Fair Stable Matchings Under Correlated Preferences

Angelina Brilliantova,[1] **Hadi Hosseini**[2]
[1] Rochester Institute of Technology, Rochester, NY, USA
[2] Pennsylvania State University, University Park, PA, USA
lb9849@rit.edu,[1] hadi@psu.edu[2]

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 $\mu : M \cup W \to M \cup W$. 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. $\underset{\mu \in S_\succ}{\operatorname{argmin}}(c(\mu))$.

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