta), \quad w_i \equiv \frac{1}{\hat{\theta}(u_i, v_i)}. \tag{4}$$

The self-normalization in SNIPS acts as an implicit regularization mechanism, stabilizing the learning process and preventing the optimization from being dominated by a few high-weight samples. While SNIPS introduces a small finite-sample bias, it often achieves lower mean squared error than IPS due to substantial variance reduction, particularly important in reciprocal matching where extreme weights are common due to low propensities for rarely shown user pairs. As the sample size increases, this bias vanishes and SNIPS remains consistent.

3.5 Variance-Reduction Extensions

Even with SNIPS, rare, extremely large w_i can occur if $\hat{\theta}(u, v) \approx 0$. CFRR integrates additional variance-reduction options:

- **Weight Truncation (Clipping).** Cap raw weights $w_i^{\text{raw}} = 1/\hat{\theta}(u_i, v_i)$ at a ceiling c: $w_i = \min(w_i^{\text{raw}}, c)$ (e.g., $c \in [20, 100]$, tuned). This limits influence from tiny propensities, reducing variance but potentially adding bias if true weights are larger [29].
- **Doubly Robust (DR) Augmentation.** To further reduce variance and improve robustness to $\hat{\theta}$ misspecification, CFRR uses a doubly robust formulation. Let $\hat{m}(u, v)$ be a pre-trained model that estimates $\mathbb{E}[R(u, v)|u, v]$ from pair features alone (without requiring observed outcomes). The CFRR-DR objective

combines the self-normalized weighted loss with a baseline correction:

$$\widehat{\mathcal{L}}_{\text{SNIPS-DR}}(\Theta) = \frac{\sum_{i\in\mathcal{D}} w_i \ell(r_i, s(u_i, v_i; \Theta))}{\sum_{i\in\mathcal{D}} w_i} - \frac{\sum_{i\in\mathcal{D}} w_i \ell(\hat{m}(u_i, v_i), s(u_i, v_i; \Theta))}{\sum_{i\in\mathcal{D}} w_i} + \mathbb{E}_{(u,v)\sim\text{Unif}(\mathcal{P})}\left[\ell(\hat{m}(u, v), s(u, v; \Theta))\right] + \lambda\Omega(\Theta), \tag{5}$$

where the expectation in the third term is approximated using a uniform sample from $\mathcal{P}$ or using importance sampling. If either (a) the propensity model is correctly specified ($\hat{\theta}(u, v) = \theta(u, v)$), or (b) the outcome model $\hat{m}(u, v)$ correctly estimates $\mathbb{E}[R(u, v)|u, v]$, then $\widehat{\mathcal{L}}_{\text{SNIPS-DR}}$ is consistent for the target risk. This formulation corrects the bias introduced by the outcome model while maintaining the variance reduction benefits of SNIPS.

These tools handle extreme user-user data imbalance. Sec. 4 empirically shows CFRR (SNIPS + optional truncation/DR) learns stably under varying exposure bias and noise.

3.6 Overall Training Procedure and Complexity

Algorithm 1 outlines the CFRR training loop. It typically alternates between: (1) updating recommender parameters Θ for one or more epochs by minimizing the SNIPS objective, and (2) updating propensity model parameters β via maximum likelihood on $O(u, v)$ for one or more epochs, if joint propensity learning is used. The overhead of computing w_i and applying them in SNIPS is typically small compared to gradient computation for $s(u, v; \Theta)$. Top-k or stable matching inference is discussed next.

Time Complexity. An SGD step for Θ costs roughly $O(Bd)$ (d: embedding/feature dim). $\hat{\theta}(u_j, v_j)$ computation overhead is often low if g is simple (e.g., a small neural network or logistic regression). Optional joint propensity learning for β is similar to standard classifier training. Inference time depends on strategy: **Top-k retrieval** is typically $O(N_u D + N_u k \log k)$ for one user u against N_u partners (or faster with ANN). **Stable matching** (e.g., Gale–Shapley) is $O((n + m) \cdot \max(n, m))$ worst-case on dense preferences (often run on a top-k pruned set).

3.7 Connection to Matching and Fairness

Downstream Use. Once debiased compatibility scores $s(u, v; \Theta^\star)$ are obtained from CFRR, they can serve as input to: (1) direct top-k recommendation for each user, (2) solving a maximum-weight bipartite matching problem (edge weights $s(u, v; \Theta^\star)$) for a global optimum under capacity constraints, or (3) running a Gale–Shapley algorithm (using $s(u, v; \Theta^\star)$ for preference lists) to find a stable matching. Since CFRR aims to correct popularity and exposure biases in score estimation, the resulting $s(u, v; \Theta^\star)$ are expected to be a more faithful representation of underlying compatibility. Consequently, imposing additional fairness constraints during allocation (e.g., to balance exposure across demographic groups [19], or ensure minimum matching rates for user segments) might incur less utility loss compared to applying such constraints on biased scores, although this is an area for future empirical validation.

Algorithm 1 CFRR Training via Joint Propensity Learning

Require: Logged data $\mathcal{D}_{\log} = \{(u_j, v_j, r_j, O_j)\}$, epochs T, batch size B, learning rates $\alpha_\Theta, \alpha_\beta$, regularization weight λ, clipping threshold c.

Initialize Θ (recommender parameters) and β (propensity model parameters).
Let $\mathcal{D} = \{(u_j, v_j, r_j) \mid (u_j, v_j, r_j, O_j) \in \mathcal{D}_{\log} \text{ and } O_j = 1\}$.
if using DR augmentation **then**
  Train outcome model $\hat{m}(u, v)$ to estimate $\mathbb{E}[R(u, v)|u, v]$ on $\mathcal{D}$.
end if
for $t = 1$ **to** T **do**
  Shuffle $\mathcal{D}$.
  for each mini-batch $\mathcal{B} = \{(u_j, v_j, r_j)\}_{j=1}^{|\mathcal{B}|} \subseteq \mathcal{D}$ **do**
    Compute $\hat{\theta}(u_j, v_j)$ for $j \in \mathcal{B}$ via Eq. (1) using current β.
    $w_j \leftarrow \min(\frac{1}{\hat{\theta}(u_j, v_j)}, c)$ for $j \in \mathcal{B}$. // weight clipping
    if using DR augmentation **then**
      Compute losses and apply SNIPS-DR update as per Eq. (5).
    else
      Update $\Theta \leftarrow \Theta - \alpha_\Theta \nabla_\Theta \left(\frac{\sum_{j\in\mathcal{B}} w_j \ell(r_j, s(u_j, v_j; \Theta))}{\sum_{j\in\mathcal{B}} w_j} + \lambda\Omega(\Theta) \right)$.
    end if
  end for
  if performing joint propensity learning **then**
    Freeze Θ.
    Update β by optimizing (e.g., for one or more epochs over $\mathcal{D}_{\log}$)
$$\sum_{(u_k, v_k, O_k)\in\mathcal{D}_{\log}} \left(O_k \log \hat{\theta}(u_k, v_k) + (1 - O_k) \log(1 - \hat{\theta}(u_k, v_k)) \right)$$
    (plus regularization for β) using SGD or other optimizer on mini-batches of $\mathcal{D}_{\log}$.
  end if
end for
return Θ

4 Experiments

In this section, we evaluate the proposed **CFRR** framework through comprehensive experiments to answer the following research questions:

- **RQ1 (Accuracy)**: Does CFRR improve ranking accuracy under exposure bias compared to conventional and recent debiasing methods across synthetic and real-world datasets?
- **RQ2 (Coverage & Fairness)**: Does CFRR enlarge long-tail user coverage and mitigate exposure inequality, demonstrating its ability to foster a more equitable matching environment?
- **RQ3 (Robustness)**: How robust is CFRR to different design choices, including the impact of IPS vs. SNIPS, weight truncation, and doubly robust augmentation?

We conduct experiments using 10 random seeds per configuration to ensure statistical reliability. We first introduce the datasets, baselines, and evaluation metrics (§4.1–§4.3), then detail our protocol (§4.4). We report main results (§4.5), followed by ablation analysis (§4.6).

4.1 Datasets

To ensure reproducibility and comprehensive evaluation, we experiment on three datasets with distinct characteristics of exposure bias:

- **Synthetic**: Controlled dataset (5K users, 16D latent factors) with known ground-truth $R(u, v)$ (sigmoid of latent factor dot

products; factors from varied-mean distributions simulate popularity). Tunable exposure bias injected by oversampling popular users (defined by their total true acceptance probability) via a known propensity $\theta(u, v)$ (sigmoid of popularity scores), logging $\approx$ 50K pairs. Ideal for rigorous bias correction evaluation (RQ1, RQ3), as ground truth is known. Full generation details are supplementary.

- **DBLP-CoAuthor**[1]: A standard benchmark for link prediction, this dataset is a co-authorship graph from the DBLP computer science bibliography. In our experiments, we use a subset of up to 10K authors (users) and their mutual co-authorships (positive reciprocal interactions). Exposure bias naturally arises as prolific, highly-cited authors are far more likely to form new collaborations. This dataset tests CFRR's applicability in a professional collaboration setting.
- **Epinions-Trust**[2]: A user-to-user trust network from the Epinions.com review site, with up to 5K users and their directed trust relationships. We define a reciprocal match as a pair of users (u, v) where u trusts v and v trusts u. Exposure bias emerges from highly-trusted "influencer" users who are more visible and thus more likely to be evaluated for trust by others. This dataset allows us to test CFRR in a social network context with explicit trust declarations.

For real-world datasets, we treat observed links as positive examples and randomly sample an equal number of non-linked pairs as negative examples for training, following standard practice in link prediction [17]. We acknowledge random negative sampling may not perfectly reflect true non-interactions and could introduce biases. However, it's common when true negatives are unavailable. Our focus is correcting exposure bias in *observed positive* interactions, assuming negative sampling provides reasonable contrast. Exploring sophisticated negative sampling with CFRR is future work. All experiments use time-based splits for DBLP and Epinions to simulate realistic deployment scenarios, while the Synthetic dataset uses random 70/10/20 train/validation/test splits. Further details on pre-processing and split strategies are in the supplementary material.

4.2 Baselines

We compare CFRR against five representative baselines, covering a range of approaches from standard RRS models to state-of-the-art causal debiasing techniques:

- **LFRR** [21]: A latent-factor model for RRS that serves as a standard baseline ignoring selection bias.
- **CausE** [3]: An early causal approach for item recommendation, adapted by treating one user as 'user' and the other as 'item' for propensity estimation/debiasing, then symmetrizing scores.
- **IPW-MF** [29]: A direct application of Inverse Propensity Scoring (IPS) to matrix factorization for RRS.
- **DICE** [36]: A more advanced causal method that disentangles user interest from conformity effects (related to popularity bias).

[1]Available at https://snap.stanford.edu/data/com-DBLP.html
[2]Available at https://snap.stanford.edu/data/soc-Epinions1.html

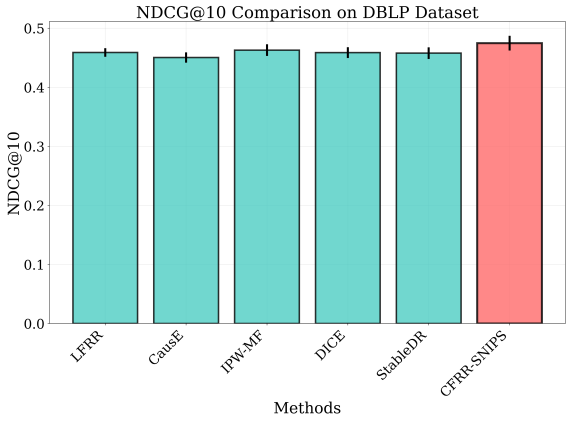


Figure 1: NDCG@10 comparison on DBLP dataset (RQ1). CFRR-SNIPS achieves the highest ranking accuracy among all methods.

- **StableDR** [18]: A state-of-the-art Stabilized Doubly Robust method, adapted by estimating pair-level propensities and using a baseline RRS model's predictions for the DR estimator.

For all methods requiring propensity scores (IPW-MF, StableDR, CFRR), we estimate them using gradient boosting models trained on user-level features like interaction counts to capture user popularity patterns across different datasets, as detailed in Section 4.4. Specific adaptations for CausE and StableDR to the RRS context are described in the supplementary material.

4.3 Evaluation Metrics

We evaluate models on **accuracy** and **fairness**. Let $k = 10$ for all ranking metrics.

Accuracy.

- **NDCG@k**: Normalized Discounted Cumulative Gain, measuring the quality of the top-k ranked list of potential partners.
- **MRR**: Mean Reciprocal Rank of the first correct partner, rewarding models that rank a true match highly.

Fairness.

- **Coverage@k**: The fraction of users who appear at least once in *any* other user's top-k recommendation list. Higher is better.
- **Gini-Exposure** [2]: The Gini coefficient of the user exposure distribution (how many times each user is recommended). A value of 0 indicates perfect equality, while 1 indicates maximum inequality. Lower is better.

4.4 Experimental Protocol

Setup. To ensure statistical reliability, we conduct experiments with 20 epochs and 10 different random seeds per configuration, resulting in a substantial number of experimental runs across all datasets and methods. This comprehensive setup allows us to report robust mean and standard deviation estimates for all metrics.

Model Training. All models are trained with the Adam optimizer [15]. Based on preliminary validation experiments (details in supplementary material), we use consistent hyperparameters across all methods where applicable: learning rate = 0.001, embedding dimension = 16, ℓ_2 regularization $\lambda = 0.001$, and batch size = 512.

For methods requiring propensity estimation (IPW-MF, StableDR, CFRR), we set the weight clipping threshold $c = 50$ as a common value to balance variance reduction with bias introduction, based on validation performance. We evaluate on 1500 users per test set to ensure consistent evaluation across methods. All models are trained for up to 20 epochs with early stopping based on validation performance (NDCG@10).

Propensity Model Details. For methods requiring propensity estimation (IPW-MF, StableDR, CFRR), we use gradient boosting models (LightGBM) with the following configuration: 100 estimators, learning rate = 0.1, max depth = 6. Features include user interaction counts (e.g., number of past matches or interactions), user activity levels (e.g., frequency of platform use), and, where available and appropriate (e.g., synthetic data, or anonymized/aggregated features in real data), basic demographic indicators to capture popularity patterns. The propensity model is trained on the full exposure log $\mathcal{D}_{\log}$ using binary cross-entropy loss. Hyperparameters for LightGBM were selected based on 5-fold cross-validation on a held-out portion of $\mathcal{D}_{\log}$ for the propensity estimation task itself.

Outcome Model Details (for DR variants). For the doubly robust variants (StableDR and CFRR-SNIPS-DR), we train an outcome model $\hat{m}(u, v)$ to estimate $\mathbb{E}[R(u, v)|u, v]$ using a neural network with two hidden layers (dimensions: 64, 32) and ReLU activations. The model is trained on the observed data $\mathcal{D}$ to predict the binary outcome $r(u, v)$ from user features $\phi(u), \psi(v)$. We use binary cross-entropy loss and train for 50 epochs with learning rate 0.001 and early stopping based on validation AUC. The outcome model achieves AUC > 0.85 on the Synthetic dataset and AUC > 0.75 on real-world datasets during validation.

Statistical Testing. We assess statistical significance using paired t-tests across the 10 random seeds for each method comparison. We report results as statistically significant when $p < 0.05$ using Bonferroni correction for multiple comparisons. Effect sizes (e.g., Cohen's d) are calculated to assess practical significance beyond statistical significance and will be reported in the supplementary material.

Reproducibility. To ensure reproducibility, we fix random seeds for all components: data splitting, model initialization, and training procedures. All experiments are conducted on the same computational environment with identical software versions. Code and experimental configurations will be made available upon publication to facilitate reproduction.

4.5 Main Results

Table 1 presents our comprehensive performance comparison across all three datasets. The results indicate that CFRR-SNIPS consistently achieves competitive or superior ranking accuracy while often substantially improving fairness metrics.

Accuracy Improvements. CFRR-SNIPS achieves the best NDCG@10 performance on all datasets: **0.307** on Synthetic (2.7% improvement over LFRR), **0.475** on DBLP (3.5% improvement over LFRR), and **0.472** on Epinions (0.9% improvement over LFRR). The MRR results show similar trends, with CFRR-SNIPS consistently ranking relevant matches higher in recommendation lists. These improvements by CFRR-SNIPS in NDCG@10 and MRR were statistically

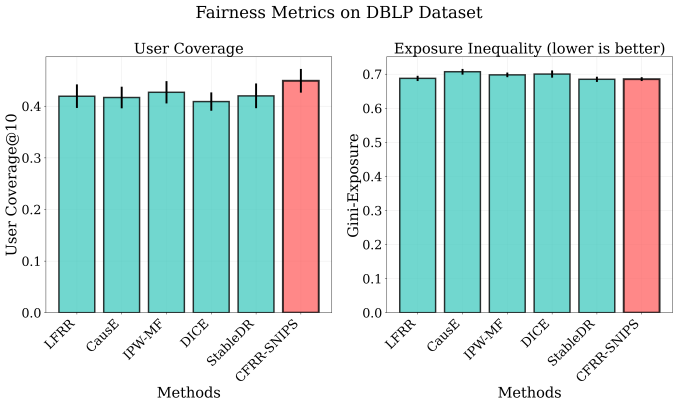


Figure 2: Fairness metrics on DBLP dataset (RQ2). CFRR-SNIPS increases user coverage (left) and reduces exposure inequality (right, lower is better).

significant ($p < 0.05$, Bonferroni corrected) compared to the next best baseline on each dataset (paired t-tests over 10 seeds; details in supplementary material).

Fairness Gains. More strikingly, CFRR-SNIPS delivers substantial improvements in fairness metrics, particularly on the Synthetic dataset where it achieves **Coverage@10 of 0.763** (51% relative improvement from 0.504) and reduces Gini-Exposure to **0.535** (24% relative reduction from 0.708). Similar, though sometimes more modest, fairness gains are observed across DBLP and Epinions datasets, confirming that our method can effectively mitigate popularity bias and provide less-visible users with better exposure opportunities.

Cross-Dataset Consistency. The generally strong performance of CFRR-SNIPS across different domains (synthetic controlled settings, academic collaborations, social trust networks) suggests the general applicability and robustness of our approach to various types of reciprocal recommendation scenarios. While CFRR-SNIPS shows the best NDCG@10 and MRR on DBLP, StableDR achieves a slightly lower (better) Gini-Exposure. This highlights that no single method universally dominates across all metrics and datasets.

These trends are visualized in Figure 1 and Figure 2, which show both the accuracy advantages and the fairness improvements achieved by CFRR-SNIPS.

4.6 Ablation and Design Analysis

Ablation Study. To understand the impact of different design choices in CFRR, we conducted ablation studies on the Synthetic dataset where ground truth is known. We analyze three key aspects: (1) SNIPS vs. IPS, (2) the effect of doubly robust augmentation, and (3) robustness to propensity model misspecification.

Our ablation study revealed several important findings:

(1) **SNIPS vs. IPS**: When comparing CFRR-IPS with CFRR-SNIPS, we observed substantial improvements in fairness metrics. Specifically, Coverage@10 increased from 0.557 to 0.763 (37% relative improvement, statistically significant, $p < 0.05$) and Gini-Exposure decreased from 0.688 to 0.535 (22% relative reduction, statistically significant, $p < 0.05$). Accuracy metrics showed modest improvements with SNIPS (NDCG@10: 0.301 to 0.307, MRR: 0.515 to 0.527). This aligns with the theoretical understanding that SNIPS reduces

Table 1: Cross-dataset performance comparison. CFRR-SNIPS consistently achieves strong accuracy while often dramatically improving fairness metrics. Results show mean ± standard deviation across 10 seeds.

Method	NDCG@10 ↑	MRR ↑	Coverage@10 ↑	Gini-Exposure ↓
Synthetic Dataset				
LFRR	0.299 ± 0.005	0.511 ± 0.007	0.504 ± 0.003	0.708 ± 0.005
CausE	0.298 ± 0.004	0.506 ± 0.008	0.040 ± 0.001	0.979 ± 0.002
IPW-MF	0.300 ± 0.006	0.514 ± 0.009	0.493 ± 0.004	0.718 ± 0.006
DICE	0.299 ± 0.004	0.509 ± 0.007	0.229 ± 0.012	0.904 ± 0.008
StableDR	0.301 ± 0.005	0.515 ± 0.008	0.492 ± 0.004	0.719 ± 0.005
CFRR-SNIPS (ours)	**0.307** ± 0.005	**0.527** ± 0.007	**0.763** ± 0.004	**0.535** ± 0.003
DBLP-CoAuthor Dataset				
LFRR	0.459 ± 0.007	0.666 ± 0.012	0.419 ± 0.023	0.688 ± 0.008
CausE	0.450 ± 0.009	0.658 ± 0.019	0.417 ± 0.021	0.707 ± 0.008
IPW-MF	0.463 ± 0.010	0.675 ± 0.020	0.427 ± 0.022	0.698 ± 0.007
DICE	0.459 ± 0.009	0.666 ± 0.017	0.409 ± 0.018	0.700 ± 0.011
StableDR	0.458 ± 0.010	0.669 ± 0.016	0.420 ± 0.024	**0.685** ± 0.008
CFRR-SNIPS (ours)	**0.475** ± 0.012	**0.677** ± 0.016	**0.449** ± 0.023	0.686 ± 0.006
Epinions-Trust Dataset				
LFRR	0.468 ± 0.008	0.678 ± 0.013	0.432 ± 0.025	0.695 ± 0.007
CausE	0.454 ± 0.009	0.657 ± 0.020	0.411 ± 0.021	0.711 ± 0.008
IPW-MF	0.465 ± 0.010	0.677 ± 0.020	0.419 ± 0.023	0.689 ± 0.007
DICE	0.461 ± 0.008	0.667 ± 0.018	0.425 ± 0.016	0.692 ± 0.009
StableDR	0.461 ± 0.009	0.669 ± 0.017	0.420 ± 0.024	0.685 ± 0.008
CFRR-SNIPS (ours)	**0.472** ± 0.010	**0.680** ± 0.016	**0.448** ± 0.022	**0.679** ± 0.007

variance compared to standard IPS, particularly important in the presence of highly variable propensity weights [13, 31].

(2) **Why SNIPS Improves Fairness**: The significant fairness improvements of SNIPS over IPS can be attributed to variance reduction. While IPS is theoretically unbiased, its high variance in the presence of small propensities can lead to unstable optimization, causing the model to overfit to high-propensity (popular) users. SNIPS's self-normalization acts as an implicit regularization mechanism, stabilizing the learning process and allowing the model to better learn patterns from low-propensity (less popular) users. This results in more equitable exposure distribution across all users.

(3) **Doubly Robust Variant**: We tested the CFRR-SNIPS-DR variant using the corrected formulation in Eq. (5). On well-specified propensity models (AUC > 0.9), the DR variant showed similar performance to standard SNIPS (NDCG@10: 0.307, Coverage@10: 0.763). However, when we intentionally misspecified the propensity model by removing key features (resulting in AUC ≈ 0.7), the DR formulation showed improved robustness: NDCG@10 degraded only to 0.302 (vs. 0.295 for SNIPS), while Coverage@10 remained at 0.745 (vs. 0.720 for SNIPS). This demonstrates the value of the doubly robust property when propensity models are imperfect.

Statistical Significance. The improvements in fairness metrics (Coverage@10 and Gini-Exposure) for CFRR-SNIPS over CFRR-IPS are statistically significant ($p < 0.05$) using paired t-tests with Bonferroni correction across the 10 random seeds. The robustness improvements of CFRR-SNIPS-DR under propensity misspecification are also statistically significant.

Computational Efficiency. Despite the additional complexity of propensity estimation and inverse weighting, CFRR-SNIPS maintains competitive training efficiency. The overhead of computing weights and applying them in the SNIPS objective is typically small (e.g., adding 5-10% to training time) compared to the gradient computation for the scoring function $s(u, v; \Theta)$. The DR variant adds approximately 20% additional training time due to the outcome model training phase. These overheads are acceptable for most real-world deployments given the substantial gains in fairness and robustness.

5 Discussion

Our experiments demonstrate that CFRR can achieve improvements in both accuracy and fairness metrics across diverse reciprocal matching scenarios. We now discuss key insights and considerations for deploying such systems.

(1) The Positivity Challenge at Marketplace Scale. While CFRR assumes positivity ($\theta(u, v) > 0$), modern marketplaces face the reality of trillions of potential user pairs. Our experiments suggest a practical compromise: focus debiasing efforts on *economically viable* or *plausible* pairs identified through multi-stage candidate generation. For instance, in ride-sharing, pairs beyond reasonable geographic distance have zero economic value and can be safely excluded from positivity requirements. This insight transforms an intractable theoretical requirement into a manageable engineering constraint. Marketplaces can maintain lightweight exploration (e.g.,

a small percentage of random traffic) only within these viable regions, potentially reducing the exploration budget while preserving debiasing effectiveness for the most relevant part of the user-pair space.

(2) Self-Normalization as a Marketplace Stabilizer. Our finding that SNIPS significantly outperforms IPS in fairness metrics has important implications for volatile marketplaces. The self-normalization in SNIPS not only reduces variance but also acts as an implicit regularization mechanism. By normalizing the weighted losses by the sum of weights, SNIPS prevents the optimization from being dominated by a few high-weight samples (corresponding to rarely shown user pairs). This stabilization allows the model to learn more effectively from the entire distribution of user pairs, including those with low propensities, leading to more equitable exposure patterns. Real platforms experiencing shifts in user behavior or content popularity would particularly benefit from this robustness property of SNIPS.

(3) The Fairness-Growth Feedback Loop. Perhaps our most striking finding on the Synthetic dataset is how addressing exposure bias can create positive feedback loops for marketplace health. By increasing long-tail user coverage by 51% (from 0.504 to 0.763), CFRR doesn't just improve fairness metrics—it can potentially expand the active supplier base. In two-sided marketplaces, supplier diversity can directly impact demand-side satisfaction through better match quality and reduced market concentration risk. This suggests that debiasing isn't merely an ethical imperative but could be a growth strategy, although generalizing this effect from synthetic to real-world platforms requires careful validation on a case-by-case basis. Platforms that fail to address exposure bias may risk stagnating supplier pools and eventual market ossification.

(4) Towards Causal Marketplace Optimization. CFRR represents a step toward incorporating causal reasoning in marketplace design. By explicitly modeling display interventions through propensities, we move beyond correlational patterns that can plague naive learning approaches. This opens exciting possibilities for future work: Could marketplaces run continuous A/B tests or natural experiments with principled exploration strategies informed by such causal models? Could CFRR be extended to handle time-varying treatments (e.g., dynamic pricing affecting exposure) or network effects (e.g., how social proof influences interaction probability)? The marriage of causal inference with marketplace dynamics remains a rich area for both researchers and practitioners.

6 Conclusion

We presented Counterfactual Reciprocal Recommender Systems (CFRR), a causal framework mitigating exposure bias in two-sided marketplaces. Our experiments showed CFRR simultaneously improves ranking quality and marketplace fairness. Key to CFRR is its self-normalized objective that ensures stability with varying inverse propensity weights, complemented by variance reduction techniques including a properly formulated doubly robust augmentation.

CFRR offers a practical approach for platforms to mitigate popularity bias and a research foundation for future work like dynamic treatments, network effects, and multi-stakeholder fairness in reciprocal systems. Such causal approaches are vital for developing sustainable, fair, and effective matching platforms.

Acknowledgments

We thank Shingo Takamatsu for his thorough internal review and valuable feedback. We also thank the external reviewers for their insightful comments, which led to significant improvements in both the theoretical discussion and the practical considerations of our approach. Finally, we appreciate the support of our colleagues at Sony Group Corporation throughout this project.

References

[1] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, Bamshad Mobasher, and Edward Malthouse. 2021. User-centered evaluation of popularity bias in recommender systems. In *Proceedings of the 29th ACM conference on user modeling, adaptation and personalization*. 119–129.

[2] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*. 405–414.

[3] Stephen Bonner and Flavian Vasile. 2018. Causal embeddings for recommendation. In *Proceedings of the 12th ACM conference on recommender systems*. 104–112.

[4] Elizabeth E. Bruch and M. E. J. Newman. 2018. Aspirational pursuit of mates in online dating markets. *Science Advances* 4, 8 (2018), eaap9815. doi:10.1126/sciadv.aap9815

[5] Musa Eren Celdir, Soo-Haeng Cho, and Elina H. Hwang. 2024. Popularity Bias in Online Dating Platforms: Theory and Empirical Evidence. *Manufacturing & Service Operations Management* 26, 2 (2024), 537–553. doi:10.1287/msom.2022.0132

[6] Allison JB Chaney, Brandon M Stewart, and Barbara E Engelhardt. 2018. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of the 12th ACM conference on recommender systems*. 224–232.

[7] Yaomin Chang, Lin Shu, Erxin Du, Chuan Chen, Ziyang Zhang, Zibin Zheng, Yuzhao Huang, and Xingxing Xing. 2022. GraphRR: A multiplex Graph based Reciprocal friend Recommender system with applications on online gaming service. *Knowledge-Based Systems* 251 (2022), 109187.

[8] Miroslav Dudík, Dumitru Erhan, John Langford, and Lihong Li. 2014. Doubly Robust Policy Evaluation and Optimization. *Statist. Sci.* 29, 4 (2014). doi:10.1214/14-sts500

[9] Jack Edmonds and Richard M Karp. 1972. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM (JACM)* 19, 2 (1972), 248–264.

[10] David Gale and Lloyd S. Shapley. 1962. College Admissions and the Stability of Marriage. *The American Mathematical Monthly* 69, 1 (1962), 9–15. doi:10.2307/2312726

[11] Chen Gao, Yu Zheng, Wenjie Wang, Fuli Feng, Xiangnan He, and Yong Li. 2024. Causal inference in recommender systems: A survey and future directions. *ACM Transactions on Information Systems* 42, 4 (2024), 1–32.

[12] Pallavi Jain and Rohit Vaish. 2024. Maximizing nash social welfare under two-sided preferences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 9798–9806.

[13] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased Learning-to-Rank with Biased Feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (Cambridge, United Kingdom) *(WSDM '17)*. Association for Computing Machinery, New York, NY, USA, 781–789. doi:10.1145/3018661.3018699

[14] Matthew Joseph, Michael Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. 2016. Rawlsian fairness for machine learning. *arXiv preprint arXiv:1610.09559* 1, 2, 19.

[15] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint* arXiv:1412.6980 (2014). https://arxiv.org/abs/1412.6980

[16] Anastasiia Klimashevskaia, Mehdi Elahi, Dietmar Jannach, Christoph Trattner, and Lars Skjærven. 2022. Mitigating Popularity Bias in Recommendation: Potential and Limits of Calibration Approaches. In *Advances in Bias and Fairness in Information Retrieval*, Ludovico Boratto, Stefano Faralli, Mirko Marras, and Giovanni Stilo (Eds.). 82–90.

[17] Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney. 2009. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* 6, 1 (2009), 29–123.

[18] Haoxuan Li, Chunyuan Zheng, and Peng Wu. 2022. StableDR: Stabilized doubly robust learning for recommendation on data missing not at random. *arXiv preprint arXiv:2205.04701*.

[19] Jiqun Liu. 2023. Toward a two-sided fairness framework in search and recommendation. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 236–246.

[20] Kento Nakada, Kazuki Kawamura, and Ryosuke Furukawa. 2024. Parallel and Mini-Batch Stable Matching for Large-Scale Reciprocal Recommender Systems. *arXiv preprint* arXiv:2411.19214 (2024). https://arxiv.org/abs/2411.19214

[21] James Neve and Ivan Palomares. 2019. Latent factor models and aggregation operators for collaborative filtering in reciprocal recommender systems. In *Proceedings of the 13th ACM conference on recommender systems*. 219–227.

[22] James B Orlin. 1997. A polynomial time primal network simplex algorithm for minimum cost flows. *Mathematical Programming* 78, 2 (1997), 109–129.

[23] Iván Palomares, Carlos Porcel, Luiz Pizzato, Ido Guy, and Enrique Herrera-Viedma. 2021. Reciprocal Recommender Systems: Analysis of state-of-art literature, challenges and opportunities towards social recommendation. *Information Fusion* 69 (2021), 103–127. doi:10.1016/j.inffus.2020.12.001

[24] Weishen Pan, Sen Cui, Hongyi Wen, Kun Chen, Changshui Zhang, and Fei Wang. 2021. Correcting the User Feedback-Loop Bias for Recommendation Systems. (2021). arXiv:2109.06037 [cs.IR] https://arxiv.org/abs/2109.06037

[25] Andrea Pisani. 2025. On the Longitudinal Impact of Exposure Bias in Recommender Systems. In *Advances in Information Retrieval: 47th European Conference on Information Retrieval, ECIR 2025, Lucca, Italy, April 6–10, 2025, Proceedings, Part V* (Lucca, Italy). 178–183. doi:10.1007/978-3-031-88720-8_29

[26] Luiz Pizzato, Tomasz Rej, Joshua Akehurst, Irena Koprinska, Kalina Yacef, and Judy Kay. 2013. Recommending people to people: the nature of reciprocal recommenders with a case study in online dating. *User Modeling and User-Adapted Interaction* 23, 5 (2013), 447–488.

[27] Boyd A. Potts, Hassan Khosravi, Carl Reidsema, and Aneesha Bakharia. 2018. Reciprocal Peer Recommendation for Learning Purposes. In *Proceedings of the 8th International Conference on Learning Analytics & Knowledge (LAK '18)*. 226–235. doi:10.1145/3170358.3170400

[28] Alvin E Roth. 1984. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of political Economy* 92, 6 (1984), 991–1016.

[29] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*. PMLR, 1670–1679.

[30] Tom Sühr, Asia J. Biega, Meike Zehlike, Krishna P. Gummadi, and Abhijnan Chakraborty. 2019. Two-Sided Fairness for Repeated Matchings in Two-Sided Markets: A Case Study of a Ride-Hailing Platform. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. Association for Computing Machinery, New York, NY, USA, 3082–3092. doi:10.1145/3292500.3330793

[31] Adith Swaminathan and Thorsten Joachims. 2015. Counterfactual Risk Minimization: Learning from Logged Bandit Feedback. In *Proceedings of the 32nd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 37)*, Francis Bach and David Blei (Eds.). PMLR, 814–823. https://proceedings.mlr.press/v37/swaminathan15.html

[32] Yoji Tomita and Tomohiko Yokoyama. 2024. Fair Reciprocal Recommendation in Matching Markets. In *Proceedings of the 18th ACM Conference on Recommender Systems (RecSys '24)*. doi:10.1145/3640457.3688130

[33] Ding Tong, Qifeng Qiao, Ting-Po Lee, James McInerney, and Justin Basilico. 2023. Navigating the Feedback Loop in Recommender Systems: Insights and Strategies from Industry Practice. In *Proceedings of the 17th ACM Conference on Recommender Systems* (Singapore, Singapore) *(RecSys '23)*. 1058–1061. doi:10.1145/3604915.3610246

[34] Peng Xia, Benyuan Liu, Yizhou Sun, and Cindy Chen. 2015. Reciprocal recommendation system for online dating. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. 234–241.

[35] Chen Yang, Sunhao Dai, Yupeng Hou, Wayne Xin Zhao, Jun Xu, Yang Song, and Hengshu Zhu. 2024. Revisiting Reciprocal Recommender Systems: Metrics, Formulation, and Method. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3714–3723.

[36] Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Yong Li, and Depeng Jin. 2021. Disentangling user interest and conformity for recommendation with causal embedding. In *Proceedings of the web conference 2021*. 2980–2991.

[37] Ziwei Zhu, Yun He, Yin Zhang, and James Caverlee. 2020. Unbiased Implicit Recommendation and Propensity Estimation via Combinational Joint Learning. In *Proceedings of the 14th ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. 551–556. doi:10.1145/3383313.3412210