

Fair Stable Matchings Under Correlated Preferences

Angelina Brilliantova,¹ Hadi Hosseini²

¹ Rochester Institute of Technology, Rochester, NY, USA

² Pennsylvania State University, University Park, PA, USA
lb9849@rit.edu,¹ hadi@psu.edu²

Abstract

Stable matching models are widely used in market design, school admission, and donor organ exchange. The classic Deferred Acceptance (DA) algorithm guarantees a stable matching that is optimal for one side (say men) and pessimal for the other (say women). A sex-equal stable matching aims at providing a fair solution to this problem. We demonstrate that under a class of correlated preferences, the DA algorithm either returns a sex-equal solution or has a very low sex-equality cost.

Introduction

Stable matching models are widely used in various domains including market design, donor organ exchange, and school admission. In its heart, the *stable matching problem* (SMP) looks for a pairwise matching between the agents of two disjoint sets (traditionally referred to as men and women) while taking into account their preferences (Gale and Shapley 1962). A key requirement is that the solutions must be *stable*, meaning that no pair of agents prefer each other to their assigned matches (Gale and Shapley 1962).

The most well-known approach for solving stable matching problems is the Deferred Acceptance algorithm (DA) (Gale and Shapley 1962). In this algorithm, the agents from one set (proposers) make proposals to the other set (receivers). When there are n agents on each side, DA is guaranteed to return a stable matching in $\mathcal{O}(n^2)$, which would be proposer-optimal and receiver-pessimal, meaning that each proposer gets matched to his most-preferred partner among all stable solutions, whereas each receiver gets her least-preferred stable alternative (Gale and Shapley 1962). One of the most prominent and well-studied fairness concepts is *sex-equality* that aims at finding a matching that equalizes the welfare of both sides. However, finding a sex-equal stable matching is a strongly NP-hard problem (Kato 1993).

For real-world problems, the set of stable matchings can be relatively small. In fact, the size of the solution space of SMP (a *stable set*) substantially depends on the correlation in the preferences of agents (Roth and Peranson 1999). We argue that for some restricted preference models, finding a sex-equal solution may be feasible. We conduct

an experimental analysis of the solution space under *correlated* (the Mallows model) and uncorrelated preferences (the Uniform model). We demonstrate that under the Mallows model, when the preferences of men and women have different noise parameters, the DA algorithm returns a sex-equal solution. Moreover, when noise parameters are equal, its outcome is substantially close to the sex-equal solution. The latter case is known to have a high asymptotic probability of exponentially-sized stable set (Levy 2017).

Preliminaries

In an instance of a stable matching problem, there is a set of men (M) and women (W), with $|M| = |W| = n$. Each agent x has a preference list $\succ_x$, which is a strict order over the other set. Agents' preferences are summarized in a *preference profile*, $\succ$. Let $r(w, m)$ be a position of woman w in $\succ_m$, $r(m, w)$ a position of man m in $\succ_w$. A *matching* between men and women is a mapping $\mu : M \cup W \rightarrow M \cup W$. If $\mu(m) = w$ then m is matched to w , which holds iff $\mu(w) = m$. Given a matching μ , a pair (m, w) is called a *blocking pair* if they prefer each other to their assigned partners in μ , i.e. $r(m, w) < r(\mu(w), w)$ and $r(w, m) < r(\mu(m), m)$. A matching is *stable* if it has no blocking pairs.

There may be several stable matchings for a given profile $\succ$. We let $S_\succ$ denote a *stable set*, consisting of all stable matchings with regard to $\succ$. A *man (woman)-optimal matching* is a stable matching μ in which each man $m \in M$ (woman $w \in W$) has a partner at least as good as in any other stable matching, i.e. $r(\mu(m), m) \leq r(\mu'(m), m)$ for $\forall \mu' \in S_\succ$.

The *sex-equality cost* measures the absolute difference between the sum of partners' ranks of men and women: $c(\mu) = |\sum_{m \in M} r(\mu(m), m) - \sum_{w \in W} r(\mu(w), w)|$. A *sex-equal matching* is the element of a stable set with the minimum sex-equality cost, i.e. $\operatorname{argmin}_{\mu \in S_\succ} (c(\mu))$.

Preference Models. Let π be a permutation of a preference list. In the *Uniform model*, the preferences of agents are not correlated: each π has $\frac{1}{n!}$ probability to be chosen. We use the well-studied *Mallows model* that provides a probabilistic model for permutations correlated with some common reference (Mallows 1957). In the Mallows model the preferences of agents are aggregated around a reference

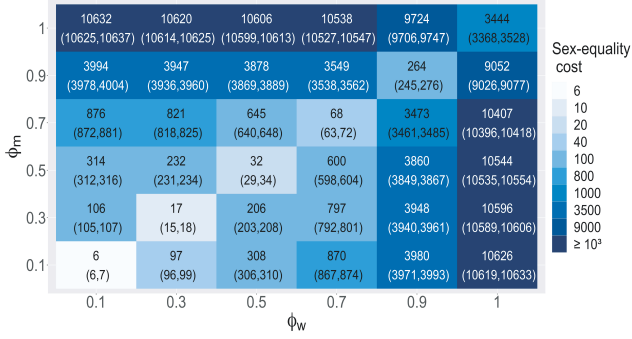


Figure 1: The median value of sex-equality costs of man-optimal matchings with the corresponding confidence intervals under the Mallows model with various ϕ_m, ϕ_w

ranking, $\hat{\pi}$. In this way, the preferences are globally correlated (as all preference lists correlate with $\hat{\pi}$). The probability for a permutation to be drawn increases as the distance to the reference ranking decreases: $p(\pi|\hat{\pi}, \phi) = \phi^{\tau(\pi, \hat{\pi})} / S_{n, \phi}$, where $\tau(\pi, \hat{\pi})$ is a Kendall-tau distance (the number of pairwise inversions) between π and $\hat{\pi}$, $S_{n, \phi}$ is a normalization constant, and $\phi \in [0, 1]$ is the *noise parameter* representing the uncertainty or variability in preferences. If $\phi = 1$ the variability is maximal, so the Mallows model is equivalent to the Uniform model; if $\phi = 0$ all samples are equal to $\hat{\pi}$.

Empirical Analysis

Data. We generate 1,000 instances of a stable matching problem for each $n \in [10, 150]$. When generating a preference profile, we draw a preference list for each agent *i.i.d* from the Mallows or Uniform distributions. In the Mallows model, men and women preferences are sampled from distinct distributions. In each instance, reference rankings are generated *i.i.d* uniformly at random and the Kendall-tau distance is measured between them. For each setting, we report median sex-equality costs and the mean probability that the DA algorithm produces a sex-equal stable matching over 1,000 instances along with the confidence intervals. Confidence intervals were obtained using a bootstrap sampling with replacement, sample size 1,000 with 100 repeats. We only report the results for $n = 150$ since for the smaller n the results are qualitatively the same.

Equitability of optimal stable matchings in the Mallows model. When the preferences of men and women have equal noise parameters, the size of a stable set is maximal and optimal matchings rarely coincide with the sex-equal solution (Figure 2). However, their sex-equality costs are considerably low and close to 0 (Figure 1). When preferences of agents from one side are correlated with a global reference, they compete for the same partners, resulting in a high sum of partners' rank. When the sides have preferences that are equally variable, their sums of partners' ranks are similar and they are equally unsatisfied with the matching.

The sex-equality cost increases with an increase in $\phi_m = \phi_w$, implying that more variable preferences of agents re-

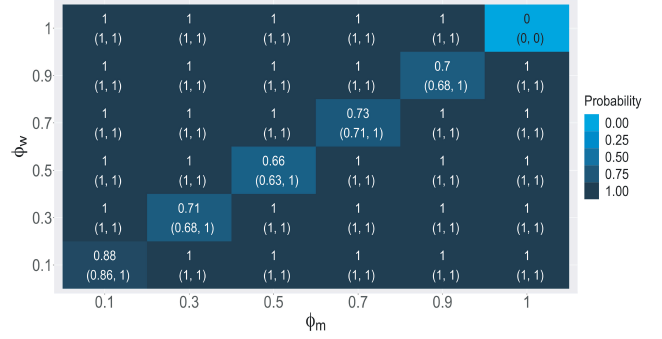


Figure 2: The mean probability that sex-equal solutions are the outcome of DA (man or woman-proposing) under the Mallows model with various ϕ_m, ϕ_w . Confidence intervals are given in parentheses

sult in less equitable solutions. Sex-equality costs of optimal matchings are the highest for uncorrelated preferences of one or both sides (with noise parameter equals 1).

When one side has more correlated preferences than the other (i.e. unequal noise parameters), the sex-equal matching coincides with one of the optimal matchings (either woman or man-optimal) with high probability (Figure 2). This observation holds for *all* generated instances with unequal ϕ , including those with large stable sets. When $\phi_w < \phi_m$, in any stable matching the sum of ranks for women partners is greater than for men. That is, women receive less preferred (worse ranked) partners than men in any stable matching, including the woman-optimal. Thus, a woman-optimal matching coincides with a sex-equal matching. Similarly, when $\phi_w > \phi_m$, a man-optimal matching is always sex-equal. All findings are independent of the Kendall-tau distance between the reference rankings for men and women.

Conclusions

In many real-world applications, agents tend to share a common but noisy reference in preferences. For such markets, the Deferred Acceptance algorithm may be considered to guarantee a stable solution in polynomial time without sacrificing fairness.

References

- Gale, D.; and Shapley, L. S. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69(1): 9–15.
- Kato, A. 1993. Complexity of the sex-equal stable marriage problem. *Japan Journal of Industrial and Applied Mathematics* 10(1): 1–19.
- Levy, A. W. 2017. *Novel uses of the Mallows model in coloring and matching*. Ph.D. thesis, Dept. of Mathematics, University of Washington, Seattle.
- Mallows, C. L. 1957. Non-null ranking models. I. *Biometrika* 44(1/2): 114–130.
- Roth, A. E.; and Peranson, E. 1999. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American economic review* 89(4): 748–780.