THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : MSTII - Mathématiques, Sciences et technologies de l'information, Informatique
Spécialité : Mathématiques Appliquées
Unité de recherche : Laboratoire d'Informatique de Grenoble

Efficacité et équité dans les problèmes d'appariement

Efficiency and Fairness in Matching Problems

Présentée par :

Rémi CASTERA

Direction de thèse :

Patrick LOISEAU — Directeur de thèse
DIRECTEUR DE RECHERCHE, CENTRE INRIA DE SACLAY

Bary PRADELSKI — Co-encadrant de thèse
Chargé de recherche, CNRS Délégation Alpes

Rapporteurs :

UTKU ÜNVER
PROFESSEUR, BOSTON COLLEGE

BRUNO ESCOFFIER
PROFESSEUR DES UNIVERSITES, SORBONNE UNIVERSITE

Thèse soutenue publiquement le **16 octobre 2024**, devant le jury composé de :

KIM THANG NGUYEN, — Président
PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES

PATRICK LOISEAU, — Directeur de thèse
DIRECTEUR DE RECHERCHE, CENTRE INRIA DE SACLAY

UTKU ÜNVER, — Rapporteur
PROFESSEUR, BOSTON COLLEGE

BRUNO ESCOFFIER, — Rapporteur
PROFESSEUR DES UNIVERSITES, SORBONNE UNIVERSITE

CLAIRE MATHIEU, — Examinatrice
DIRECTRICE DE RECHERCHE, CNRS ILE-DEFRANCE VILLEJUIF

BETTINA KLAUS, — Examinatrice
PROFESSEURE, UNIVERSITE DE LAUSANNE

RIDA LARAKI, — Examinateur
PROFESSEUR, UNIVERSITE MOHAMMED VI POLYTECHNIQUE

Invités :

BARY PRADELSKI
CHARGE DE RECHERCHE, CNRS/MAISON FRANÇAISE D'OXFORD

Abstract

Matching problems are ubiquitous, as they arise in school choice, college admission, job markets, or even refugee resettlement. The literature about matching historically focuses on efficiency and stability, but recently interest has been growing around fairness questions. This thesis aims to contribute to this emergent work. Assume that a population is divided into groups, representing different demographics, and the aim is to treat all groups fairly when choosing a matching. I consider various definitions of fairness towards those groups (a concept called group fairness), and which parameters of each studied model play a role in the existence or the optimality of a fair matching. In particular, I study the trade-off between fairness and efficiency, defined as the size of the matching when there are no preferences, or the satisfaction of agents when there are preferences. In the setting with no preferences, I propose an original geometric representation of the problem that allows me to give conditions for the existence of a matching that is maximal and fair, and when it does not exist, I provide tight bounds on the ratio between the size of the largest matching and the size of the largest fair matching (I call this ratio the Price of Fairness). In the setting with preferences, that I model as a college admission problem with students on one side and colleges on the other side, I study the role of the correlation between colleges' rankings of students. I show that correlation improves the efficiency of the stable matching. Moreover, when different groups have different levels of correlation in their rankings by the colleges, it creates disparities in each group's rate of students that remain unassigned, even when each college's individual ranking is completely fair towards each group. I also show that when colleges' rankings are fair, there is no trade-off between efficiency and fairness, in the sense that both can be achieved simultaneously.

Résumé

Les problèmes d'appariement sont omniprésents, dans les choix d'école, les admissions dans l'enseignement supérieur, le marché du travail, ou encore l'installation de réfugiés. La littérature, sur l'appariement s'est historiquement concentrée sur les questions de l'efficacité et de la stabilité, mais il y a depuis peu un intérêt grandissant pour les questions d'équité. Cette thèse a pour objectif de contribuer à ce nouveau champs de recherche. En supposant qu'une population est divisée en groupes, qui peuvent représenter différents groupes démographiques, le but est de les traiter équitablement au moment de choisir un appariement. J'étudie différentes notions d'équité entre les groupes, et quels sont les paramètres de chaque modèle étudié qui influencent l'existence ou l'optimalité d'un appariement équitable. En particulier, je m'intéresse au compromis entre équité et efficacité, définie comme la taille de l'appariement dans les problèmes sans préférences ou la satisfaction des agents dans les problèmes avec préférences. Dans le modèle sans préférences, je propose une représentation géométrique originale du problème qui permet de donner des conditions à l'existence d'un appariement qui soit à la fois de taille maximale et équitable, et quand ce n'est pas le cas je donne des bornes étroites sur le ratio entre la taille d'un appariement maximal et la taille du plus grand appariement équitable. Dans le cas avec préférences, que je modélise comme un problème d'admission à l'université où les agents d'un côté sont des étudiants et de l'autre des universités, j'étudie le rôle de la corrélation entre les classements que chaque université fait des étudiants. Je montre que la corrélation améliore l'efficacité de l'appariement stable. De plus, quand différents groupes ont des niveaux de corrélation différents dans leurs classements par les universités, cela crée des différences de taux d'admission entre ces groupes, même quand le classement de chaque université pris individuellement est parfaitement équitable. Je montre également que quand les classements de chaque université sont équitables, il n'y a pas de compromis entre efficacité et équité, dans le sens où ces deux objectifs peuvent être atteints simultanément.

Acknowledgement

First and foremost, I would like to thank my two advisors Patrick Loiseau and Bary Pradelski for those three and a half years during which they introduced me to the research world and and taught me so many things. Thanks to both of you for the time and energy you dedicated to me and the papers we wrote together, and for always encouraging me to travel and meet as many fellow researchers as possible. I really enjoyed working with you and the relation we have built, in spite of the distance that separated us for most of my PhD, but at least I know that leaving France will not prevent us from keeping in touch!

I would like to thank once again Utku Ünver and Bruno Escoffier for taking the time to review this thesis, and for their precious comments and suggestions. I am also very grateful to the other members of the jury Claire Mathieu, Kim Thang Nguyễn, Bettina Klaus and Rida Laraki for agreeing to assess my work during the defense.

In addition to my advisors, I had the chance to work and write papers with amazing and talented people, that I hope to be able to continue working with in the future. I would like to thank my co-authors Mathieu Molina, Felipe Garrido-Lucero, Simon Mauras and Vianney Perchet.

I am extremely grateful to all the people that helped me improve my papers and my research overall for their help, starting first and foremost with Julien Combe who agreed to be the "outside expert" for my PhD and who has as such followed my work all along and gave me precious feedback and advice. I will add all the people I met in various conferences and workshops that gave me great advice , to name a few Nick Arnosti, Jakob Weissteiner, Vitali Emelianov, Emil Temnyalov, Faidra Monachou, Marek Pycia, Itai Ashlagi, the participants and organizers of EC'22& 23, CED'22, Stony Brook'22, From Matching to Markets, 19th Matching in Practice, Matching Markets and Inequality, and GAIMSS.

The past three years would not have been so pleasant without my amazing colleagues Victor, Hugo, Julien, Pierre, Achille, Mathilde, Manon, Mathieu, Felipe, Simon M., my conference buddy Simon J., and many other members of my two teams POLARIS and Fairplay. I want to thank you all for the many (maybe too many) good times we spent at EVE, playing petanque or coinche, having a drink, going out to eat or more generally chatting and taking much deserved breaks. I have a special thought for our Diplomacy game that ending up lasting way too long and endangered more than one PhD...

Arrivant à la fin de mes études, j'aimerais remercier tous les professeurs qui m'ont donné le goût de la recherche, m'ont encouragé et m'ont aidé à arriver là où je suis. Je pense en particulier à Stéphane Attal, Jérôme Germoni, Claude Inserra, Pierre Lavaurs, Gaëlle Dejou, Stéphane Vignoli, Sébastien Gauthier, Bruno Ziliotto, Stéphane Gaubert et Quentin Mérigot.

En dehors du cadre académique, je tiens évidemment à remercier tous les amis, récents et anciens, qui m'ont accompagné tout au long de cette aventure (je m'excuse d'avance à tous ceux que je vais oublier de citer). Les grenoblois: Manon, Romain, Julien, Camille, Guillaume, Gaël, Bruno, Margot, Guillaume, Pierre, Maureen, Miguel, Nausicaa, Nathis, Adrien, Alice; les confinés: Louis, Dorine, Pierre, Eric, Emile, Laura, Jérémie, Dana, Alban, Audrey, Elie; les fanfarons: Nicolas, Nell, Louis, Tristan; et bien évidemment les Fripons[1] Achille, Flore, Bédi, Florent, Valentine, Florian, Théo, Valou, Alexandre et Alexandre, Maxime, Cricri, Laurie, et Martin, qui ont l'immense mérite d'encore me supporter après plus de vingt ans.

Le déterminisme social ne pourrait pas fonctionner aussi bien sans une famille excpetionnelle, je remercie donc évidemment mes parents qui m'ont toujours soutenu dans tous mes choix et projets, ma soeur Gwenaëlle qui a mis un high score au bac pas facile à aller chercher, et mon frère Camille qui a passé les vingt dernières années à me dire "là tu fais le malin mais tu verras l'année prochaine ça sera pas la même" (et il n'avait pas tort).

Je remercie notre chat Kiwi, qui a contribué à cette thèse en tapant sur le clavier probablement plus de caractères que moi, qui n'ont malheureusement pas trouvé leur place dans la version finale.

Enfin, je remercie Tiphaine d'être à mes côtés tous les jours, les bons comme les mauvais, et de rendre ce monde de fous un peu plus doux.

[1] nom original censuré pour cause de vulgarité

Bon, ça ne veut absolument rien dire, mais je trouve que c'est assez dans le ton.

François Rollin

Contents

1 Introduction: Scientific background and related literature

Contents

Decision-makers are often faced with the intricate task of optimizing the allocation of potentially scarce resources or opportunities among a population. This challenge is faced in real-world applications such as dispatching students to universities [GS13], selling goods in markets [CS98], granting loans to businesses [HRW00], or displaying online advertisement [Meh13]. Many of these allocation problems can be framed as matching problems, a theoretical model based on graph theory extensively studied in the economics [EIV23], operations research [Der88], and computer science literature [Nis+07]. The main goal is usually to efficiently find a "good" matching, where "good" is understood as maximizing some form of social welfare, for example matching the maximum number of individuals.

Outcome inequalities for different demographic or social groups are ubiquitous, for example, in college admission, job assignment, or investment allocation. [AKR22] find that Asian-American applicants have lower admission chances at Harvard than white applicants for a similar academic record, [NR19] find significantly lower inflows in female-managed mutual funds than in male-managed mutual funds, and [BM04] find race-based discrimination in callback decisions by job advertisers. Consequently, the sources of observed outcome inequalities remain the subject of frequent and continued controversy and political debate.

Matching problems, however, can also be highly sensitive. For instance, the European Union has recently proposed the creation of a job market matching platform between employers and migrants (European Commission press brief [Com23]) to address labor shortage. Migrants can belong to different demographic groups defined by sensitive or protected attributes such as age, race, gender, or wealth; and it is essential that matching decisions preclude discrimination across such groups. Hence, for such complex decision-making problems the definition of "good" matching cannot purely be based on utility requirements: ethical, political and legal considerations must be taken into account, and careful policies and allocations have to be designed to avoid harmful impact due to various forms of discrimination. In a similar vein, the global refugees resettlement crisis has reached new heights with more than 2 millions projected resettlement needs in 2023, a 35% increase from 2021 and 2022 according to the UNHCR [Ref23]. As this is a pressing issue that affects increasingly larger populations, the way refugees are assigned to resettlement sites needs to be addressed carefully. This problem has recently received attention from the matching community (see [DKT23a] and subsequent related works [Fre+23; Aha+21]). Again here, issues of discrimination across different demographic groups of refugees need to be addressed, be it for legal, ethical, or political reasons.

In this thesis, I[1] look into the fairness aspects of matching problems, and the potential trade-off between fairness and efficiency. In this first chapter, I introduce the scientific context of the thesis, as well as the fundamental technical elements. I first recall fundamentals of graph and matching theory, then explore the different meanings of the word "fairness" in mathematics, computer science and economics. Finally, I discuss the existing related work on group fairness in matching, give an outline of the content of this thesis, and explain how it fits in this literature.

1.1 Background: Matching

A **matching** problem is a situation where a central authority or clearinghouse has to make pairs from a pool of objects or agents. Matching theory has emerged as a part of graph theory in the beginning of the 20th century, with seminal results from Hall [Hal35] and Kőnig [Kőn31]. Matching became an important subject in combinatorial optimization due to its link with fundamental results of optimization theory (notably, Hall and Kőnig theorems are precursors of the max-flow/min-cut theorem, which is itself a special case of the duality theorem for linear programming). The main purpose of the classical results in matching theory is to find matchings of maximal size [Wes18; Blu90; MV80; GT91].

As described by Roth and Sotomayor [RS92], the history of **two sided matching** starts during the 1940's, when the Association of American Medical Colleges (AAMC) tried to find a way to control the competition among hospitals for medical interns. Since there were more open

[1] The introduction and conclusion of this thesis are written with the singular pronoun "I" as they are my personal work. The chapters, being based on articles written with co-authors, are written with the plural pronoun "we".

positions than candidates, hospitals were fighting to enroll the best students, and they tried to do so by starting their admission process before the other hospitals. After a few years, hospitals had ended up recruiting interns almost two years before the beginning of the internship. The problem was fixed by keeping students' academic records confidential until a few months before the internship. However, the new problem was that when a student knew they were admitted to an hospital, but also on waiting list at another one they preferred, they were inclined to wait until the last moment before accepting their admission in case they would receive an offer from the hospital they preferred. Following this, students were given ten days to answer an offer, then eight, up to the point where students had only twelve hours to answer, and were forbidden to have a phone call during these twelve hours, and hospitals as well. This example illustrates the problems that arise when an admission process involving several competing hospitals/colleges/companies is not performed in a centralized way. The first algorithm addressing this problem since known as the **college admission problem** was proposed in 1962 by Gale and Shapley [GS62]. Their seminal work opened a new branch of matching theory, this time linked to game theory rather than combinatorial optimization. The college admission problem has been extensively studied [Rot86; RS92; APR09], as well as many variants adapted to similar problems. Indeed, situations where agents from two sides need to be matched according to their preferences are ubiquitous: students and universities [GS62], doctors and hospitals [NRM], internet users and servers [MS15], or even rabbis and congregations [BP03].

1.1.1 Definition and properties

Definition 1.1 (Graph)**.** A **graph** is a couple $\mathcal{G} = (V, E)$, where V is a finite set, the elements of which are called **vertices**, and $E \subseteq V \times V$, the elements of which are called **edges**.

An example of a graph is given in Figure 1.1. If every vertex in the graph is connected to all other vertices, I say that the graph is **complete**.

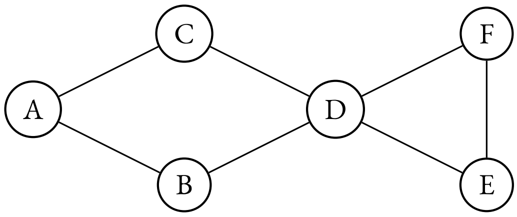


Fig. 1.1: Example of a graph with six vertices and seven edges.

Definition 1.2 (Matching)**.** We say that $\mu \subseteq E$ is a **matching** if each vertex in V belongs to at most one edge in μ.

A matching is simply a set of edges that do not share any endpoints. If a vertex v belongs to an edge in a matching μ, I say that v is **matched** in μ. Figure 1.2 shows two different sets of edges

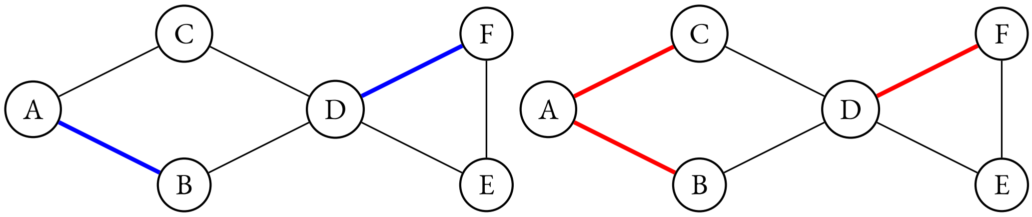


Fig. 1.2: Left: the two thick blue edges form a matching because no vertex is used more than once. Right: the three thick red edges do not form a matching because vertex A is used twice.

on the same graph, one that verifies the definition of a matching and one that does not. We call $\mathcal{M}(\mathcal{G})$ the set of all possible matchings on $\mathcal{G}$, and simply write $\mathcal{M}$, omitting the dependence on $\mathcal{G}$, when there is no ambiguity.

Proposition 1.1. *$\mathcal{M}$ is downward closed, i.e., if $\mu \in \mathcal{M}$ and $\mu' \subseteq \mu$, then $\mu' \in \mathcal{M}$.*

In other words, if I remove edges from a matching it remains a matching. However, adding edges may violate the definition of a matching, so finding large matchings is a non-trivial problem.

Definition 1.3 (Maximal and maximum matchings)**.** A matching μ is:

- **maximal** if there is no matching that contains μ other than itself.
- **maximum** if no matching has a strictly larger cardinality.

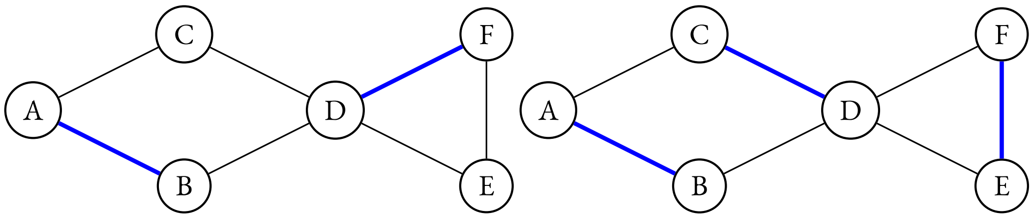


Fig. 1.3: Left: the highlighted matching is maximal; no edge can be added without using some vertex twice. Right: the highlighted matching is maximum; there is no larger matching on this graph.

By definition, there always exist at least one maximal matching (that might be the empty set). A graph might have several maximal matchings, and maximal matchings might even be of different size. On the other hand, there always exists a maximum matching, and there might be more than one, but contrarily to maximal matchings they all have the same size. Figure 1.3 gives examples of maximal and maximum matchings on the same example graph as I used before.

Proposition 1.2. *If μ is a maximal matching, then every edge in E has an endpoint in common with an edge in μ.*

Proof. Suppose that some edge e has no endpoint in common with any edge in μ, then $\mu \cup \{e\} \in \mathcal{M}$, which contradicts μ's maximality. ■

Definition 1.4 (Perfect matchings)**.** A matching μ is:

- **perfect** if all vertices are matched.
- **near-prefect** if only one vertex is not matched.

Clearly, a perfect matching can only exist if V has an even cardinality and a near-perfect matching can only exist if V has an odd cardinality. In either case, there might exist no perfect (or near-perfect) matching. When they exist, perfect and near-perfect matchings are maximum and thus maximal.

1.1.2 Finding maximum matchings

The simplest algorithm to find a maximum matching in any graph is called the Blossom algorithm and was proposed by Edmonds [Edm65]. It relies on the concept of augmenting paths.

Definition 1.5 (Paths)**.** A **path** is a sequence of edges $(e_1 = (u_1, v_1), ..., e_k = (u_k, v_k))$, for any $k \in \mathbb{N}^*$, such that for all $i \in [k-1] := \{1, \ldots, k-1\}$, $v_i = u_{i+1}$, and for all $i, j \in [k]$, $u_i \neq u_j$.

- A path that ends on the vertex it started on ($v_k = u_1$) is called a **cycle**.
- Given a matching μ, a path is said to be **alternating** if at least one of its extremities is unmatched and its edges are alternatively in μ and not in μ.
- An alternating path such that the first and last edges of the path are in $E \setminus \mu$ is called an **augmenting path**.

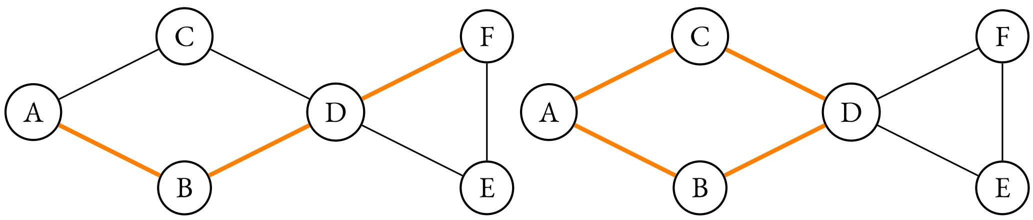


Fig. 1.4: Left: the highlighted edges form a path. Right: the highlighted edges form a cycle.

A path and a cycle are illustrated on Figure 1.4. Given a matching μ and an augmenting path P, I can find a larger matching simply by removing from μ the edges that are in P and adding all edges in P that were not in μ. Formally, $\mu' = (\mu \setminus P) \cup (P \setminus \mu)$ is a matching that has exactly one more element than μ. Figure 1.5 illustrates this: starting from the blue matching μ of size two, I can find an augmenting path P as shown on the right, and removing the edges that are in $P \cap \mu$ and replacing them by those that are in $P \setminus \mu$ gives a matching larger than μ by one edge.

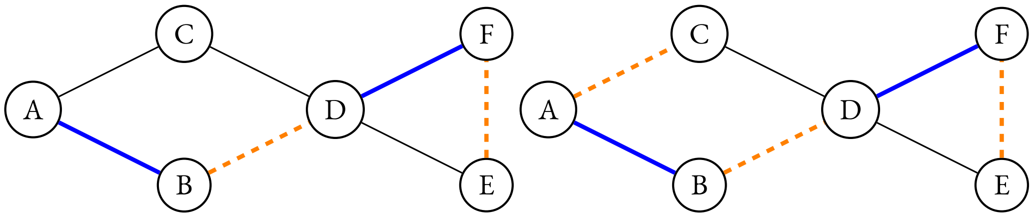


Fig. 1.5: The highlighted blue edges are a matching μ. Left: the path formed by μ and the dashed orange edges is an alternating path for μ. Right: the path formed by μ and the dashed orange edges is an augmenting path for μ, i.e., an alternating path that starts and ends with edges not in μ.

Starting from some matching, I can then greedily look for augmenting paths to increase its size. This leads to the Blossom algorithm (see Algorithm 1).

Algorithm 1: Blossom algorithm

Input:
Graph $\mathcal{G}$ and matching μ
Output:
A maximal matching

while *There exists an augmenting path* **do**
    Augment μ along any augmenting path
end
Return μ.

To prove that this algorithm always output a maximum matching, I only need Berge's theorem [Ber57]:

Theorem 1.3 (Berge, 1957). *A matching is maximum if and only if it has no augmenting path.*

Proof. Let μ be a matching. Suppose there exists an augmenting path, I already showed that it implies the existence of a larger matching. Conversely, suppose there exists a larger matching μ'. Consider the graph $\mathcal{G}' = (V, E')$ with $E' = (\mu \setminus \mu') \cup (\mu' \setminus \mu)$, called the symmetric difference between μ and μ'. This graph consists of isolated vertices and alternating paths. If some of those paths are cycles, then they are of even length since they are alternating. Since μ' is larger than μ there must exist a path of odd length, that starts and ends with edges from μ', which is therefore an augmenting path, concluding the proof. ■

This result proves that the Blossom algorithm always outputs a maximum matching, because it stops when it finds no augmenting path, meaning that its current matching is maximum. The operation of finding an augmenting path is not trivial but efficient methods exist, which will not be detailed here. The overall complexity of this algorithm is $\mathcal{O}(|E||V|^2)$.

In this thesis, I will focus on a specific class of graphs, bipartite graphs, for which finding maximum matchings happen to be easier.

Definition 1.6 (Bipartite graph)**.** We say that a graph $\mathcal{G}$ is **bipartite** if the set of vertices can be partitioned into two parts U and V such that all edges consist of one element of U and one element of V.

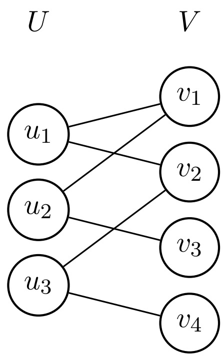


Fig. 1.6: Example of a bipartite graph.

We write bipartite graphs as $\mathcal{G} = (U, V, E)$. An example of a bipartite graph is shown on Figure 1.6. To find a maximum matching in a bipartite graph I can use the Ford-Fulkerson algorithm [FF56] (see Algorithm 2). This algorithm is originally used to find the maximal flow between two vertices labeled as **source** and **sink**, but it can be used to find maximum matchings by adding a source connected to all edges in U and a sink connected to all edges in V (U and V play interchangeable roles here).

Algorithm 2: Ford-Fulkerson algorithm for bipartite graphs

Input:
Bipartite graph $\mathcal{G}$
Output:
A maximum matching

Initialization:
Add a vertex s and connect it to all vertices in U, and a vertex t connected to all vertices in V. Change all edges to be arrows, pointing from s to U, from U to V and from V to t.
while *There exists a path from s to t following arrows* **do**
  Find a path P from s to t following arrows
  For every arrow in P, invert the direction of the arrow
end
Return the arrows going from V to U.

At the beginning of the algorithm, all arrows go from U to V; the maximum matching is given by the arrows that are in the opposite direction when the algorithm stops, as illustrated in Figure

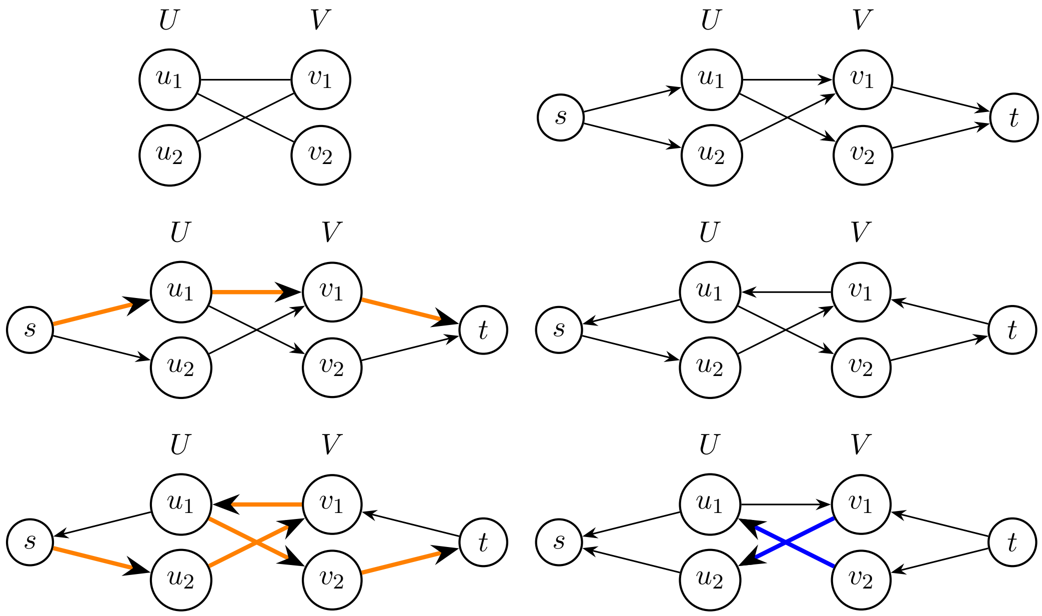


Fig. 1.7: Illustration of the Ford-Fulkerson algorithm. Top row: initialization, the initial bipartite graph is on the left, on the right I add the source s and the sink t, and turn edges into arrows from left to right. Middle row: first iteration, a path from s to t is found through the edge (u_1, v_1) (left), I invert the arrows along it (right). Bottom row: second iteration, a new path is found (left), once again I invert the arrows(right). There is no more paths from s to t, the algorithm stops and the maximum matching found is $((u_1, v_2), (u_2, v_1))$ as given by the arrows going from V to U.

1.7. This algorithm converges in $\mathcal{O}(|E|(|U| + |V|))$, which seems faster than the Blossom algorithm. However, the blossom algorithm when used on a bipartite graph also converges in $\mathcal{O}(|E|(|U| + |V|))$ (instead of $\mathcal{O}(|E|(|U| + |V|)^2)$ in the general case) because augmenting paths are easier to find in bipartite graphs. In practice, even though the convergence time is the same, Ford-Fulkerson is often preferred for bipartite graphs because it is much easier to implement (finding a path from s to t following arrows is more straightforward than finding an augmenting path). More complex algorithms have been developed since with better complexity, notably the Hopcroft–Karp–Karzanov [HK73; Kar73] algorithm that achieves $\mathcal{O}(|E|\sqrt{|U| + |V|})$, or the Chandran-Hochbaum algorithm [CH11] that achieves $\mathcal{O}\left(\min\{|U|k, E\} + \sqrt{k}\min\{k^2, E\}\right)$, where k is the size of the maximum matching.

1.1.3 Two-sided matching

The matching model I introduced so far is quite generic and can be used to model many situations, but in many problems from computer science and economics there is a feature that I did not yet take into account: preferences. So far I looked for matchings as large as possible, assuming that all pairs edges of the graph were equally good to the matching, but that might not always be the case, especially when vertices represent people or institutions. In this section, I introduce the classical model of bipartite matching with preferences. We consider that vertices on each side have preferences over the vertices of the other side, that is why I call this **two-sided** matching, as opposed to a model where only one side has preferences and the other is indifferent (which is called one-sided matching and will not be considered in this thesis).

Consider a mixed doubles tennis tournament. There are n men and n women, and one needs to make teams taking into account that each player p has preferences $\succ_p$ over the players of the other gender they want to be teamed-up with, which is a total order with no ties. There is an edge between two players of opposite genders if and only if they both prefer being teamed-up rather than remaining alone. Each player can only be matched to at most one player (I call this model **one-to-one** matching). It is obvious that it is often impossible to assign to each player their favorite partner if the preferences do not perfectly line up. Then, which matchings respect players preferences and which do not? An answer was proposed by [GS62] with the notion of stability.

Definition 1.7 (Stable matching)**.** Let μ be a matching.

1. If there exists two players p, p' of opposite genders who are connected by an edge and who are both not matched, the matching is said **wasteful**.

2. If players p and p' are connected and both prefer each other rather than their respective matches (or if them is unmatched), they experience **justified envy**.

A pair of players in situation 1. or 2. is called a **blocking pair**. If μ contains no blocking pair, it is called **stable**.

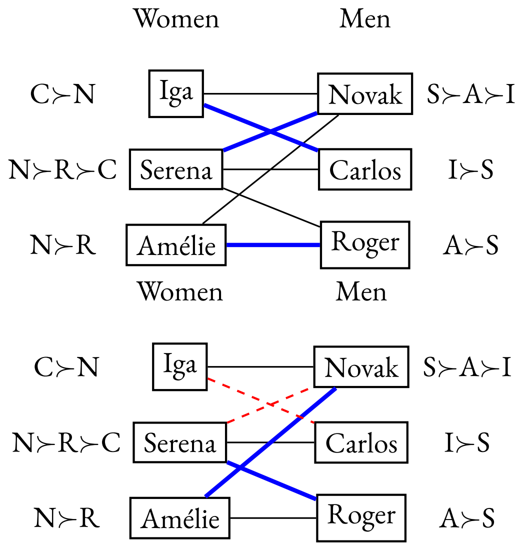


Fig. 1.8: Example of a two-sided, one-to-one matching problem. Left: the matching in blue (IC, SN, AR) is stable. Right: the matching in blue (SR, AN) is unstable because of two blocking pairs, represented by the dashed red lines. One is between Serena and Novak who would prefer being matched together (justified envy), the other one is between Iga and Carlos who could be matched together but are both unassigned (waste).

An illustration of a two-sided matching problem is shown in Figure 1.8, with an example of a stable matching on the left and an unstable matching on the right. Notice that a wasteful matching is equivalent to a matching that is not maximal, since it is possible to add an edge. The notion of stability is named this way because if two players form a blocking pair, then they will want to team up and leave their current teammate.

The question that naturally arises after defining such a property is whether there always exists a stable matching. The answer was also given in [GS62]

Theorem 1.4 ([GS62])**.** *There always exists a stable matching.*

To do so, they propose the **Deferred Acceptance** algorithm (shortened DA) that outputs a stable matching and prove that the algorithm always terminate (see Algorithm 3).

Algorithm 3: Deferred Acceptance algorithm (one-to-one, women-proposing)

Input:
Bipartite graph $\mathcal{G}$ with preferences
Output:
A stable matching

while *Some woman is not matched and has not been rejected by all men connected to her* **do**
    Every woman tries to match with the man she prefers among those that are connected to her and have not rejected her yet. Every man that receives proposals temporarily keeps the partner he prefers among those who are proposing and his current one if he already had one.
end
Return the obtained matching.

This version of the algorithm is called "women-proposing, men, disposing" or shorter "women-proposing". Obviously the roles of men and women can be reversed to obtain the "men-proposing" version of the algorithm. An illustration of the steps of the algorithm is shown in Figure 1.9.

To show that Algorithm 3 produces a stable matching, I only need the following observation: suppose there exists a blocking pair (p, p') where p is the woman. Since p prefers p' to her match, it means that she proposed to him at some point and he declined. If he declined, it was necessarily because he preferred his partner at the time to p. Moreover, throughout the algorithm, men only switch partners for ones they prefer, so he necessarily prefers his final match to p, so they cannot form a blocking pair.

To prove that there always exists a stable matching, I now only need to prove that the algorithm always terminate.

Proof. The algorithm terminates when every woman either has a partner or has been rejected by every potential partner. This implies that at every step, if it has not terminated, some woman is

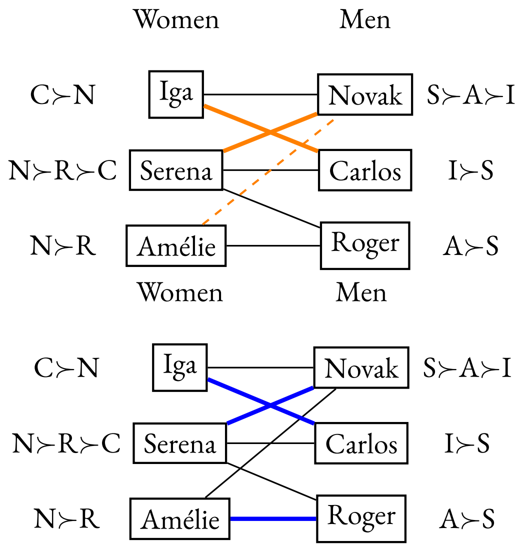


Fig. 1.9: Illustration of the execution of the Deferred Acceptance algorithm. Left: first iteration, every woman proposes to her preferred man. Serena and Amélie both propose to Novak, who prefers Serena so he keeps her. Right: second iteration, Iga and Serena do not propose since they are matched for now, Amélie proposes to her second choice Roger, who accepts since he was unassigned. Every woman is matched so the algorithm stops and the stable matching (in blue) is (IC, SN, AR).

not matched and has a man to propose to, so there is at least one proposal sent at each step. The number of proposals is bounded by n^2 since each woman sends at most n proposals. Since the number of proposals sent is an integer, strictly increasing, and bounded, the algorithm necessarily terminates. ∎

Now that we understand stability and how to find stable matchings, I can extend the model to a more general situation. Suppose that instead of tennis players, I need to match a set of students S to colleges. Similarly, all students have preferences over the colleges they find acceptable, and colleges have priorities over students (I use two different words - preferences for students and priorities for colleges - to make future discussions clearer and not too heavy). Each student can be matched to at most one college, however each college C can admit several students, at most a number q_C. This kind of problem is called **many-to-one** matching, as opposed to the tennis tournament which was one-to-one matching. Stability is defined in the exact same way as before: there is waste if a college has an empty seat and some student would prefer this college to their current match (including if they are unmatched), and there is justified envy if some student would prefer a college C to their current one and has a better rank at C than a student currently attending.

The Deferred Acceptance algorithm needs only to be slightly modified to accommodate for this change (see Algorithm 4). Since the proofs of correctness and termination of this version of the algorithm are almost identical to the previous one they are omitted. Figure 1.10 illustrates the steps of the algorithm.

Algorithm 4: Deferred Acceptance algorithm (many-to-one, students-proposing)

Input:
Bipartite graph $\mathcal{G}$ with preferences, and capacities on one side
Output:
A stable matching

while *Some student is not matched and has not been rejected by all colleges connected to them* **do**
    Every student applies to the college they prefer among those that are connected to them and have not rejected them yet. Every college that receives applications temporarily enrolls, in the limit of its capacity, the students it prefers among those it had previously enrolled and those that are now applying.
end
Return the obtained matching.

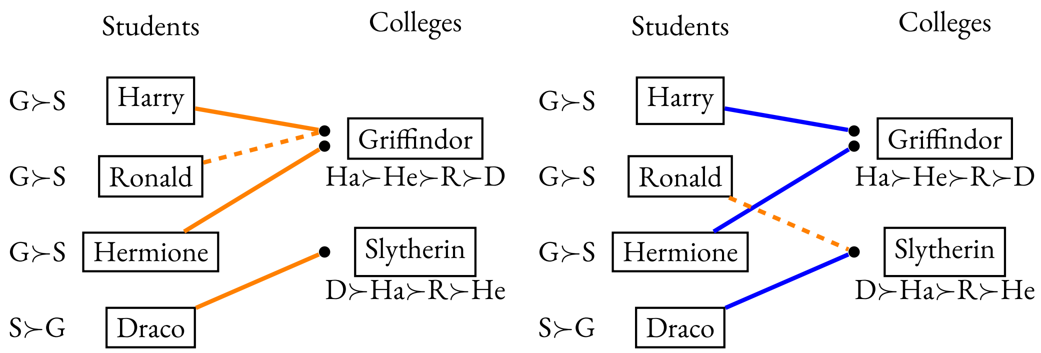


Fig. 1.10: Illustration of the execution of the Deferred Acceptance algorithm. We assume that the graph is complete and therefore do not represent all edges for clarity, instead at each step I only draw the current matching and applications. Left: first iteration, all students apply to their first choice. Harry, Ronald and Hermione all apply to Griffindor but there are only two seats, so Ronald is rejected. Right: second iteration, Ronald applies to Slytherin, but Draco is already there and Slytherin prefers Draco to Ronald, so he is rejected. Every students is either matched or has been rejected from all colleges, so the algorithm stops and gives the stable matching (in blue)((G, (Ha, He)), (S, (D))).

Remark 1.1. One may wonder whether agents could get a better outcome from the algorithm by misreporting their preferences. This question was answered by Dubins and Freedman [DF81] and Roth [Rot82]: with students-proposing DA, no student can improve their outcome by misreporting their preferences (and conversely for colleges with colleges-proposing DA). We say that the mechanism is **strategy-proof** for students (respectively for colleges). However, the other side can always manipulate. Roth further proved in [Rot82] that no mechanism that outputs a stable matching can be strategy-proof for both sides.

We now state some fundamental results regarding the structure of the set of stable matchings. We state them with the students and colleges terminology but they also apply to one-to-one matching problems.

Theorem 1.5.

- *[GS62]: The matching given by students-proposing DA is optimal for students, in the sense that every student weakly prefers[2] this matching to any other stable matching. The converse is true for colleges-proposing DA.*

- *[Knu76]: The students-optimal stable matching is pessimal for colleges (every college weakly prefers[3] any stable matching to this one). The converse is true: the colleges-optimal stable matching is pessimal for students.*

Definition 1.8 (Join and meet)**.** Let μ and μ' be two stable matchings. We define the join operator $\vee$ that gives an allocation where each student gets their favorite partner between those they get in μ and μ'. Similarly, the meet operator $\wedge$ gives every student their least favorite partner between the two matchings.

Theorem 1.6 (Conway & Knuth, [Knu76])**.**

- *The join and meet of two stable matchings are still matchings and still stable.*

- *The set of stable matchings endowed with those two operators forms a distributive lattice.*

A distributive lattice is an algebraic structure that has a partial order given by $\mu \leq \mu' \Leftrightarrow (\mu \wedge \mu' = \mu) \Leftrightarrow (\mu \vee \mu' = \mu')$. With this partial order, the student-optimal stable matching is the maximal element of the set of stable matchings and the college-optimal stable matching is the minimal element. For a better understanding of the structure of a lattice an example is provided in Figure 1.11

Remark 1.2. The set of stable matchings does not have a more particular structure than that of a distributive lattice: Blair [Bla84] showed that any finite lattice is isomorphic to the set of stable matchings of some instance of the problem.

We finally state a result describing the sets of matched and unmatched agents when the numbers of students and seats are different.

Theorem 1.7 (McVittie & Wilson [MW71], Roth [Rot84; Rot86])**.** *For a given instance of the college admission problem,*

1. *the sets of matched and unmatched students are the same in every stable matching,*

[2] A matching is weakly preferred to another by a student if they get either the same or a better outcome.

[3] A matching is weakly preferred to another by a college if it gets the same set of students, or a set that Pareto-dominates the other w.r.t. the ranking of the college.

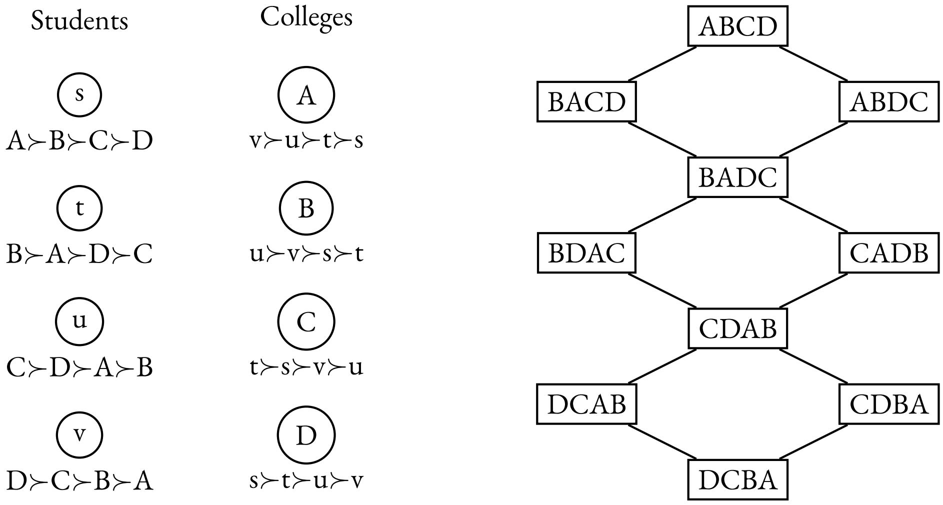


Fig. 1.11: Example of a college admission problem and its set of stable matchings.Left: 4 students and 4 colleges with capacity 1, each with their preferences written below them (I assume the graph is complete and do not draw the edges for clarity). Right: the corresponding lattice of stable matchings. Matchings are represented by which college is matched to s, t, u and v in that order, "BDAC" for instance represents the matching ((s,B),(t,D),(u,A),(v,C)). If a matching is higher than another, it means that the top one is preferred by all student to the bottom one. When two matchings are on the same level, the one on top of them is their "join" and the one below is their "meet", those relations are represented by the edges.

2. *the number of unfilled seats in each college is the same in every stable matching,*

3. *the colleges that have empty seats get the same set of students in every stable matching.*

This result shows that the only students that get different outcomes across stable matchings are those matched to colleges that fill all their seats. Unassigned students and those matched to colleges with remaining seats will get the same outcome in every stable matching.

1.2 Background: Fairness

Fairness is defined as "the quality of treating people equally or in a way that is reasonable"[4]. This definition is quite vague, and has lead to very different interpretations in many areas of computer science and economics every time researchers tried to study the fairness of a situation or to impose fairness constraints to a mechanism. In this section, I give an overview of the most common fairness notions used in the scientific literature and classify them in two main categories, **individual fairness** and **group fairness**. We finally introduce in more detail a particular source of unfairness (related to group fairness), called **statistical discrimination**.

[4] Oxford advanced learner's dictionary

1.2.1 Individual fairness

Individual fairness names a class of fairness notions related to merit. Rooted in philosophy, it can be traced back to Aristotle, "equals should be treated equally and unequals unequally". The main idea is that similar agents should get similar outcomes, and was more recently imported in economics theory by Young [You94] and Roemer [Roe98]. Since then, it has been considered by many authors in the economics and computer science literatures.

In the past decade, stemming from the AI literature, the most used framework has been the one proposed by Dwork et al. [Dwo+12]. Let S be a set of individuals, and A the set of possible outcomes. Define a distance between individuals $d : V \times V \to \mathbb{R}$. Agents are mapped to distributions of outcomes by $\mu : V \to \Delta(A)$, and $D : \Delta(A) \times \Delta(A) \to \mathbb{R}$ is a distance between distributions of outcomes. They then propose the following definition.

Definition 1.9 (Individual fairness)**.** Given two distances d and D between individuals and distributions of outcomes respectively, a mapping $M : V \to \Delta(A)$ is said individually fair if it is 1-Lipschitz, i.e., if $\forall x, y \in V, D(M(x), M(y)) \leq d(x, y)$.

Indeed, this definition is a natural formalization of the idea that similar agents have to get similar outcomes. However, this framework leaves much freedom in the choice of the distances d and D. In order to obtain a relevant notion of fairness, those functions must be wisely chosen. Notice that A could represent absolute outcomes, or individual-dependent utilities. For instance, in a matching problem, two students getting the same outcome could either mean getting the same college, or both getting their k-th choice in their respective preference lists for some k (as is done in [Dev+23]), depending on how A is defined.

While many authors who studied individual fairness claim that it "captures the intuitive notion of fairness" [Dwo+12] and prevents the disparate treatment of individuals from different demographic groups (cf. next subsection "Group fairness"), it appears to be only partially true. As noted by Fleisher [Fle21], there are several limitations to individual fairness. He first notices that individual fairness does not encompass efficiency, in the sense that not allocating any resource to any one is individually fair, but very wasteful. Second, the similarity metric used to compare individuals is designed by humans, and can therefore encompass bias towards some groups of individuals. In the same vein, designing such a metric requires to choose which features are relevant to compare individuals and which are not, which requires a moral judgment over what constitutes fairness, making the definition circular.

Those observations indicate that individual fairness alone is not sufficient to encompass the common idea of fairness. Fleisher's first observation tells us that we need to pay attention to the efficiency of mechanisms since individual fairness offers no guaranty on this side. Even more importantly, we must consider fairness notions that ensure that different demographic groups are treated fairly independently of human bias.

1.2.2 Group fairness

To quantify discrimination towards a group of individuals by a mechanism, it seems quite natural to look at how they are treated as a group by this mechanism compared to other groups or the global population. The definition of treatment is dependent on the context and of course subject to interpretation, but should relate to some measure of welfare or satisfaction, or in the most simplistic case a binary positive/negative outcome.

The machine learning community has taken an interest in the past few years in this question, especially in the context of classification, i.e., the binary setting I just mentioned. Several metrics have been introduced in this context to measure unequal treatment between groups, almost all of them based on the error rates of a classifier. Let me briefly describe this model.

Consider a population of individuals S, partitioned in groups $G_1, \dots, G_K$. Every individual s has a **true label** $y_s \in \{-1, 1\}$. A mechanism has to classify the individuals without knowing their true labels, for each one it outputs a **predicted label** $\hat{y}_s$. The mechanism, called a **classifier**, can be compared to the ground truth by some metrics.

Definition 1.10. Given a set S of individuals with true labels y_s and predicted labels $\hat{y}_s$,

- the **True Positives rate** is $\mathbb{P}(\hat{y}_s = 1 | y_s = 1)$, the **True Negatives rate** is $\mathbb{P}(\hat{y}_s = -1 | y_s = -1)$,
- the **False Positives rate** is $\mathbb{P}(\hat{y}_s = 1 | y_s = -1)$, the **False Negatives rate** is $\mathbb{P}(\hat{y}_s = -1 | y_s = 1)$,
- the **Predicted Positives rate** is $\mathbb{P}(\hat{y}_s = 1)$, the **Predicted Negatives rate** is $\mathbb{P}(\hat{y}_s = -1)$.

For a group $G \subset S$, I define the same measures conditioned on belonging to G, e.g., G's True Positives rate is $\mathbb{P}(\hat{y}_s = 1 | y_s = 1, s \in G)$.

Those metrics are illustrated in Figure 1.12. By comparing those metrics between groups, I can quantify the differential treatment. A growing literature is dedicated to the design of machine learning-based classifiers that have, as a constraint, that one or several of those metrics must be equal across groups. Each combination of metrics leading to a different fairness constraint, I here list the most commons:

		Predicted labels		
		+1	-1	
True label	+1	True Positive	False Negative	Actual Positive
	-1	False Positive	True Negative	Actual Negative
		Predicted Positive	Predicted Negative	

Fig. 1.12: The different types of classification outcomes

- **Demographic Parity**: equalize Predicted Positive rates [BS20; Eme+22].
- **Equal Opportunity**: equalize True Positive Rates [Har+16; BS20].
- **Equal Odds**: equalize True Positive and False positive rates [BS20].

Notice that Demographic Parity does not require to know the true labels to be computed. Actually, it does not even require the existence of some ground truth, and can therefore be applied to any situation that involves selecting people, even when the selection is arbitrary.

Notably, Barocas et al. [BHN19] have designed a framework that encompasses all the fairness constraints that can be defined based on those metrics. Chouldechova [Cho17] and Kleinberg [KMR16] have proved a fundamental impossibility result: outside of some trivial settings, it is impossible to simultaneously equalize False Positive rates, False Negative rates, and Predicted Positive rates across groups.

Outside of machine learning, those definitions are also widely used, especially Demographic Parity. The reason is that in many situations the outcome is binary: being admitted to a university, hired in a company, getting a grant or a loan, and many others. In all of those situations, it is quite straightforward to compute the acceptance rate inside each group and to compare them to evaluate if there is unequal treatment.

1.2.3 Statistical discrimination

In economics, when studying discrimination, a distinction is often made between **taste-based discrimination** and **statistical discrimination**. Taste-based discrimination occurs when a decision maker gives on average worse outcomes to specific group because of a negative opinion they hold about this group. On the other hand, statistical discrimination occurs when a decision maker or mechanism has no intentional bias towards a group but still discriminates against them because of imperfect information.

Statistical discrimination was first proposed by economists Phelps [Phe72] and Arrow [Arr73]. The Arrovian model assumes that some groups in the population of workers expect to be discriminated against, and therefore have a low incentive to invest time, effort or money in their applications. On the other hand, some groups do not expect to be discriminated, or even believe they will be advantaged, and therefore invest more in their applications since they have higher chances of being hired. As a result, companies will optimally hire those who have invested, reinforcing the prior belief of discriminated and advantages groups. On the other hand, Phelps models statistical discrimination as a result of noisy estimations of candidates qualities. If a decision-maker has very precise information about the quality of applicants from group 1, and very little information about the quality of candidates from group 2, it might lead to a different treatment of the two groups even when the latent qualities are equally distributed across groups.

The Phelpsian model has recently found an echo in the growing literature about discrimination related to both the generalized use of algorithmic decision making and the controversies around college admission in many countries. Kleinberg and Raghavan [KR18] study what they call **implicit bias**, i.e., a model where every candidate has a latent quality that follows the same distribution for every candidate, but for some group among them, the quality estimate made by the decision maker is their latent quality divided by some constant. This leads to obvious inequalities in the probabilities to be selected for those candidates. In the same fashion, Emelianov et al. [Eme+20; Eme+22], followed by Garg et al. [GLM21], proposed the following simple model for what they call **differential variance**. An employer needs to hire a fraction α of a set of applicants S, partitioned in two groups G_1, G_2. Each applicant s has a latent quality W_s, drawn according to a Gaussian distribution $\mathcal{N}(0,1)$. The employer only has access to a noisy estimate $\hat{W}_s$ of W_s, such that $\hat{W}_s = W_s + \sigma_{G(s)}\varepsilon_s$, where $G(s)$ is s' group, σ_1, σ_2 are the amount of noise for each group, and $\varepsilon_s \sim \mathcal{N}(0,1)$. They consider two types of decision-makers. The first one is called **group-oblivious**, because they do not take into account the noise and rank the applicants directly from the estimates $\hat{W}$. The second one is called **Bayesian**, because they know that the estimates are noisy and do not directly use $\hat{W}$ but rather the expected value of W conditioned on $\hat{W}$ and the group of the applicant:

$$\tilde{W}_s := \mathbb{E}[W_s|\hat{W}_s, G(s)] = \hat{W}_s/(1+\sigma^2_{G(s)}) \tag{1.1}$$

Consider the group-oblivious decision-maker. The distribution of observed qualities is $\mathcal{N}(0, 1+\sigma_1^2)$ for group G_1 and $\mathcal{N}(0, 1+\sigma_2^2)$ for group G_2. As illustrated on the left part of Figure 1.13, the highest values belong to the high noise group, and so do the lowest. As a consequence, if the number of positions is less than half of the population ($\alpha < 0.5$), then the high noise group will be overrepresented in the selected applicants, otherwise it is the low noise group that will be advantaged.

Conversely, with the Bayesian decision-maker, the distribution of observed qualities is $\mathcal{N}(0, 1/(1+\sigma_1^2))$, so the roles are reversed: the group with the highest noise variance becomes the group with the lowest variance of estimated quality. By the same principle, and as represented on the right of Figure 1.13, if $\alpha < 0.5$ the low noise group will be overrepresented, and if $\alpha > 0.5$ the high noise group will be overrepresented.

This model provides an explanation to the intuition that applicants with noisier information will be discriminated against: in many situations, the capacity is less than half of the applicants, and the decision maker is somewhat Bayesian in the sense that candidates with low information will be compressed to the mean value, while candidates with more information will be more likely to get extreme grades. Emelianov et al. go on to show that applying a demographic parity constraint (i.e., the fraction of applicants inside G_1 that are hired is equal to the fraction of applicants inside G_2 that are hired) improves the utility of the group-oblivious decision-maker, with utility taken as the sum of latent qualities of hired applicants. However, the utility of the Bayesian decision-maker cannot be improved: they already make the optimal use of the available information.

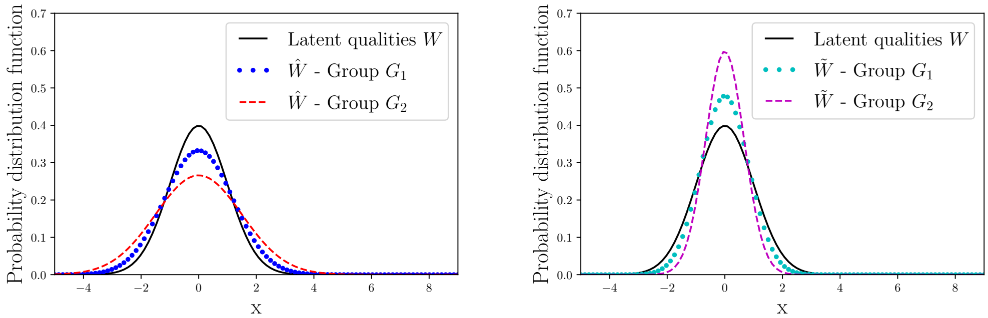


Fig. 1.13: Distributions of the estimated qualities for each group, with the latent qualities as a reference, $\sigma_1 = 0.2, \sigma_2 = 0.5$. Left: Group oblivious decision-maker. Right: Bayesian decision-maker.

1.3 Related literature

In this section, I present the existing literature related to the subject of this thesis, i.e., group fairness in matching problems, in order to understand the existing results, and what remains to be studied. Each chapter has in addition its own introduction to complete the context and related literature specific to that chapter.

1.3.1 Matching under fairness constraints

The bipartite matching literature has seen a rise in recent works pertaining to the conception and deployment of fair matching algorithms that prevent discrimination: [Chi+19; Ban+23] propose algorithms that efficiently approximate optimal fair matchings, and many authors [MX20; MXX23; Hos+23; San+21; Ban+23; Esm+22] consider fair matching in an online setting where individuals must be matched or discarded irrevocably once they become available.

Works on fairness in matching with preferences have considered various affirmative action policies, including upper and lower quotas, to reduce discrimination ([Abd05; KK15; DKT23b; Kri+22; DPS20]). For instance, Abdulkadiroglu [Abd05] defines a variant of the classical college admission model of [GS62], by adding the constraint that colleges cannot admit more than a certain number of students from each group. He then defines a notion of weak stability: a blocking pair (s, C) only counts if the college C has not reached its quota for the group of the student s. He then proposes a very slightly modified version of the DA algorithm (cf. Algorithm 4) that outputs the student optimal weakly stable matching, and shows that this mechanism is strategy-proof for students.

Kamada and Kojima [KK23] developed a general framework to implement constraints of any type in the college admission model, and in particular fairness constraints. To do this, they assume that each college C has a feasibility set $F_C \subseteq \mathbb{P}(S)$, i.e., they can only admit sets of students that belong to their feasibility set. When each college's feasibility set is a capacity constraint ($I \in F_C \Leftrightarrow |I| \leq q_C$ for some $q_C \in \mathbb{N}$), then this is the classical college admission problem. However, this framework allows to implement any constraint by explicitly specifying which sets of students are feasible. This could for instance allow to implement budget constraint, in the case where some students cost more money than others if admitted, e.g., disabled students, and therefore implement fairness rules towards those students while respecting colleges' budgets. They show that when the constraints are not only capacity constraints, a stable matching does not necessarily exist. They redefine a notion of weak stability by allowing waste, i.e., only justified envy constitutes a blocking pair, not waste. They show hat a weakly stable matching always exist when the feasibility sets are downward closed, i.e., any subset of a feasible set of students is also feasible. They then provide two algorithms that output the student-optimal weakly stable matching, one of them being quite close to DA. They further show that whenever the feasibility set of some college is not downward closed, there exists a set of students preferences such that there is not weakly stable matching, proving that downward closedness is the necessary and sufficient condition for the existence of a weakly stable matching.

1.3.2 Fairness of the unconstrained models

Another approach to study fairness in matching problems is to investigate the mechanisms that cause inequalities in the final matching. As we saw with statistical discrimination, inequalities do not always result from a conscious choice of the decision maker or the creator of an algorithm. If similar mechanisms exist in the matching setting, understanding them could lead to more efficient ways to root out inequalities.

One of the major findings in this area is due to Karni et al. [KRY22], who show that a fair ranking does not necessarily lead to a fair matching, in the sense of individual fairness. In their model, colleges rankings of students contain ties. They show that when using naive extensions of DA to this situation, including classical tie-breaking rules, students that are tied can have different outcome distributions, which is deemed (individually) unfair. They propose two new algorithms that extend respectively Student-proposing DA and College-proposing DA and produce matchings that are ε-close to a stable matching and ex-ante ε-close to a fair matching. Recently, [Dev+23] presented a framework to study fairness in matching problems. Their approach is original in the fact it acknowledges that the classical definition of individual fairness ensures that agents get similar outcomes but not necessarily with respect to their preferences. Instead, they introduce a new notion of individual fairness that states that similar agents should get resources that are in the same position in their respective preference lists.

On the group fairness side, Bommasani et al. [Bom+22] defines a notion of *systemic failure*, that measures how many applicants will get rejected by all decision-makers, as a metric describing the harm of algorithmic monoculture (all colleges using the same algorithm to evaluate students) to applicants. They assume that applicants belong to different groups and compute this metric on each group separately to show that some groups suffer from higher rates of systemic failure. On the bipartite matching side, [Chi+19] studies a model where vertices on one side of the graph are divided into groups, and looks into which matchings respect demographic parity. Their model in fact extends more generally to matroids, a notion I explore in more details in Chapter 2.

1.4 Purpose and structure

While the literature on matching is vast, and so is the literature on fairness, the question of group fairness in matching problems has received less attention, and most works in this area focus on designing algorithms that enforce some fairness constraint in an existing matching model. This thesis provides an original theoretical analysis of the sources of unfairness that lie at the core of matching models, of the amplitude of those inequalities, and of their dynamics with respect to the parameters of the model.

In Chapter 2, I study bipartite matching. Like [Chi+19], we assume that vertices on one side of the graph are partitioned into groups. We propose an original geometrical representation of the set of feasible matchings. We then define a broad class of fairness notions, and use both geometric and combinatorial arguments to study the Price of Fairness, i.e., the distance between the set of maximum matchings and the set of fair matchings, for different fairness notions.

In Chapter 3, I focus on two-sided matching, and more particularly on the role played by correlation of colleges rankings of students. We introduce a model based on copulas that allows a precise and flexible modeling of this correlation. We then study the influence of correlation on efficiency, i.e., the amount of students getting their top choice, and inequality, defined as the difference in the rates at which students from different groups go unassigned.

I conclude with a discussion of the accomplished and remaining work, and a more general reflection about the importance of fairness and the role of mechanism design for society.

2

Bipartite matching: Matroid representation and Price of Fairness

Contents

This chapter is based on the article [Cas+24] by Mathieu Molina, Felipe Garrido-Lucero, Simon Mauras, Patrick Loiseau, Vianney Perchet and myself. I provided a major contribution to all the contents presented here except Section 2.6.3 that is nonetheless included for completeness.

2.1 Introduction

In this chapter, we study the Price of Fairness in the context of bipartite matching problems (without preferences). We focus on the Price of Fairness, a quantity that measures the loss in utility due to fairness requirements. The Price of Fairness has been studied in other problems of resource allocation [BFT11], kidney exchange [DPS14], and fair division [Bei+19]. We show that certain fairness notions incur no loss of utility, while others may have a Price of Fairness bounded or unbounded.

2.1.1 Our contributions

We consider a model of cardinal bipartite matching with agents and jobs, where fairness is required on the agents side. The agents are partitioned into K disjoint groups, that could for instance be based on sensitive attributes such as gender or ethnicity. For most of the applications mentioned, matching algorithms are deployed on a large market, hence fractional matchings can serve as good approximations. For this reason, we mainly work (except when stated otherwise) with fractional matchings.

First, we frame group fairness as a geometric problem, where each matching is represented by a point $x \in \mathbb{R}^K$ where x_i denotes the number of jobs matched to agents of group i. We characterize the set of feasible and maximum matchings as a polytope. Specifically, we show that the set of feasible matchings when taking into account the number of agents matched per group is a polymatroid – an intricate extension of (transversal) matroids. This implies that the Pareto frontier of the feasible matchings in $\mathbb{R}^K$, the set of maximal matchings, and the convex hull of all lexicographic maximal matchings are all equal. Hence many natural fairness notions, as for example the leximin egalitarian rule from fair division, can be achieved with a maximal matching at no utility loss. We also show, using again the structure of polymatroid, that such leximin optimal fair point satisfies additional properties like minimizing the variance of utility across groups.

Second, we introduce w-weighted group fairness notions, which seek to equalize the fraction of the entitlement w_i that each group receives. Compared to the leximin fairness, this stronger fairness notion does not allow for groups to be better off than others. Weigthed fairness is quite flexible, and can encompass many fairness concepts inspired from the fair machine learning literature, such as demographic paritity and equality of opportunity. In particular, we define opportunity fairness by setting w_i as the maximum number of agents from group i who can be matched. Defining the Price of Fairness (PoF) as the ratio between the optimum without and with fairness constraints (1 corresponds to no utility loss), we show that the worst case opportunity-PoF is equal to $K - 1$. This bound is independent of the number of jobs and agents in the graph (the size of the graph), and only depends linearly on the number of groups. As a significant consequence, any instance with only two groups has no diminution of the maximal matching size under opportunity fairness. Finally, we provide refined bounds under specific conditions such as (1) having a fixed ratio between the optimum and the ideal objective or (2) having graphs sampled from an Erdös-Rényi model.

2.1.2 Further related works

Matchings and matroids In terms of techniques, our characterizations use tools from bipartite matching and matroid theory [Oxl22]. In a bipartite graph, sets of left endpoints of matchings form a transversal matroid, from which one can define the independence polytope (convex hull of feasible sets) and the basis polytope (convex hull of maximal sets). This corresponds to a special case of our model when groups have size 1. We show that this construction extends to arbitrary groups size, where the (transversal) matroid becomes a discrete polymatroid [HH02], and the independence polytope becomes a continuous polymatroid [Edm70]. Notably, our work is the first to define this particular (discrete) polymatroid. Matroids are also used to study connected matching problems, such as kidney exchange or college admission with reserves [SY22]. For both settings [EF65] showed that the set of acceptable matchings (respecting reserves for college admission, individually rational for kidneys) are matroids. In the kidney problem, [RSU05] further showed that a matching is Pareto efficient and individually rational if and only if it is a basis of the associated matroid. As for college admission with reserves, since reserves can be used for affirmative action purposes, it is an example of how matching problems with fairness constraints can be represented by matroids. The main difference is that the matroid in [SY22] is the set of matchings that respect the reserve rule, while in our model the matroid contains all feasible matchings and the fairness constraints are imposed on the matroid afterwards.

Fair division An entire literature is dedicated to the fair division of items between K players, with guarantees either share-based (e.g., proportionality where each players gets a $1/K$ fraction of all items) or envy-based (e.g., envy-freeness where no players prefers the bundle allocated to another player). In mathematics, fair division emerged from the problem of dividing continuous goods, through the seminal works of Steinhaus [Ste49], who defined proportionality, and Foley [Fol66] and Variant [Var74], who defined envy-freeness, later generalized by Weller [Wel85].

More recently, computer scientists have considered the discrete version of the problem of sharing indivisible goods, with relaxations of share-based guarantees such as MMS [Bud11] and envy-based guarantees such as EF1 [Lip+04; Bud11] or EFX [Car+19]. This framework is closely related to our setting where groups can be seen as players who benefit from bundles of jobs and have valuations given by the number of agents they can match. Looking closely at this reduction, the valuation of each player (group) is a matroid rank function for which improved guarantees can be obtained [BEF21; Ben+21; BV22; VZ23]. Notably, these recent works propose efficient algorithms that select maximal size matchings which satisfy various fairness properties (leximin, EF1, MMS).

Our model distinguishes from these existing results as we consider a restricted continuous setting (fractional matchings) for which strong guarantees such as proportionality and envy-freeness are easily achievable, and for which we aim at an stronger fairness property by equalizing matching rates. In particular, we remark that our maximal fair fractional matching easily satisfies proportionality and envy-freeness.

2.2 Model

In this section we introduce the model considered in this chapter. A summary of the main notation used in the chapter is provided in Table 2.1.

Consider a bipartite non-directed graph $\mathcal{G} = (U, V, E)$, with U the set of **jobs**, V the set of **agents**, and $E \subseteq U \times V$ set of **edges**. We assume the graph $\mathcal{G}$ **known**.

The set of **feasible matchings** on the graph $\mathcal{G} = (U, V, E)$, denoted by $\mathcal{M}(\mathcal{G})$, is defined as the family of all subsets of E that do not include two edges with common extremes. As equivalently showed by Edmonds [Edm79], $\mathcal{M}(\mathcal{G})$ corresponds to all binary matrices $\mu \in \{0,1\}^{|U|\times|V|}$ satisfying,

$$\forall u \in U, \sum_{v' \in V} \mu(u, v') \leq 1, \forall v \in V, \sum_{u' \in U} \mu(u', v) \leq 1, \text{ and } \mu(u,v) = 1 \implies (u,v) \in E.$$

We drop the dependence on $\mathcal{G}$ in any posterior definition. Notice from the matching definition that jobs and agents may remain unmatched. **Fractional matchings** are obtained when relaxing the integrality in the previous definition.

Agents are partitioned into K **groups** $G_1, ..., G_K$, where $G_i \cap G_j = \varnothing$ for $i \neq j \in [K]$, and $V = \bigcup_{i\in[K]} G_i$, with $[K] := \{1, 2, ..., K\}$. An illustration is given in Figure 2.1 with three groups represented by the three different shapes.

We will be interested in studying the number of agents matched per group by a given matching μ. In order to do it, we will consider the following **geometric approach**. Let $X : \mathcal{M} \to \mathbb{R}^K$ be the mapping given by $\mu \longmapsto X(\mu) = (X_1(\mu), X_2(\mu), ..., X_K(\mu))$, where $X_i(\mu) := \sum_{u\in U} \sum_{v\in G_i} \mu(u, v)$ denotes the number of G_i-agents matched by μ. The mapping X captures **anonymity** as it does not keep track of the matched agents' identity but only the number of agents per group matched.

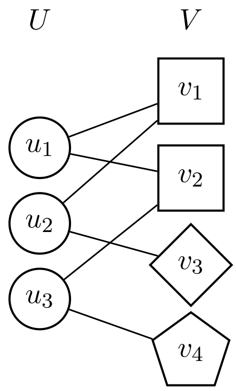


Fig. 2.1: Example of a bipartite graph with 3 groups on the right side, represented by the different shapes.

Definition 2.1. A point $x \in \mathbb{R}^K$ is **realizable** if there exists $\mu \in \mathcal{M}$ such that $X(\mu) = x$. We denote the set of **realizable points** in $\mathbb{R}^K$, i.e., the image set $X(\mathcal{M})$, as $\mathbf{M}$[1]. The mapping X being linear in each coordinate, the convex hull of $\mathbf{M}$ corresponds to the image through X of the set of fractional matchings, i.e., $\mathrm{co}(\mathbf{M})$ is the set of **fractional realizable points**.

The classical **maximum size matching problem** focuses on matching as many agents as possible given a graph $\mathcal{G}$. Using the geometric approach, the set of **maximum size matchings** $\mathcal{P}$ in the graph $\mathcal{G}$ can be written as

$$\mathcal{P} := \operatorname*{argmax}_{\mu \in \mathcal{M}} \left\{ \sum_{i=1}^{K} X_i(\mu) \right\} = \operatorname*{argmax}_{\mu \in \mathcal{M}} \{ \|X(\mu)\|_1 \}.$$

The set $\mathcal{P}$ is always non-empty due to the finiteness of the graph $\mathcal{G}$. We denote $\mathbf{P}$ its image through the map X, i.e., $\mathbf{P} = X(\mathcal{P})$. We denote by M_Λ the maximum number of agents of groups $\Lambda \subseteq [K]$ that can be ever matched, that is,

$$M_\Lambda := \max_{\mu \in \mathcal{M}} \sum_{i \in \Lambda} X_i(\mu), \quad \forall \Lambda \subseteq [K].$$

In particular, the maximum number of agents we can match is equal to $M_{[K]}$. Finally, we denote $e_i \in \mathbb{R}^K$ the i-th canonical vector.

We will now see how we can take into account the membership of agents in the maximization problem.

[1]Remark the notation for the sets $\mathcal{M}$ and $\mathbf{M}$. A cursive capital letter will always represent a set of matrices, while a capital bold letter will always represent a set of points in $\mathbb{R}^K$.

2.3 Geometry of integral and fractional matchings

The maximum size matching problem can actually be expressed as a multi-objective optimization problem (MOOP) given by the maximization of the number of agents matched for each group G_i, which corresponds to the maximization of all entries of the K-dimension vector $x \in \mathbf{M}$:

$$\max\{x_1 \in \mathbb{R}, x_2 \in \mathbb{R}, ..., x_K \in \mathbb{R} \mid x = (x_1, x_2, ..., x_K) \in \mathbf{M}\},$$

As we will prove later, the set of maximal matchings $\mathbf{P}$ will correspond to the Pareto-frontier of the set $\mathbf{M}$, which motivates its notation.

Hence, we dedicate this section to the characterization of the sets $\mathbf{M}$ and $\mathbf{P}$ and their convex hull, respectively.

2.3.1 The discrete polymatroid M

The set $\mathbf{M}$ results to be a discrete polymatroid [HH02], the generalization of matroids to multisets. Remark we are the first in the literature to study this construction for the set $\mathbf{M}$. The proof is a non-trivial extension of the fact that sets of agents who can be matched in some matching form a (transversal) matroid.

Theorem 2.1. *The set* $\mathbf{M}$ *is a* discrete polymatroid*, that is,*

- *Whenever* $x \in \mathbb{N}^K$ *and* $y \in \mathbf{M}$ *such that* $x \leq y$ *(coordinate wise), then* $x \in \mathbf{M}$*,*
- *Whenever* $x, y \in \mathbf{M}$ *and* $\|x\|_1 < \|y\|_1$*, there exists* $i \in [K]$ *such that* $x_i < y_i$ *and* $x + e_i \in \mathbf{M}$*.*

Proof. The first point is direct as unmatching agents does not affect the realizability of a matching. To prove the second point, select two matchings $\mu \subseteq E$ and $\nu \subseteq E$ such that $x = X(\mu)$ and $y = X(\nu)$. First, we build the symmetric difference $\delta = \mu \triangle \nu$. Observe that δ is a subgraph where all vertices have degree at most 2 and, therefore, it corresponds to a collection of cycles and paths.

Because $\|x\|_1 < \|y\|_1$, the pigeon-hole principle implies the existence of at least one path with one endpoint G_1 in V and the other in U, such that G_1 is matched by ν. If G_1 is in group i_1, then swapping the edges along the path shows that $x + e_{i_1}$ is feasible. However, notice that it does not necessarily hold that $x_{i_1} < y_{i_1}$ (see Figure 2.2 for an example).

To obtain i such that $x + e_i$ is feasible and $x_i < y_i$, we build the “exchange graph”, with one vertex per group $i \in [K]$, plus one special vertex 0.

- For each path in δ such that both endpoints are in V, we draw an arc from i to j, where i is the group of the endpoint matched by μ, and j is the group of the endpoint matched by ν.
- For each path in δ with exactly one endpoint in V, of group $i \in [K]$ and matched by μ, we draw an arc from i to 0.
- For each path in δ with exactly one endpoint in V, of group $i \in [K]$ and matched by ν, we draw an arc from 0 to i.

Notice that multiple edges may exist between two vertices. Starting at $i_0 = 0$, we pick the outgoing arc going to group i_1, and we continue to build a path greedily, until to get stuck at some node i_ℓ as we have exhausted all outgoing arcs. First, observe $x + e_{i_\ell}$ is feasible (swap all the paths in δ corresponding to the arcs used in the exchange graph).

Second, we show that $x_{i_\ell} < y_{i_\ell}$. We denote by $\deg^-(i)$ (resp. $\deg^+(i)$) the in-degree (resp. out-degree) of node $i \in \{0, 1, \ldots, K\}$. By construction, we have that $\deg^+(0) - \deg^-(0) = \|y\|_1 - \|x\|_1 > 0$, and that $\deg^+(i) - \deg^-(i) = x_i - y_i$ for each group $i \in [K]$. If we are stuck at i_ℓ, it is because $\deg^+(i_\ell) < \deg^-(i_\ell)$, i.e. because $x_{i_\ell} < y_{i_\ell}$. ■

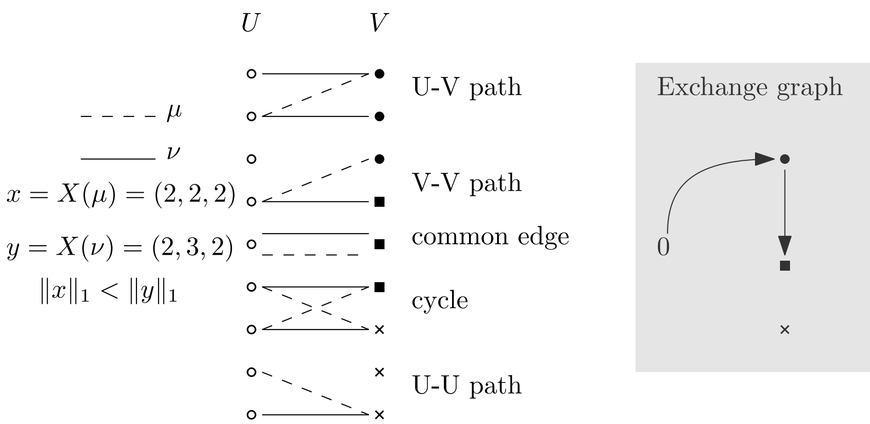


Fig. 2.2: Proof of Theorem 2.1. Matchings μ and ν are drawn on the left, other edges are not represented. V-V paths create arcs between the corresponding groups in the exchange graph from 0 while U-V paths creates arcs to 0.

Corollary 2.2. *The set* $\mathbf{P}$ *of points in* $\mathbf{M}$ *with maximum* $\|\cdot\|_1$ *corresponds to the Pareto frontier of* $\mathbf{M}$*, that is, the set of non-Pareto dominated points in* $\mathbf{M}$.

Proof. Direct from applying the augmentation property of Theorem 2.1. ■

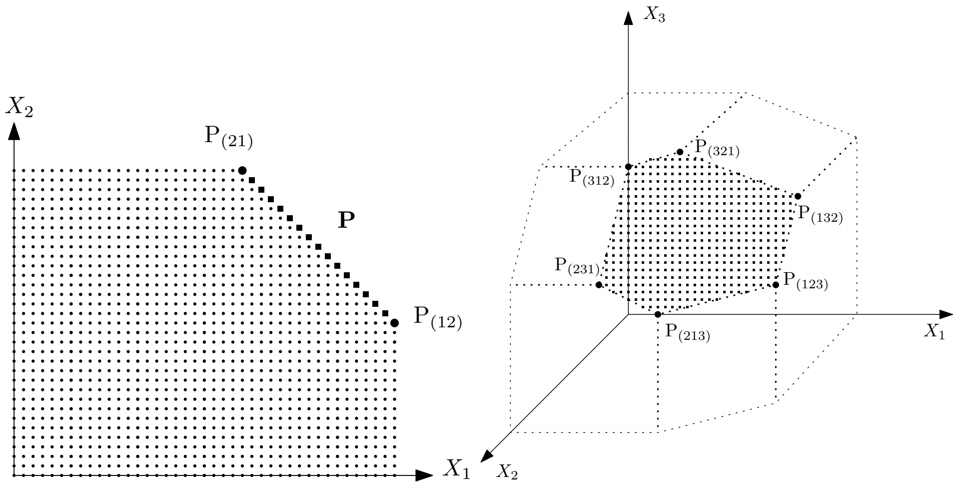


Fig. 2.3: Sets $\mathbf{M}$, $P_\sigma, \forall \sigma \in \Sigma([K])$, and $\mathbf{P}$, for $K = 2$ and $K = 3$.

2.3.2 The set of lexicographic maximum size matchings

A **permutation** of $[K]$ is a bijection function $\sigma : [K] \to [K]$. The set of permutations of $[K]$ is denoted $\Sigma([K])$. For $K = 3$, we write $\sigma = (132)$ to denote $\sigma(1) = 1, \sigma(2) = 3$, and $\sigma(3) = 2$.

Definition 2.2. Let $\sigma \in \Sigma([K])$ be fixed. We define the set of **lexicographically maximum size matching** as $\mathcal{P}_\sigma := \text{argmax}_{>_{\mathcal{L}(\sigma)}} \{X(\mu) : \mu \in \mathcal{M}\}$, where $>_{\mathcal{L}(\sigma)}$ denotes the lexicographic order in σ. We denote $P_\sigma = X(\mathcal{P}_\sigma)$ to the image through X of the set $\mathcal{P}_\sigma$.

Notice that the finiteness of $\mathcal{G}$ always implies the non-emptiness of $\mathcal{P}_\sigma$, for any permutation $\sigma \in \Sigma([K])$. Moreover, P_σ corresponds to a unique point in $\mathbb{R}^K$. Figure 2.3 illustrates graphs with $K = 2$ and $K = 3$ groups, respectively. The set of realizable points $\mathbf{M}$ is represented by the whole integer polytope, the points P_σ, for $\sigma \in \Sigma([K])$, by the circles, and the set $\mathbf{P}$ by the squares together with the circles.

Remark 2.1. Given a permutation $\sigma \in \Sigma([K])$, computing P_σ amounts to taking $x \equiv 0_K \in \mathbb{R}^K$ and **sequentially** maximizing its entries in the order given by σ. In particular, each P_σ can be **computed in polynomial time** on the size of the graph $\mathcal{G}$ by running K sequential flow algorithms (such as Ford-Fulkerson [FF56]). From a geometrical point of view, finding P_σ is done by running a serial dictatorship process (Algorithm 5) as illustrated in Figure 2.4 for $K = 3$ and $\sigma = (132)$.

We conclude this section with the following useful results.

Proposition 2.3. P_σ *is the only point* $x \in \mathbf{M}$ *which maximizes each of the following two (equivalent) objectives:*

Algorithm 5: Serial dictatorship

Input:
Graph $\mathcal{G}$ and permutation $\sigma \in \Sigma([K])$
Output:
The lexicographic maximum size point P_σ
Initialization:
$x \leftarrow 0_K \in \mathbb{R}^K$.
for $i \in [K]$ **do**
  maximize $x_{\sigma(i)}$ such that x is feasible:
    $x_{\sigma(i)} \leftarrow \max\{t \geq 0 \,|\, x + t \cdot e_{\sigma(i)} \in \mathbf{M}\}$
end
Return x.

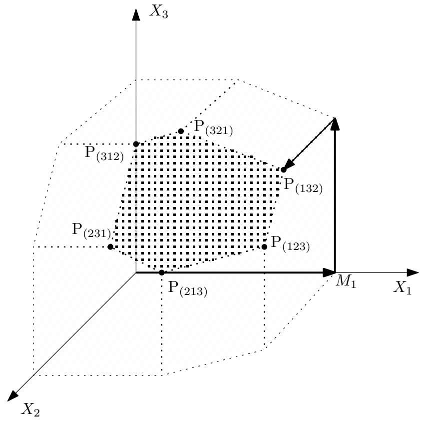


Fig. 2.4: Serial dictatorship, $K = 3$ and $\sigma = (132)$.

1. $\sum_{j \in [i]} x_{\sigma(j)}$, *for all* $i \in [K]$.
2. $\sum_{j \in [K]} \lambda_j x_{\sigma(j)}$, *for all* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq 0$.

Proof sketch. We assume the result does not hold and use the augmentation property of the polymatroid to show a contradiction. The full proof is provided in Section 2.9.4. ∎

Corollary 2.4. *For each permutation* σ *the point* P_σ *is equal to*

$$\mathrm{P}_\sigma = (M_{\sigma([i])} - M_{\sigma([i-1])})_{i \in [K]},$$

where $\sigma([i]) = \{\sigma(1), \ldots, \sigma(i)\}$ *is the set containing the first* i *agents of* σ.

Proof. Corollary of property (1) of Proposition 2.3. ∎

2.3.3 The polytope of fractional matchings

As Edmonds did for matchings [Edm79], we can characterize the convex hull of the set of realizable points $\mathbf{M}$ as the intersection of finitely many hyperplanes. To do this, we show that the convex hull of $\mathbf{M}$ is a polymatroid [Edm70], which generalizes the matroid polytopes to multisets.

Proposition 2.5. *The convex hull of* $\mathbf{M}$*, denoted* $\mathrm{co}(\mathbf{M})$*, is a* polymatroid*, that is,*

- *Whenever* $x \in \mathbb{R}^K$ *and* $y \in \text{co}(\mathbf{M})$ *such that* $x \leq y$ *(coordinate wise), then* $x \in \text{co}(\mathbf{M})$*,*

- *Whenever* $x, y \in \text{co}(\mathbf{M})$ *and* $\|x\|_1 < \|y\|_1$*, there exists* $i \in [K]$ *and* $\varepsilon > 0$ *such that* $x_i < y_i$ *and* $x + \varepsilon e_i \in \text{co}(\mathbf{M})$*.*

Proof. Using Theorem 2.1 and [HH02, Theorem 3.4]. ■

Proposition 2.6. *A point* $x \in \mathbb{R}_+^K$ *belongs to* $\text{co}(\mathbf{M})$*, the convex hull of the set* $\mathbf{M}$*, if and only if,*

$$\forall \Lambda \subseteq [K], \quad \sum_{i \in \Lambda} x_i \leq M_\Lambda.$$

Proof. Using Proposition 2.5 and [HH02, Proposition 1.2]. ■

Finally, we prove that the set of maximum size matchings corresponds to the convex combination of the lexicographic maximum size matchings.

Proposition 2.7. *The Pareto frontier* $\text{co}(\mathbf{P})$ *has the following characterizations.*

- *inequalities:* $\text{co}(\mathbf{P}) = \text{co}(\mathbf{M}) \cap \{x \in \mathbb{R}_+^K : \sum_{i \in [K]} x_i = M_{[K]}\}$

- *extreme points:* $\text{co}(\mathbf{P}) = \text{co}(\{\mathrm{P}_\sigma : \sigma \in \Sigma([K])\})$

Proof. The characterization with inequalities follows from Proposition 2.6. To prove the characterization with extreme points, we use [BCT85, Theorem 2.4]. ■

As it will be shown when studying fair maximal matchings or the Price of Fairness (Section 2.6.1), working with fractional matchings will allow us to exploit the geometric properties of the set $\text{co}(\mathbf{P})$ and to bound the Price of Fairness even for graphs where the only integral fair matching will be the empty one. Therefore, for most of the fairness discussions, **we will relax the integrality condition** and focus on the polytope of fractional matchings.

2.4 The fairest optimal matching

Once seen that all Pareto optimal matchings have the same size, we turn our attention to the question of finding which of them is the "fairest" one. Indeed, due to the large number of options that the set of maximal matchings represents, there is an interest for the central planner to select only among those ones which satisfy some additional criteria, such as fairness. Therefore, we present two fairness notions which can always (for any graph) be guaranteed at no loss of optimality. For the two solution concepts, we propose their definitions and discuss their relative geometric, procedural, and axiomatic fairness.

1. Among all possible Pareto optimal matchings, one might be tempted to chose one *in the middle*. For that matter, we consider the barycenter of the extreme points of $\mathrm{co}(\mathbf{P})$. The intuition behind the fairness of this point comes from two sources: (1) it corresponds to the expected output of the random serial dictatorship procedure which makes it algorithmically fair and (2) it guarantees each group their Shapley value in a given cooperative game, a standard notion from game theory, making it axiomatically fair.

2. One rule often used in social choice [Sen17] is the egalitarian rule (or Rawlsian fairness): the selected matching needs to maximize $\min_{i \in [K]} x_i$, i.e., the goal is to ensure that worse off groups do as good as possible. Due to the multiplicity of solutions, a tie-break rule is to choose the one simultaneously maximizing the second minimum. In case of having several options, the third minimum is considered, and so on. This solution concept is known as the leximin rule, which has also been studied in the social choice literature [dG77; DG78].

2.4.1 The Shapley fairness

Denote $v(\Lambda) = M_\Lambda$, for any $\Lambda \subseteq [K]$. The pair $([K], v)$ defines a cooperative-game with a sub-additive value function. A classical fair solution concept in cooperative game theory is the Shapley value where players are rewarded their average marginal contribution. Formally, we define the **Shapley value** φ_i of group $i \in [K]$ as,

$$\varphi_i := \frac{1}{K!} \sum_{\sigma \in \Sigma([K])} M_{\sigma([i])} - M_{\sigma([i-1])}.$$

Notice that $M_{\sigma([i])} - M_{\sigma([i-1])} = (\mathrm{P}_\sigma)_i$ corresponds to the number of G_i-agents matched after sequentially matching the agents in $G_{\sigma(1)}, ..., G_{\sigma(i-1)}$. In particular, the vector $x \in \mathbb{R}^K$ given by $x_i = \varphi_i, \forall i \in [K]$, corresponds to the barycenter of the extremal points[2] P_σ of the set of maximal matchings, as illustrated in Figure 2.5.

[2] It is important to remark that the point $(\varphi_1, ..., \varphi_K) \in \mathbb{R}^K$ corresponds to the barycenter of the extreme points of $\mathbf{P}$ and not the barycenter of $\mathbf{P}$, as the multiplicity of these points must be considered if some of them coincide.

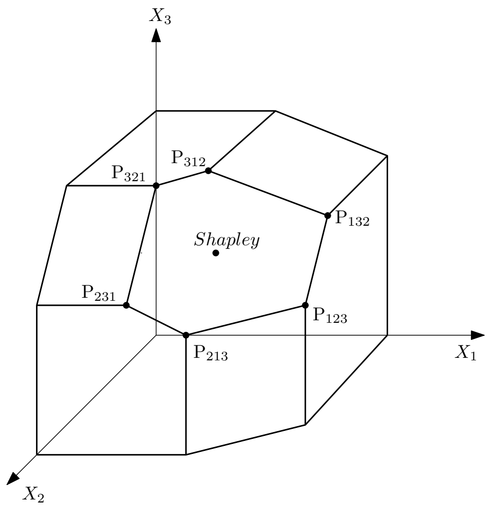


Fig. 2.5: Shapley value

We call to this vector the **Shapley matching**. The following proposition is immediate from the barycenter characterization.

Proposition 2.8. *The vector $x \in \mathbb{R}^K$ defined by $x_i = \varphi_i, \forall i \in [K]$, is always realizable, it has maximum size, and it lies in the barycenter of the extreme points* $\{\mathrm{P}_\sigma, \forall \sigma \in \Sigma([K])\}$.

2.4.2 Leximin rule

Among the optimal points in $\mathbf{P}$, the leximin maximum matching will find an interesting algorithmic interpretation and many geometrical properties. We dedicate this section to its study.

Definition 2.3. Given $x \in \mathbb{R}_+^K$ and $i \in [K]$, we define $x_{(i)}$ as the i-th smallest coordinate of x. We define the *leximin* ordering $>_{\min}$ over $\mathbb{R}_+^K$, comparing sorted vectors lexicographically. Formally,

$$x >_{\min} y \iff (x_{(1)}, \ldots, x_{(K)}) >_{\mathcal{L}} (y_{(1)}, \ldots, y_{(K)}).$$

Using the fact that $\mathbf{M}$ is convex and compact, the leximin preorder has a unique maximum [Beh77], which belongs to the Pareto frontier $\mathbf{P}$, and which we will denote $\mathrm{P}_{\text{leximin}}$. This implies, in particular, that the leximin has maximum size.

Remark 2.2. Computing $\mathrm{P}_{\text{leximin}}$ amounts to taking $x \equiv 0_K \in \mathbb{R}^K$ and continuously increasing all entries at rate 1, until reaching a facet of $\mathbf{M}$, i.e., until a constraint in Proposition 2.6 becomes tight. Then, freezing all entries in this tight constraint, continue increasing the others. Repeat the procedure until reaching the Pareto frontier $\mathbf{P}$. This "waterfilling" algorithm is sometimes referred to as probabilistic serial [BM01]. We state its pseudo-code in Algorithm 6. Unlike

Algorithm 6: Probabilistic serial

Input: Graph $\mathcal{G}$.

Output: The leximin maximum point $\mathrm{P}_{\text{leximin}}$.

Initialization:

$x \leftarrow 0_K \in \mathbb{R}^K$.

$S \leftarrow [K]$

while $S \neq \emptyset$ **do**

  maximize t such that $x + t \cdot \mathbb{1}_S \in \mathrm{co}(\mathbf{M})$,

    $x \leftarrow x + t \cdot \mathbb{1}_S$.

  for $\Lambda \subset [K]$ such that $\sum_{i \in \Lambda} x_i = M_\Lambda$,

    $S \leftarrow S \setminus \Lambda$.

end

Return x.

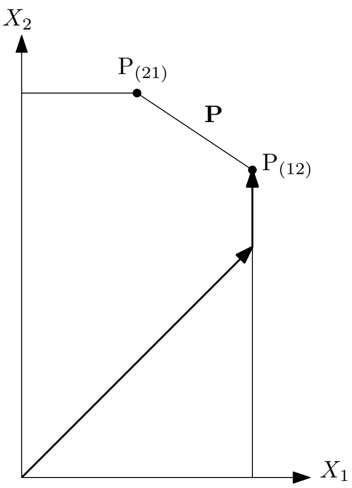


Fig. 2.6: Probabilistic serial, $K = 2$.

most matroid problems, our setting allows to compute the maximal t and Λ in Algorithm 6 in **polynomial time** (through linear programming) as constraints come from a matching problem. Section 2.9.3 proves it formally.

Proposition 2.9. $\mathrm{P}_{\text{leximin}}$ *is the only point* $x \in \mathbf{M}$ *which maximizes each of the following objectives:*

1. $\sum_{i \in [K]} \min(t, x_i)$, *for all* $t \in \mathbb{R}_+$.
2. $\sum_{i \in [j]} x_{(i)}$, *for all* $j \in [K]$.
3. $\sum_{i \in [K]} \lambda_i x_{(i)}$, *for all* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq 0$.

Proof sketch. Properties (2) and (3) are implied by the water filling maximization (1), which itself is a result of the augmentation property. The full proof is provided in Section 2.9.4. ■

Proposition 2.10. *The point* $x = \mathrm{P}_{\text{leximin}}$ *is the unique point in* $\mathbf{P}$ *that* minimizes *each of the following objectives:*

1. $\|x\|_p$, *for all* $p > 1$.
2. $\mathbf{Var}(x) := \frac{1}{K} \sum_{i \in [K]} (x_i - \frac{1}{K} \sum_{\ell \in [K]} x_\ell)^2$

Proof sketch. Using Proposition 2.9 for $\mathrm{P}_{\text{leximin}}$, we can deduce that it majorizes all points in $\mathbf{P}$. By applying Karamata's inequality and the strict convexity of the norms considered, we can deduce the first result. The second result comes from all points in $\mathbf{P}$ summing to the same quantity. The full proof is provided in Section 2.9.4. ■

By the same arguments as above, the leximin optimal point also uniquely minimizes the Gini coefficient over $\mathbf{P}$ as it is strictly Schur-Convex. An analogous property has been proved in the fair division literature [BEF21; Ben+21], who show that in the integral setting, leximin minimizes symmetric strictly convex function (such as sum of squares) among all utilitarian optimal allocations. Here we show that this holds true when matchings are fractional, using a slightly different proof technique.

Denote $\mathbf{1} \in \mathbb{R}^K$ the vector with only ones. Interestingly, we can project back and forth between $F_{\mathbf{1}} := \{t\mathbf{1} \mid t \geq 0\}$ and $\mathbf{P}$ depending on whether the fairest-optimal or the optimum fair point should be selected.

Proposition 2.11. *Let* $x = \mathrm{P}_{\text{leximin}}$ *and* $y = t^*\mathbf{1}$, *where* $t^* = \max\{t \geq 0 | t\mathbf{1} \in \mathrm{co}(\mathbf{M})\}$. *It holds that* x *is the* $\|\cdot\|_2$ *projection on* $\mathbf{P}$ *for any point in* $F_{\mathbf{1}}$, *and that* y *is the* $\|\cdot\|_2$ *projection on* $F_{\mathbf{1}} \cap \mathrm{co}\,(\mathbf{M})$ *for any point in* $\mathbf{P}$.

Proof sketch. The proof relies on Proposition 2.10 and the fact that $F_{\mathbf{1}}$ and $\mathbf{P}$ are orthogonal. The full proof is provided in Section 2.9.4. ■

In this work, we are interested in two (possibly competing) objectives: fairness and matching size. So far we have characterized some of the possible fair matchings among the maximal ones. The following section will focus on the opposite approach: fix a set of matchings which satisfy a given fairness property first and then, choose the largest one among those.

2.5 Weighted fairness and price of fairness

A basic and straightforward fairness notion was presented in the previous section through $F_{\mathbf{1}}$ by requiring to match all groups equally. However, this is unlikely to fit many use cases where the size of some groups might be much smaller than others, resulting in an important loss in optimality. We define a general class of fairness rules which are able to take into account the graph properties and to encompass many classical rules.

2.5.1 Weighted Fairness

Definition 2.4. Let $w = (w_i)_{i\in[K]} \in \mathbb{R}_+^K$ be a fixed weighted vector. We say that an element $x \in \mathbb{R}^K$ is **weighted-fair**, or simply w-**fair**, if for any $i, j \in [K]$,

$$\frac{x_i}{w_i} = \frac{x_j}{w_j}.$$

The set of w-fair points in $\mathbb{R}^K$ is denoted $\mathbf{F}_w$. Notice that weighted-fairness can be represented in $\mathbb{R}^K$ as the only line connecting the origin and the point $(w_1, w_2, ..., w_K)$, as showed in Figure 2.7 for the following three examples of weighted-fairness notions.

1. **Egalitarian fairness**: for any $i \in [K]$, $w_i = 1$.

2. **Demographic fairness**: for any $i \in [K]$, $w_i = |G_i|$.

3. **Opportunity fairness**: for any $i \in [K]$, $w_i = M_i$.

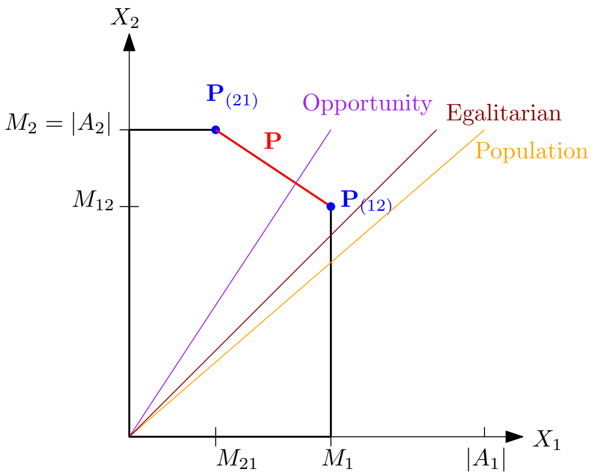


Fig. 2.7: Fairness notions

Relationship with fairness concepts of Section 2.4: Weighted-fairness is intricately linked to the Shapley fairness and leximin rule previously presented. Shapley fairness corresponds to considering weights $w_i = \varphi_i$ for any $i \in [K]$. Regarding the leximin rule, there are two possible interpretations. Firstly, extending the leximin notion to its weighted version where the objective is to sequentially maximize the entries of the vector $x^w := (x_i/w_i)_{i\in[K]}$ (the previous definition is recovered for egalitarian weights $w_i = 1$), it is clear that the optimum among all w-fair matchings corresponds to the one where all entries are equal to the minimum of the weighted leximin. Indeed, when running the water-filling process, instead of continuing to optimize other entries, the optimal w-fair matching stops at the first constraint saturation. Weighted fairness is, somehow, a *strong* fairness notion as no group envies the others' allocation based on their respective entitlements while the weighted leximin rule is a *weak* fairness notion that allows for some groups to be better off.

Remarkably, the set of fair matchings and the (sub)-set of optimal matchings are also linked geometrically through the leximin and the optimum among w-fair matchings: as can be seen in Proposition 2.11 for the egalitarian fairness, the projection of F_1 over the set of maximal matchings corresponds to the leximin point, while the projection of the maximal matchings over F_1 corresponds to the fair optimum. Thus, it is possible to *project back and forth* between the two notions. This is shown more generally for any $w > 0$ in Section 2.9.2, where instead of maximal matchings, we work with the set of **w**-weighted maximal matchings.

Weighted-fairness relates to group-fairness in Machine Learning when looking at predictions as selected matchings ($\hat{Y} = 1$ if the agent is matched) and true labels as entitlement (e.g., $Y = 1$ if the agent would be matched if only her group was present). For a group containing N_i agents, M_i of which could possibly be matched ignoring other groups, and x_i of them being currently matched, then demographic parity corresponds to equalizing x_i/N_i (demographic fairness), and equal opportunity corresponds to equalizing x_i/M_i (opportunity fairness). This justifies the names selection for the corresponding weighted fairness notions.

2.5.2 Price of Fairness

The following measure is at the heart of our analysis and measures the loss on optimality suffered when weighted fairness constraints are imposed.

Definition 2.5. Given $w \in \mathbb{R}_+^K$ fixed, we define the w-Price of Fairness (PoF_w) as,

$$\mathrm{PoF}_w = \frac{\max_{x \in \mathbf{M}} \|x\|_1}{\max_{x \in \mathbf{F}_\mathrm{w} \cap \mathrm{co}(\mathbf{M})} \|x\|_1}.$$

The set $\mathbf{F}_\mathrm{w} \cap \mathrm{co}(\mathbf{M})$ is always non-empty as the empty matching is always w-fair. Whenever $\mathbf{F}_\mathrm{w} \cap \mathrm{co}(\mathbf{M}) = \{0_{\mathbb{R}^K}\}$ we say that $\mathrm{PoF}_w = \infty$. The weighted PoF, and more generally the PoF for any fairness notion defining a set $\mathbf{F}_\mathrm{w} \cap \mathrm{co}(\mathbf{M})$ that is closed and not reduced to a singleton, is bounded for fractional matchings. This is not always the case for integer matchings. We prove all this in Section 2.6.1.

We can also be interested in an additive difference between the optimum and fair optimum. Interestingly, for most fairness notions (even beyond weighted fairness), using the structure of the maximum matchings being exactly the Pareto optimal matchings, we can obtain a characterization of this additive gap in terms of L_1 distance between $\mathbf{P}$ and $\mathbf{F} \cap \mathrm{co}(\mathbf{M})$.

Proposition 2.12. *Let $\mathbf{F}$ be a set of fair points that is closed and non-empty. Denote $\mathbf{H} := \mathbf{F} \cap \mathrm{co}(\mathbf{M})$ and $d_1(\mathrm{co}(\mathbf{P}), \mathbf{H}) := \inf_{(x,y) \in \mathrm{co}(\mathbf{P}) \times \mathbf{H}} \|x - y\|_1$. It holds,*

$$d_1(\mathrm{co}(\mathbf{P}), \mathbf{H}) = \max_{x \in \mathrm{co}(\mathbf{M})} \|x\|_1 - \max_{y \in \mathbf{H}} \|y\|_1,$$

In addition, the infimum in $d_1(\mathrm{co}(\mathbf{P}), \mathbf{H})$ is always attained.

Proof sketch. The proof is provided in Section 2.9.4. ■

Nevertheless, one might be interested in exactly computing this additive gap or the PoF_w. The following section discusses how to efficiently (on the the graph size) do it.

2.5.3 Maximum size fair matching: A linear programming approach

The exact computation of the PoF_w for a given graph $\mathcal{G}$ and vector w is a polynomial problem on the size of $\mathcal{G}$ and on K. The computation of $\max_{x\in\mathbf{M}}\|x\|_1$ can be done through any maximum matching procedure, such as the Hungarian method [Kuh55]. To compute $\max_{x\in\mathbf{F}_{\mathrm{w}}\cap\mathrm{co}(\mathbf{M})}\|x\|_1$, we present a linear programming formulation.

Let $w\in\mathbb{R}_+^K$ be a vector of weights. w-Fairness imposes that $\forall i,j\in[K]$, $x_i/w_i = x_j/w_j = c$, for c some constant. Having this in mind, any w-fair vector $x\in\mathbb{R}^K$ satisfies $x=(cw_1,cw_2,\ldots,cw_K)$. In order to find a maximum size w-fair matching therefore, it is enough with solving the following linear program (LP),

$$\max\left\{c>0 \mid c\sum_{i\in\Lambda} w_i \le M_\Lambda, \forall\Lambda\subseteq[K]\right\}. \tag{2.1}$$

The optimal value of the previous LP corresponds to,

$$c^* := \min_{\Lambda\subseteq[K]} \frac{M_\Lambda}{\sum_{i\in\Lambda} w_i}$$

and then,

$$\max_{x\in\mathbf{F}_{\mathrm{w}}\cap\mathrm{co}(\mathbf{M})}\|x\|_1 = c^*\sum_{i\in[K]} w_i.$$

Remark 2.3. Notice that under this formulation, computing c^* is exponential on the number of groups. However, our constraints come from an underlying matching problem and thus, it can be computed in **polynomial time** through a linear program having only a polynomial number of constraints, as demonstrated in Section 2.9.3.

From an algorithmic point of view, Equation (2.1) can be interpreted as running Algorithm 6 with rates $(w_i)_{i\in[K]}$ and stopping at the first saturation of the constraints. Alternatively, by being able to compute all M_Λ efficiently, we can give an efficient oracle to the membership of $x\in\mathbf{M}$ in time $\mathcal{O}(2^K)$ using the characterization of $\mathbf{M}$ given in Proposition 2.6.

This result can also be compared to the one in [Chi+19; Ban+23] who consider computationally harder problems with integral matching, and can provide approximations of $\max_{x\in\mathbf{F}_{\mathrm{w}}\cap\mathbf{M}}\|x\|_1$ with an exponential dependency in K.

Once showed how to efficiently compute the PoF_w it remains the question of how large it can be. One major downside of Egalitarian and Demographic Fairness is the fact that they suffer by the disparity in the size of groups or the addition of isolated vertex to the graph. This is not an issue for the opportunity fairness notion which takes into account the whole graph structure. Because of this, we choose to extend the study of the opportunity PoF in the following section in its worst case setting and beyond.

2.6 Opportunity price of fairness

Recall a point $x \in \mathbb{R}_+^K$ is opportunity fair if it verifies

$$\frac{x_i}{M_i} = \frac{x_j}{M_j}, \forall i, j \in [K], \tag{2.2}$$

where M_i is the maximum number of G_i agents that can be ever matched. We denote $\mathbf{F}_\text{O}$ the set of opportunity fair points. We dedicate this section to study the Opportunity PoF in three settings: worst case, beyond the worst case, and random graphs.

2.6.1 Worst-case analysis

Opportunity fairness will achieve different results depending on the dimension of the problem, i.e., the number of groups K. Moreover, we start by showing that considering fractional matchings is crucial for the worst case analysis as the Opportunity-PoF (PoF_O) is known to remain always bounded (Proposition 2.12) while for integer matchings we can construct graphs where the only fair matching is the empty one, yielding an unbounded PoF_O.

Proposition 2.13. *For any constant $M > 0$, there exists a graph $\mathcal{G}$ such that the only element in $\mathbf{F}_\text{O} \cap \mathbf{M}(\mathcal{G})$ is the null vector $\mathbf{0}_K$, while there exist matchings $\mu \in \mathcal{M}(\mathcal{G})$ with $X(\mu) \geq M$.*

Proof. Take $K = 2$ and let M_1 and M_2 be two different prime numbers such that $M_1+M_2-1 \geq M$. Consider a graph $\mathcal{G}$ with $M_1 - 1$ jobs connected to all agents in G_1 but no agent in G_2, $M_2 - 1$ jobs connected to all agents in G_2 but no agent in G_1, and one additional job connected with everybody, as in Figure 2.8 left. The maximum size matching has size $M_1 + M_2 - 1$. A point $x \in \mathbf{M}$ has both x_1 and x_2 in $\mathbb{N}$ yet, as M_1 and M_2 are primes and different, for both coordinates to be integer and satisfy opportunity fairness, either $x = (0, 0)$ or $x = (M_1, M_2)$. Since $(M_1, M_2) \notin \mathbf{M}$, it holds $\mathbf{F}_\text{O} \cap \mathbf{M} = (0, 0)$. Figure 2.8 right illustrates these points. ■

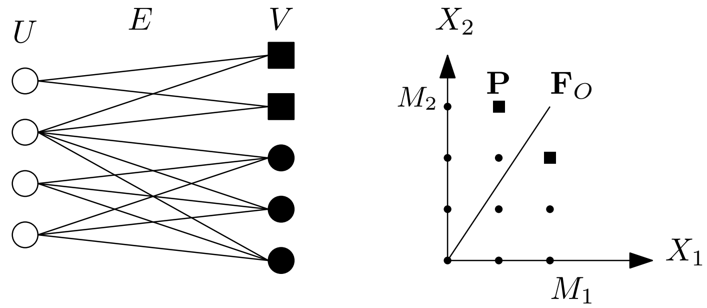


Fig. 2.8: Integral PoF_O

Proposition 2.13 implies the following result.

Corollary 2.14. *For integer matchings it holds* $\sup_{\mathcal{G}\ graph} \text{PoF}_\text{O}(\mathcal{G}) = \infty$.

The issue exposed in Corollary 2.14 is solved when working with fractional matchings.

Theorem 2.15. *For fractional matchings, it holds that* $\sup_{\mathcal{G}\ graph} \text{PoF}_\text{O}(\mathcal{G}) = K - 1$.

Proof. First let us show the upper bound. Suppose $\mathbf{F}_\text{O} \cap \text{co}(\mathbf{P}) = \varnothing$, otherwise $\text{PoF}_\text{O} = 1 \leq K - 1$. The PoF_O is given by,

$$\text{PoF}_\text{O} = \frac{M_{[K]}}{M_{\Lambda^*}} \cdot \frac{\sum_{i\in\Lambda^*} M_i}{\sum_{i\in[K]} M_i},$$

where Λ^* achieves the solution of Equation (2.1). Since $\mathbf{F}_\text{O} \cap \text{co}(\mathbf{P}) = \varnothing$, $\Lambda^* \neq [K]$. Setting $\lambda := \text{argmax}(M_i : i \in \Lambda^*)$, notice that $M_{\Lambda^*} \geq M_\lambda \geq \frac{1}{|\Lambda^*|} \sum_{i\in\Lambda^*} M_i$. Plugging this bound in the PoF_O definition, it follows,

$$\text{PoF}_\text{O} \leq |\Lambda^*| \cdot \frac{M_{[K]}}{\sum_{i\in\Lambda^*} M_i} \cdot \frac{\sum_{i\in\Lambda^*} M_i}{\sum_{i\in[K]} M_i} \leq |\Lambda^*| \leq K - 1,$$

where we have used that $M_{[K]} \leq \sum_{i\in[K]} M_i$ and $|\Lambda^*| \leq K - 1$ as $\Lambda^* \neq [K]$. To show the tightness of the bound, let M and N be two values. Consider next a graph where a group is independently connected to M jobs and $K-1$ groups connected to N jobs at the detriment of the other groups (see Figure 2.9 for an example when $K = 3$). It holds,

$$\text{PoF}_\text{O} = \frac{M + N}{\frac{M}{K-1} + N} \xrightarrow{M\to\infty} K - 1.$$

■

Theorem 2.15 implies the following remarkable result.

Corollary 2.16. *For any graph $\mathcal{G}$ with $K = 2$,* $\text{PoF}_\text{O}(\mathcal{G}) = 1$.

This can also be derived from a simple geometrical argument as it can be proved that the Pareto frontier and the line $[(0,0),(M_1,M_2)]$ always intersect. Unfortunately, as shown in Theorem 2.15 and illustrated by Figure 2.9 (μ_O^* denotes an opportunity fair maximum size matching), the property does not necessarily hold for larger values of K. Remark, however, that the worst case example relies on disparate maximum number of matchable agents among different groups. It seems intuitive that whenever the values $(M_i)_{i\in[K]}$ are restricted to be equal, we should rule out such a worst case example. This motivates a beyond worst case study for PoF_O under additional structure either on the parameters or on the geometry.

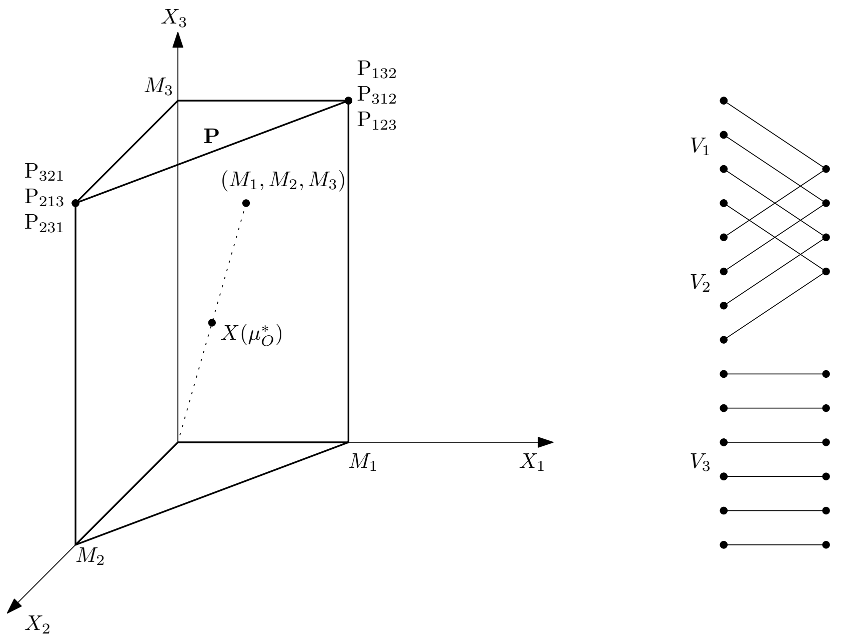


Fig. 2.9: Toblerone graph

2.6.2 Beyond the worst case analysis

We start the analysis by considering upper bounds which depend on the relative opportunity levels of the groups $(M_i)_{i\in[K]}$.

Proposition 2.17. *The* PoF_O *is never greater than*

$$\frac{\max_{i\in[K]} M_i}{2\min_{i\in[K]} M_i} + \frac{K}{4}\left(\frac{\max_{i\in[K]} M_i}{\min_{i\in[K]} M_i}\right)^2 + \frac{1}{4K}\mathbb{1}_{K odd}. \tag{2.3}$$

Moreover, whenever $M_i = M_j$ for all $i, j \in [K]$, the bound is tight

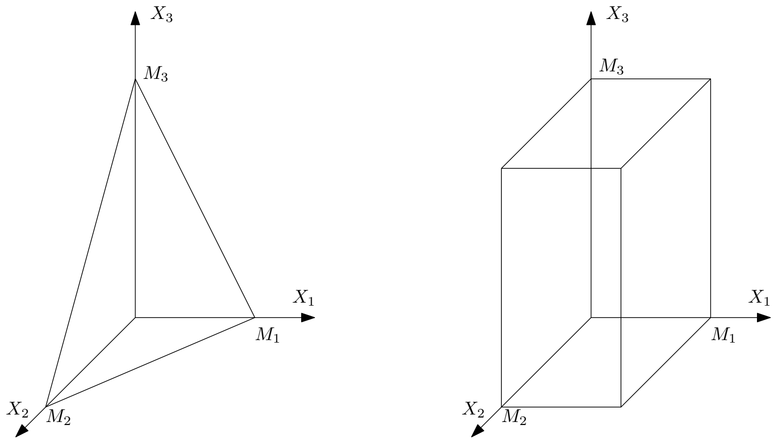


Fig. 2.10: Set $\mathrm{co}(\mathbf{M})$ shape. Left pyramid $\rho = 1/K$, right hyper-rectangle $\rho = 1$.

Proof sketch. Consider the maximal opportunity-fair matching, the facet of $\mathrm{co}(\mathbf{M})$ where it lies, and the set of indices corresponding to the facet. We express the PoF_O as a function of the cardinal of this subset and differentiate the obtained expression to find its maximum. The maximum being reached when the subset contains half of the groups, the upper bound is derived. The full proof is provided in Section 2.9.4. ■

The previous bound concerns the size of the polytope $\mathbf{M}$. Alternatively, we study an upper-bound related to the geometry of $\mathbf{M}$ captured by the parameter $\rho := M_{[K]} / \sum_{i \in [K]} M_i$, that is, the ratio between the size of a maximum size matching and the *ideal optimum*, the Utopian matching where each group gets as many matched agents as if they were the only group on the graph. It is direct that $\rho \in [1/K, 1]$ and that the extreme cases correspond to $\mathbf{M}$ be shaped as an inverted pyramid ($\rho = 1/K$, Figure 2.10 left) with the origin as the top of the pyramid and $\mathbf{P}$ as the pyramid basis, and as a hyper-rectangle ($\rho = 1$, Figure 2.10 right). In both cases, $\mathbf{F}_\mathrm{O}$ intersects $\mathbf{P}$ independent on the value of K, yielding $\mathrm{PoF}_\mathrm{O} = 1$. The following proposition gives a quantitative result for intermediary cases where the geometry of $\mathbf{M}$ differs from a perfect inverted pyramid or a hyper-rectangle.

Proposition 2.18. *Suppose that for all* $i \in [K]$, $M_i = M > 0$, *and let* $\rho = \frac{M_{[K]}}{\sum_{i \in [K]} M_i} = \frac{M_{[K]}}{KM}$ *be the rate between the size of a maximum size of matching and the size of the Utopian matching. If* $\rho \in [\frac{1}{K}, \frac{1}{K-1}]$, *then* $\mathrm{PoF}_\mathrm{O} = 1$. *Otherwise, for* $\rho \in [\frac{1}{K-1}, 1]$, *we have that*

$$\sup_{\mathcal{G} \text{ s.t. } M_i = M,\, \rho \text{ is fixed}} \mathrm{PoF}_\mathrm{O}(\mathcal{G}) = \rho \max\left(\frac{K - \lfloor K\rho \rfloor + 1}{K\rho - \lfloor K\rho \rfloor + 1}, K - \lfloor K\rho \rfloor\right) \leq \rho((1-\rho)K + 1)$$

Proof sketch. The bound is showed by determining a lower bound on the quantity c^* as a function of ρ. The tight example is constructed from a continuous parametrization in terms of competition between groups from Figure 2.9. The proof is provided in Section 2.9.4. ■

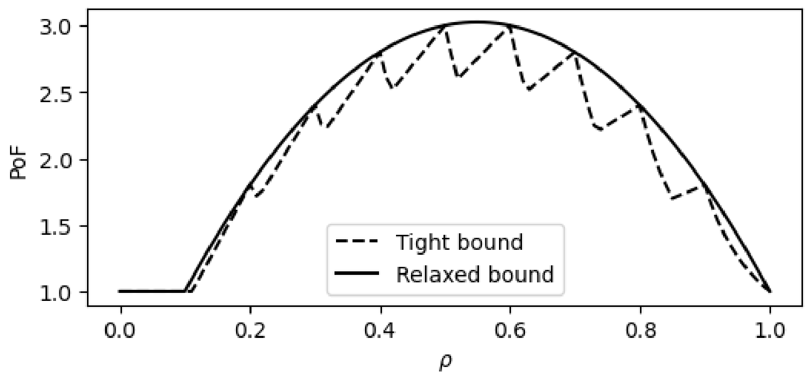


Fig. 2.11: Plot of maximum PoF_O and relaxed bound as a function of ρ for $K = 10$ and $M_i = M$

This tight bound is surprisingly multimodal as it can be seen in Figure 2.11. The tight bound is recovered when all values M_i are equal (Proposition 2.17) by simply maximizing the above bound over ρ. Notice that a slight modification of the proof generalizes the bound by taking into account the quantity $\max M_i / \min M_i$. However, the range for which the bound remains valid is smaller in $\max M_i / \min M_i$ than for Proposition 2.17.

Rather than deriving refined inequalities we prefer to find sufficient conditions to identify graphs with a PoF_O equal to 1. To do this, given $\sigma \in \Sigma([K])$, we denote $M^\sigma := \left(\frac{M_{\sigma([\ell])} - M_{\sigma([\ell-1])}}{M_{\sigma(\ell)}}\right)_{\ell \in [K]}$ (recall $M_{\sigma([\ell])}$ corresponds to the size of a maximum size matching when only considering the first ℓ groups on $\sigma([K])$).

Proposition 2.19. *Whenever all sequences* $\{M^\sigma, \sigma \in \Sigma([K])\}$ *are non-increasing, it holds* $\text{PoF}_\text{O} = 1$.

Proof sketch. From the sequence M^σ being non-increasing, we can show that the function $\Lambda \subset [K] \mapsto M_\Lambda / \sum_{i \in \Lambda} M_i$ is non-increasing in Λ. This implies that the set which solves Equation (2.1) is $[K]$, obtaining a PoF_O of 1. The proof is provided in Section 2.9.4. ■

Proposition 2.19 gives a sufficient condition for a graph to have a PoF_O equal to 1. Notice that for $K = 2$, the sequences M^σ are always non-increasing, recovering Corollary 2.16. Remark as well that, although sufficient, the condition is not necessary to obtain an opportunity PoF equal to 1. Indeed, consider a graph $\mathcal{G}$ with four groups, one agent per group, two jobs, such that G_1 and G_2 are connected to one of them, and G_3 and G_4 are connected to the other one. It holds $\text{PoF}_\text{O}(\mathcal{G}) = 1$ and yet $\mathcal{G}$ has increasing sequences. Nonetheless, the monotonicity condition is useful to study, for instance, complete graphs.

Proposition 2.20. *Let* $\mathcal{G}$ *be a complete graph. Then,* $\text{PoF}_\text{O}(\mathcal{G}) = 1$.

Proof sketch. To obtain this result, we show that M^σ is non-increasing and apply the previous proposition. The proof is provided in Section 2.9.4. ■

Real-life applications can rarely be modeled by worst case settings, even under extra structure as studied in this section. This motivates us to go beyond these approaches and to consider a stochastic setting with random graphs, as presented in the following section.

2.6.3 Stochastic model

We dedicate this sub-section to study the Opportunity Price of Fairness for Erdös-Rényi bipartite graphs and to determine regimes where a PoF_O equal to 1 is asymptotically[3] achievable. We define a bipartite Erdös-Rényi graph with one side V of size n and another side U of size $\lfloor \beta n \rfloor$, for $\beta \in (0, 1)$ fixed and known, such that,

- for any $v \in V$, v belongs to the group G_i, for $i \in [K]$, with probability $\alpha_i \geq 0$, such that $\sum_{i\in[K]} \alpha_i = 1$. Each node in V is assigned to one and only one group,

- each node $v \in G_i$ has probability $p_i \in (0, 1)$, known and fixed, to be connected to each node in U, independently.

We denote $\mathbb{G}_{n,\beta,\vec{\alpha},\vec{p}} = (U, (G_i)_{i\in[K]}, E)$ the random bipartite graph just described, where $\vec{\alpha} := (\alpha_i)_{i\in[K]}$ and $\vec{p} := (p_i)_{i\in[K]}$. For simplicity, we drop the indices and just denote $\mathbb{G}$ to the Erdös-Rényi bipartite graph. We aim at characterizing the regimes for which $\mathrm{PoF}_\mathrm{O}(\mathbb{G}) = 1$. In order to do it, we recall in Section 2.9.4 two results from the literature of random graphs [FK16] which characterize, respectively, the edge probability for which random graphs become sparse, and for which random graphs become dense enough to ensure the existence of perfect matchings. We show that in both cases, with high probability, $\mathrm{PoF}_\mathrm{O}(\mathbb{G})$ is equal to 1.

Proposition 2.21. *Consider an Erdös-Rényi bipartite graph* $\mathbb{G}$ *such that* $\max_{i\in[K]} p_i \leq \frac{1}{\omega n^{3/2}}$ *for* $\omega = \omega(n) \to \infty$ *arbitrarily slow as* $n \to \infty$*. Then, with high probability,* $\mathrm{PoF}_\mathrm{O}(\mathbb{G}) = 1$.

Proof sketch. The bipartite graph generated by $\max_{i\in[K]} p_i$ can be shown to be sparse using properties from random graphs. This implies that with high probability no two agents have the possibility of being matched to the same jobs. The proposition with respect to different p_i is implied by stochastic domination of the random graph with $\max_{i\in[K]} p_i$. The proof is provided in Section 2.9.4. ■

[3] Recall that classical results in random graphs are stated as the number of vertices grows to $+\infty$.

Proposition 2.22. *Consider an Erdös-Rényi bipartite graph* $\mathbb{G}$ *such that* $p_i \geq \log^2(n)/n$ *for any* $i \in [K]$. *Then* $\lim_{n\to\infty} \mathbb{P}(\mathrm{PoF}_O(\mathbb{G}) = 1) = 1$.

Proof sketch. The proof goes as follows: First, we show $|G_i|$ concentrates around $\alpha_i n$. Second, we show that $M_i = \min(\alpha_i n, n)$. Third, we prove that for any $\sigma \in \Sigma([K])$, the sequence M^σ is non-increasing by running Algorithm 5. We can then apply Proposition 2.19. See Section 2.9.4 for the full proof. ∎

2.7 Extension to general matroids

In Section 2.3.1 we used the fact that the set $\mathbf{M}$ defines a polymatroid, and used polymatroid properties to prove Theorem 2.1. In fact, many of our results come from the very particular shape of $\mathbf{M}$, and this shape is not specific to bipartite matching problems but to a broad class of polymatroids. In this section, we define matroids, polymatroids and we explain how we can extend our fairness analysis to a broad class a combinatorial problems beyond matching.

2.7.1 Matroids and polymatroids

We start by giving the definition of a matroid.

Definition 2.6. A pair $(E, \mathcal{M})$, with E a finite set and $\mathcal{M} \subseteq 2^E$ a family of subsets of E that contains the empty set $\varnothing$, is called a **matroid** if the following properties hold,

1. $\forall A \in \mathcal{M}$ and $A' \subseteq A$, $A' \in \mathcal{M}$,
2. $\forall A, B \in \mathcal{M}$ such that $|A| < |B|$, $\exists e \in B \setminus A$ such that $A \cup \{e\} \in \mathcal{M}$.

E is called the **ground set** and $\mathcal{M}$ the family of **independent sets**.

Some classes of matroids are defined by properties that give them specific structures, here we provide some examples.

- **Transversal matroid**. Let $G = (U, V, E)$ be a bipartite graph. For a matching $\mu \subseteq E$ we denote $\mu(E) := \{v \in V \mid \exists u \in U, (u, v) \in \mu\}$. If $\mathcal{M} := \{\mu(E), \forall \mu \subseteq E \text{ matching}\}$, then the pair $(U, \mathcal{M})$ is a transversal matroid.
- **Graphic matroid**. Given $G = (V, E)$ a graph, the graphic matroid is the pair $(E, \mathcal{M})$ such that $\mathcal{M} := \{A \subseteq E \mid A \text{ is acyclic}\}$.

- **Uniform matroid**. Given E a finite set and $r \in \mathbb{N}$, the r-uniform matroid is the pair $(E, \mathcal{M})$ such that $\mathcal{M} := \{A \subseteq E \mid |A| \leq r\}$.

The bipartite matching problem can therefore be modeled by a transversal matroid (notice that the notation we introduced here are consistent with those used in the rest of the chapter). The communication network problem, the goal of which is to maintain a communication network between different locations in a city, can be modeled as a graphic matroid. The state can request bids from various types of companies (public and private, for example) to externalize the task. If V corresponds to the city locations and E to the possible communication connections between them, finding an acyclic subset of edges ensures to minimize the material costs.

Definition 2.7. Given a matroid $(E, \mathcal{M})$, a **basis** is a maximal independent set, i.e., any $A \in \mathcal{M}$ such that, for any $e \in E \setminus A$, $A \cup \{e\} \notin \mathcal{M}$.

By the augmentation property, all bases of $(E, \mathcal{M})$ have the same size and are therefore maximum. This number is called the **rank** of the matroid and it is denoted $\operatorname{rank}(\mathcal{M})$. For a fixed subset of $A \subseteq E$, we can define the **sub-matroid** $(A, \mathcal{M}_A)$ by $\mathcal{M}_A := \{A' \subseteq A \mid A' \in \mathcal{M}\}$. In particular, this allows to extend $\operatorname{rank}(\cdot)$ to all sets of 2^E.

Definition 2.8. Given $f : 2^E \to \mathbb{R}_+$ a submodular function (i.e., $\forall A, B \in E, f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$), the **polymatroid** associated to f is the polytope defined as,

$$Q_f := \left\{ x \in \mathbb{R}_+^E \mid \sum_{e \in E'} x_e \leq f(E'), \forall E' \subseteq E \right\}.$$

The integer polytope $Q_f \cap \mathbb{N}^E$ is called a **discrete polymatroid**. Equivalently, a polytope $Q \subseteq \mathbb{R}_+^E$ is a polymatroid if it verifies the following properties,

1. For any $x \in \mathbb{R}_+^E$ and $y \in Q$ such that $x \leq y$ (component wise), $x \in Q$,
2. If $x, y \in Q$ with $\|x\|_1 < \|y\|_1$, there exists $z \in Q$ such that $\forall e \in E$, $x_e \leq z_e \leq \max\{x_e, y_e\}$ with at least one strict equality between x and z.

We will show that when the ground set E of a matroid is divided into groups, and we project it on $\mathbb{R}^K$ with the mapping X as we did before, we obtain a polymatroid.

2.7.2 Colored matroids

We now introduce colored matroids [Chi+19].

Definition 2.9. Let $(E, \mathcal{M})$ be a matroid, $K \in \mathbb{N}$ be a fixed integer, and $(E_i, i \in [K])$ (called **groups** or **colors**) be a partition of E. We call $((E_i)_{i\in[K]}, \mathcal{M})$ a **colored matroid** (with K colors).

Given a colored matroid, we define the mapping

$$\begin{aligned} X : \mathcal{M} &\longrightarrow \mathbb{R}_+^K \\ A &\longrightarrow (X_1(A), X_2(A), ..., X_K(A)) := (|A \cap E_1|, |A \cap E_2|, ..., |A \cap E_K|), \end{aligned}$$

that is, X maps independent sets into vectors of dimension K counting the number of elements of each color.

Theorem 2.23. *Let $((E_i)_{i\in[K]}, \mathcal{M})$ be a colored matroid. The set of allocations $X(\mathcal{M})$ verifies,*

1. *For any $x \in \mathbb{R}_+^K$ and $y \in X(\mathcal{M})$ such that $x \le y$ (component wise), $x \in X(\mathcal{M})$.*
2. *For any $x, y \in X(\mathcal{M})$, with $\|x\|_1 < \|y\|_1$, there exists $i \in [K]$ such that $x_i < y_i$ and $x + \vec{e}_i \in X(\mathcal{M})$.*

This theorem is the generalization of Theorem 2.5 to general matroids. The proof is analogous to the proof of Theorem 2.5 and is therefore omitted. Theorem 2.23 implies two facts: (i) $X(\mathcal{M})$ is a discrete polymatroid and (ii) for any two $x, y \in \mathcal{M}$ with $|x| > |y|$, there always exists an agent from a less represented group in x that can be added to y.

From Section 2.3.1 we immediately obtain the following result:

Corollary 2.24. *Let $((E_i)_{i\in[K]}, \mathcal{M})$ be a colored matroid. It follows, that $\mathrm{co}(X(\mathcal{M}))$ is a polymatroid which can be represented as*

$$\mathrm{co}(X(\mathcal{M})) := \{x \in \mathbb{R}_+^K \mid \forall \Lambda \subseteq [K], \textstyle\sum_{i\in\Lambda} x_i \le \mathrm{rank}(\Lambda)\},$$

where $\mathrm{rank}(\Lambda) := \mathrm{rank}(\cup_{i\in\Lambda} E_i)$ is the size of a basis on the sub-matroid generated by the elements with labels in Λ.

2.7.3 Extension of our results

We will not state and prove again all the results from this chapter that extend to colored matroids, but only go over which results extend and for which classes of matroids. The results that are extensions of numbered statements are given with the reference of the original statement.

True for:	Statement:	Extends:
Any (colored) matroid	The Pareto front of $\text{co}(X(\mathcal{M}))$ is equal to the set $\{x \in X(\mathcal{M}))\|\|x\|_1 = \text{rank}(\mathcal{M}\}$	Corollary 2.2
	The Shapley allocation exists, is on the Pareto front, and is the barycenter of the P_σ, $\sigma \in \Sigma([K])$	Prop. 2.8
	$\text{PoF}_\text{O} \leq K - 1$, and if $K = 2$ then $\text{PoF}_\text{O} = 1$	Theorem 2.15
	$\text{PoF}_\text{O} \leq \frac{1}{2} \cdot \frac{\max_{i \in [K]} \text{rank}(i)}{\min_{i \in [K]} \text{rank}(i)} + \frac{K}{4} \cdot \left(\frac{\max_{i \in [K]} \text{rank}(i)}{\min_{i \in [K]} \text{rank}(i)} \right)^2$ $+ \frac{1}{4K} \cdot \mathbb{1}_{K \text{odd}}$	Prop. 2.17
	$\text{PoF}_\text{O} \leq \rho \max \left(\frac{K - \lfloor k\rho \rfloor + 1}{K\rho - \lfloor K\rho \rfloor + 1}, K - \lfloor K\rho \rfloor \right)$ $\leq \rho((1-\rho)K + 1)$	Prop. 2.18
Uniform, transversal & graphic matroids	Algorithm 5 is polynomial in $\|E\|$	Remark 2.1
	For any weight vector w, a w-fair allocation can be found in polynomial time w.r.t. $\|E\|$ and K	Remark 2.3
Transversal & graphic	All the bounds on PoF_O given above are tight	
Uniform	$\text{PoF}_\text{O} = 1$ for any K	

2.8 Discussion

While we have focused on maximizing the cardinality of a bipartite matching with fairness constraints on the agents side, it is natural to consider various generalizations of this problem. In terms of utility objectives, it could be reasonable to consider weighted matching. If the edges weights are based on agents groups, this simply leads to a skewed polytope, and some results can be obtained (see Section 2.9.2). However, whenever the weights depend on the agents or on the edges, it is not clear whether similar general results can be obtained, as anonymity between agents and the augmentation property, which is crucial to the polymatroid structure, are both lost.

Fairness requirements can also be modified to consider possible discrimination on both sides of the bipartite matching by assigning types to jobs, and seeking matchings such that the number of matched pairs between groups and types are equal for all possible pairs. This two-sided fairness unfortunately cannot be directly encoded into a larger bipartite graph, as it requires to take into account that jobs can only be matched once.

Finally, a third possible generalization is on the structure of the problem, where general graphs are considered instead of bipartite graphs. Clearly, it also cannot be encoded into the setting studied in this chapter for the same previously mentioned reasons.

2.9 Appendices

2.9.1 Notation

Table 2.1 provides a summary of the notation used throughout the chapter.

Tab. 2.1: Notation for Chapter 2

Graph and groups:	
$\mathcal{G}$	Graph
U	Jobs
V	Agents
E	Edges
G_i	Group i
K	Number of groups
$[K]$	List of all groups ($= \{0, \ldots, K\}$)
Λ	Subset of groups ($\subseteq [K]$)
Matchings:	
μ	Matching
$\mathcal{M}$	Feasible matchings
$\mathcal{P}$	Maximum matchings
X	Mapping from $\mathcal{M}$ to $\mathbb{R}^K$
$\mathbf{M}$	Realizable points ($= X(\mathcal{M})$)
$\mathbf{P}$	Maximum points and Pareto front ($= X(\mathcal{P})$)
$\mathrm{co}(\mathbf{M}), \mathrm{co}(\mathbf{P})$	Convex hulls of $\mathbf{M}, \mathbf{P}$
$>_{\mathcal{L}(\sigma)}$	Lexicographic order along permutation σ
Permutations and extremal points:	
σ	Permutation of $[K]$
$\Sigma([K])$	Set of all permutations
$\mathcal{P}_\sigma$	Lexicographic maximum matching for permutation σ
P_σ	Extremal point of $\mathbf{P}$ corresponding to σ ($= X(\mathcal{P}_\sigma)$)
M_Λ	Size of a maximum matching when restricting V to $\bigcup_{i \in \Lambda} G_i$
Fairness:	
w-fair	Fair w.r.t. Definition 2.4 with vector $w \in \mathbb{R}_+^K$
$\mathbf{F}_{\mathrm{w}}$	Set of w-fair matchings
PoF_w	Price of fairness, for the w-fairness rule

2.9.2 Generalization of projection properties for weighted fairness and weighted leximin

Let us start this section by discussing a third possibility to define a 'fairest' optimal matching. Let $w = (w_1, \ldots, w_K)$, with $w_i > 0, \forall i \in [K]$, be a vector of positive weights, corresponding to the entitlement of each group. Denote by $\frac{1}{w} = (\frac{1}{w_i})_{i \in [K]}$.

Definition 2.10. We define $\mathbf{P}^w$ as the set of weighted maximal matchings, that is,

$$\mathbf{P}^w := \underset{x \in \mathbf{M}}{\operatorname{argmax}} \left(\frac{1}{w}\right)^\top x. \tag{2.4}$$

Remark the set of maximal matchings corresponds to $\mathbf{P} = \mathbf{P}^{\mathbf{1}}$.

We denote by $\Sigma^w([K]) \subseteq \Sigma([K])$ the set of permutations which are consistent with w, namely for any $\sigma \in \Sigma^w([K])$ and for any $(i,j) \in [K]^2$, $\sigma^{-1}(i) < \sigma^{-1}(j)$ if and only if $1/w_i \geq 1/w_j$. This ensures that whenever i appears before than j in σ, the weight associated to i is at least as high as the one of j. The number of permutations which are consistent with w depends on the number of unique weights values, and the number of groups which have this same unique value. More specifically,

$$|\Sigma^w([K])| = \prod_{u \in \{w_i | i \in [K]\}} |\{i \mid w_i = u\}|.$$

In particular, if all w_i are distinct, $\Sigma^w([K])$ is reduced to a singleton, while for equal weights $\Sigma^w([K]) = \Sigma([K])$.

We have the following property:

Proposition 2.25. *The set of weighted maximal matchings is equal to the convex hull of the lexicographic maximal matchings for all w-consistent permutations. More precisely:*

$$\mathbf{P}^w = \mathrm{co}(\{\mathrm{P}_\sigma \mid \sigma \in \Sigma^w([K])\}). \tag{2.5}$$

Proof. First, by Proposition 2.3 taking $\lambda_i = 1/w_i$, it is immediate that all points P_σ which are optimal for the weighted objective are exactly those with $\sigma \in \Sigma^w([K])$. Thus, the linearity of the mapping X implies that $\mathrm{co}(\{\mathrm{P}_\sigma \mid \sigma \in \Sigma^w([K])\}) \subset \mathbf{P}^w$.

Conversely, as weights are positive, $\mathbf{P}^w \subseteq \mathbf{P}$. Since $\mathbf{P}$ corresponds to the convex hull of all lexicographic maximum size points P_σ, but those P_σ with $\sigma \notin \Sigma^w([K])$ are sub-optimal, their coefficient on the convex combination must be necessarily zero, implying that $\mathbf{P} = \mathrm{co}(\{\mathrm{P}_\sigma \mid \sigma \in \Sigma^w([K])\})$. ∎

As points in $\mathbf{P}^w$ are also maximal in the sense of cardinality, this provides another possible notion of fairness which induces no optimality loss. The less unique values of w, the smaller the dimension of $\mathbf{P}^w$ is. Let us see how some of our results translate to these new weighted optimal maximizers. We denote by x^w the re-scaled point $(x_1/w_1, \ldots, x_K/w_K)$, and for the definite positive matrix $D_w = Diag(1/w)$, the associated norm $\|x\|_w = x^\top D_w^\top D_w x$. Let $\mathrm{P}^{\mathrm{w}}_{\mathrm{leximin}}$ be the w-weighted leximin maximum point. By considering the scaled fractional polytope $\mathbf{M}^w := \{x^w \mid x \in \mathbf{M}\}$, it is clear that the augmentation property of Proposition 2.5 still applies, hence the set of maximal points $\mathbf{P}^w$ is exactly the Pareto frontier in $\mathbf{M}^w$. Because the weighted leximin optimal matching belongs to the Pareto frontier of $\mathbf{M}^w$, by definition, it is also maximal and therefore, it belongs to $\mathbf{P}^w$. We are ready to state the generalization of Proposition 2.10.

Proposition 2.26. *The point $x = \mathrm{P}^{\mathrm{w}}_{\mathrm{leximin}} \in \mathbf{P}^w$ is the unique point that minimizes each of the following objectives among all points in $\mathbf{P}^w$:*

1. $\|x\|_w = x^\top D_w^\top D_w x$.
2. $\mathbf{Var}(x) := \frac{1}{K}\sum_{i\in[K]}(\frac{x_i}{w_i} - \frac{1}{K}\sum_{\ell\in[K]}\frac{x_\ell}{w_\ell})^2$

Proof. The proof is identical as in the main body: (1) It can be shown that for any point $z \in \mathbf{P}^w$ it holds $\|z\|_w \geq \|x\|_w$, (2) all points in $\mathbf{P}^w$ sum up the same as x^w, and (3) $\|x\|_w^2 = \sum_{i\in K}(x_i/w_i)^2$. The second point follows similarly. ■

The projection property can also be extended, with the correspondent proof by scaling the vectors by w.

Proposition 2.27. *The $\|\cdot\|_w$-projection of $\mathbf{P}^{\mathbf{w}}$ onto $\mathbf{F}_w$ is the fair optimum $y = t^* w$ where $t^* = \max\{t \geq 0 \mid tw \in \mathrm{co}(\mathbf{M})\}$, and the $\|\cdot\|_w$-projection of $\mathbf{F}_w$ onto $\mathbf{P}^{\mathbf{w}}$ is $\mathrm{P}^{\mathrm{w}}_{\mathrm{leximin}}$.*

We observe that $\mathbf{P}^w$ and $\mathbf{F}_w$, when projected one on another with $\|\cdot\|_w$, it reduces to either the fair optimum or the optimum fair. The main difference compared to $w = \mathbf{1}$, is that a different distance is used as we scale the polytope by w.

2.9.3 Computational remarks

In this section, we show that several computational tasks involving fractional matching can be performed in polynomial time, using a linear programming approach. First, recall that X is a linear mapping that maps fractional matchings $\mu \in \mathrm{co}(\mathcal{M})$ to the corresponding points $X(\mu) \in \mathrm{co}(\mathbf{M})$. A natural question is: given $x \in \mathbb{R}_+^K$, can we decide in polynomial time if $x \in \mathrm{co}(\mathbf{M})$, and if so can we build a fractional matching μ such that $X(\mu) = x$?

Proposition 2.28. *Given $x \in \mathbb{R}_+^K$, there exists a fractional matching $\mu \in \mathrm{co}(\mathcal{M})$ such that $X(\mu) = x$ if and only if the following linear program has a feasible solution with value $\|x\|_1$,*

$$\begin{aligned}
\textbf{\textit{maximize}} \quad & \sum_{u\in U}\sum_{v\in V} \mu_{u,v} & \\
\textbf{\textit{such that}} \quad & 0 \leq \mu_{u,v} \leq \mathbb{1}_{\{(u,v)\in E\}} & (\forall u \in U, v \in V)\\
& \sum_{u\in U} \mu_{u,v} \leq 1 & (\forall v \in V)\\
& \sum_{v\in V} \mu_{u,v} \leq 1 & (\forall u \in U)\\
& \sum_{v\in G_i}\sum_{u\in U} \mu_{u,v} = x_i & (\forall i \in [K])
\end{aligned}$$

Proof. The proof follows from the definition of X. ■

Next, in various settings (to compute the leximin optimal, or a maximal fair matching), we want to start from a point $x \in \mathrm{co}(\mathbf{M})$, and increase continuously each x_i at a rate of w_i, until some constraint $\sum_{i\in\Lambda} x_i \leq M_\Lambda$ is saturated.

Proposition 2.29. *Given $x \in \mathrm{co}(\mathbf{M})$ and a vector of weights $w \in \mathbb{R}_+^k$, we define*

$$c^* = \max\{c \geq 0 \,|\, x + c \cdot w \in \mathrm{co}(\mathbf{M}\}) = \max\left\{c \geq 0 \,|\, \forall \Lambda \subset [K], \sum_{i\in\Lambda}(x_i + c\cdot w_i) \leq M_\Lambda\right\}.$$

Then, c^* is equal to the optimal solution of the following linear program.

$$\begin{aligned}
\textbf{\textit{maximize}} \quad & c & \\
\textbf{\textit{such that}} \quad & c \geq 0 & \\
& 0 \leq \mu_{u,v} \leq \mathbb{1}_{\{(u,v)\in E\}} & (\forall u \in U, v \in V)\\
& \sum_{u\in U} \mu_{u,v} \leq 1 & (\forall v \in V)\\
& \sum_{v\in V} \mu_{u,v} \leq 1 & (\forall u \in U)\\
& x_i + c\cdot w_i = \sum_{v\in G_i}\sum_{u\in U} \mu_{u,v} & (\forall i \in [K])
\end{aligned}$$

Proof. Given a feasible solution c to the linear program, the variables $\mu_{u,v}$ provide a fractional matching proving that $x + c \cdot w \in \mathrm{co}(\mathbf{M})$. Conversely, the optimal c^* has a corresponding fractional matching, which yields a feasible solution to the linear program. ■

Finally, when increasing each coordinate x_i at a rate of w_i, we sometimes need to compute the sets $\Lambda \subseteq [K]$ for which the constraint $\sum_{i\in\Lambda}(x_i + c\cdot w_i) \leq M_\Lambda$ is saturated (holds with equality).

Importantly, using [BCT85, Lemma 2.2] the set of Λ's for which the constraint is tight is closed under union and intersection. That implies that there exists a maximal (inclusion-wise) set Λ^* which saturates the constraint.

To compute Λ^*, we start by computing the optimal solution to the dual linear of Proposition 2.29. Each constraint $x_i + c \cdot w_i \leq \sum_{v \in G_i} \sum_{u \in U} \mu_{u,v}$ has a dual variable which, by complementary slackness, will be positive only if the constraint is tight. Finally, we define Λ^* as the set of groups for which the dual variable is positive. By construction, we have that $\sum_{i \in \Lambda^*} x_i = M_{\Lambda^*}$.

2.9.4 Omitted proofs

In this section, we present the omitted proofs of the chapter.

Proof of Proposition 2.3

Without loss of generality take $\sigma = \mathrm{I}_K$ the identity permutation and $x = \mathrm{P}_\sigma$. To prove property (1) it is enough to show that for any $i \in [K]$, $x_1 + x_2 + \cdots + x_i = M_{[i]}$. Take $i \in [K]$, define $y = (x_1, x_2, \ldots, x_i, 0, \ldots, 0)$, and $z = (z_1, \ldots, z_i, 0, \ldots, 0)$ such that $\|z\|_1 = M_{[i]}$. By definition of $M_{[i]}$, it holds $\|y\|_1 \leq \|z\|_1$. If the inequality is strict, the augmentation property of Theorem 2.1 contradicts the lexicographic optimality of x, hence proves point (1). To prove (2), take $y \in \mathbf{M}$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq \lambda_{K+1} = 0$, and write

$$\begin{aligned} \lambda_1 x_1 + \cdots + \lambda_i x_K &= (\lambda_1 - \lambda_2) x_1 + \cdots + (\lambda_K - \lambda_{K+1})(x_1 + \cdots + x_K) \\ &\geq (\lambda_1 - \lambda_2) y_1 + \cdots + (\lambda_K - \lambda_{K+1})(y_1 + \cdots + y_K) \\ &= \lambda_1 y_1 + \cdots + \lambda_K y_K, \end{aligned}$$

where the inequality is a (positive) linear combination of inequalities from property (1).

Proof of Proposition 2.9

For convenience, define $x = \mathrm{P}_{\text{leximin}}$. To prove (1), assume for the sake of contradiction that there exists $t \geq 0$ and $y \in \mathbf{M}$ such that $\sum_{i \in [K]} \min(t, x_i) < \sum_{i \in [K]} \min(t, y_i)$. We define two points $\tilde{x}, \tilde{y} \in \mathbf{M}$, where $\tilde{x}_i = \min(t, x_i)$ and $\tilde{y}_i = \min(t, y_i)$, and we get $\|\tilde{x}\|_1 < \|\tilde{y}\|_1$. Using the augmentation property of Proposition 2.5, there exists $i \in [K]$ such that $\tilde{x}_i < \tilde{y}_i$ and $\tilde{x} + \varepsilon e_i \in \mathbf{M}$. Finally, observe that $x <_{\min} \tilde{x} + \varepsilon e_i$ which contradicts the leximin optimality of x.

To prove (2), we proceed by induction. The property holds at $j = 1$ by definition of $\mathrm{P}_{\text{leximin}}$. Assume the property holds at $j-1$. For the sake of contradiction, assume that there exists $y \in \mathbf{M}$ such that $\sum_{i\in[j]} x_{(i)} < \sum_{i\in[j]} y_{(i)}$. Then, we have $x_{(i)} < y_{(i)}$. Setting $t = y_{(i)}$, observe that

$$\sum_{j=1}^{K} \min(t, x_j) \leq \sum_{j=1}^{i} x_{(i)} + (K-i+1)\cdot t < \sum_{j=1}^{i} y_{(i)} + (K-i+1)\cdot t = \sum_{j=1}^{K} \min(t, y_j).$$

This contradicts (1), which concludes the induction proof of (2). To prove (3), take $y \in \mathbf{M}$, $\lambda_1 \geq \lambda_2 \geq \dots \lambda_K \geq \lambda_{K+1} = 0$, and write

$$\begin{aligned}
\lambda_1 x_{(1)} + \dots + \lambda_K x_{(K)} = \quad & (\lambda_1 - \lambda_2)x_{(1)} + (\lambda_2 - \lambda_3)(x_{(1)} + x_{(2)}) + \dots \\
& + (\lambda_K - \lambda_{K+1})(x_{(1)} + \dots + x_{(K)}) \\
\geq \quad & (\lambda_1 - \lambda_2)y_{(1)} + (\lambda_2 - \lambda_3)(y_{(1)} + y_{(2)}) + \dots \\
& + (\lambda_K - \lambda_{K+1})(y_{(1)} + \dots + y_{(K)}) \\
= \quad & \lambda_1 y_{(1)} + \dots + \lambda_K y_{(K)},
\end{aligned}$$

where the inequality is a (positive) linear combination of inequalities from property (2).

Proof of Proposition 2.10

Let $z \in P$. Because P corresponds also to the set of maximum matchings, we have that $\|z\|_1$ is equal for all points in P, in particular $\|x\|_1 = \|z\|_1$. Using Proposition 2.9, we know that for all $j \in [K]$, $\sum_{i\in[j]}$, $\sum_{i\in[j]} x_{(i)} \geq \sum_{i\in[j]} z_{(i)}$. Using additionally that both points z and x sum to the same quantity, this implies that for every $j \in [K]$, $\sum_{i=j}^{K} x_{(i)} \leq \sum_{i\in[j]} z_{(i)}$. The previous inequalities taken with $\|z\|_1 = \|x\|_1$ means that x is majorized by any point $z \in P$. Therefore by Karamata's inequality, because $x \mapsto x^p$ is strictly convex for $p > 1$, x is the unique minimizer in P of the function $\sum_{i\in K} z_i^p$, which is simply $\|z\|_p^p$.

To see that x also uniquely minimize the variance, it is enough to remark that for any $z \in P$, $\mathbf{Var}(z) = \frac{1}{K}\sum_{i\in[K]}(z_i - \frac{1}{K}\sum_{l\in[K]} z_l)^2 = \frac{1}{K}\sum_{i\in[K]}(z_i - \|x\|_1)^2 = \frac{1}{K}\|z\|_2^2 + (1 - 2/K)\|x\|_1^2$, which only depend on z through $\|z\|_2^2$, which is uniquely minimized by the leximin optimal x.

Proof of Proposition 2.11

Let $h = \frac{M_{[K]}}{K}\mathbf{1} \in F_1$. For any $w \in \mathbf{P}$ and $z = t\mathbf{1} \in F_1$, observe that

$$\|w - z\|_2^2 = \|(w-h) - (h-z)\|_2^2 = \|w-h\|_2^2 + \|h-z\|_2^2 + 2\langle w-h | h-z\rangle$$

Because F_1 and $\mathbf{P}$ are orthogonal, the scalar product is equal to zero:

$$\langle w-h|h-z\rangle = \sum_{i\in K}(w_i - \frac{M_{[K]}}{K})(\frac{M_{[K]}}{K} - t) = (\|w\|_1 - M_{[K]})(\frac{M_{[K]}}{K} - t) = 0.$$

Using the fact that $\|w\|_2^2 = \|w-h\|_2^2 + \|h\|_2^2$, we have that $\|w-z\|_2^2 = \|w-h\|_2^2 + \|h-z\|_2^2 = \|w\|_2^2 - \|h\|_2^2 + \|h-z\|_2^2$. Therefore, when projecting z onto $\mathbf{P}$, we have that $\|w-z\|_2$ is minimized when $\|w\|_2$ is maximized, that is when $w = x$ using Proposition 2.10. When projecting w onto $F_1 \cap \text{co}(\mathbf{M})$, we have that $\|w-z\|_2$ is minimized when $\|h-z\|_2$ is minimized, that is when $t = t^*$ because $t^* \le M_{[K]}/K$.

Proof of Proposition 2.12

We prove the desired equality by proving both inequalities. Let $x^* \in \text{co}(\mathbf{P})$ and $y^* \in \mathbf{H}$. By reverse triangle inequality it follows,

$$\max_{x\in\text{co}(\mathbf{M})} \|x\|_1 - \max_{y\in\mathbf{H}} \|y\|_1 = \|x^*\|_1 - \max_{y\in\mathbf{H}} \|y\|_1 \le \|x^*\|_1 - \|y^*\|_1 \le \|x^* - y^*\|_1.$$

Since the left-hand side does not depend on the points x^* and y^* chosen, taking infimum on both sides we obtain the first of the two inequalities. Conversely, let $y^* \in \mathbf{H}$ so that $\|y^*\|_1 = \max_{y\in\mathbf{H}} \|y\|_1$ and $x^* \in \text{co}(\mathbf{P})$ satisfying that $x_i^* \ge y_i^*$ for all $i \in [K]$. Then,

$$\begin{aligned}\|x^* - y^*\|_1 &= \sum_{i\in[K]} |x_i^* - y_i^*| = \sum_{i\in[K]} x_i^* - y_i^* = \|x^*\|_1 - \|y^*\|_1 \\ &= \max_{x\in\text{co}(\mathbf{M})} \|x\|_1 - \max_{y\in\mathbf{H}} \|y\|_1,\end{aligned}$$

from where the second inequality follows by taking infimum. Finally, the existence of a point $(x, y) \in \text{co}(\mathbf{P}) \times \mathbf{H}$ attaining the infimum is due to $\text{co}(\mathbf{P})$ and $\mathbf{H}$ being compact sets on $\mathbb{R}^K$ and $\|\cdot\|_1$ being a continuous function. Indeed, notice that both sets are bounded as they are subsets of $\mathbf{M}$ (which is bounded as well as $\mathcal{G}$ is a finite graph), $\text{co}(\mathbf{P})$ is the closed convex hull of $\mathbf{P}$, and $\mathbf{H}$ is closed by assumption.

Proof of Proposition 2.17

Let c^* be the solution to Equation (2.1) and Λ^* the corresponding set of groups. Without loss of generality, let us take $\Lambda = [\ell]$, for $\ell \le K$, and denote $\bar{M} := \max_{i\in[K]} M_i$, $\bar{m} := \min_{i\in[K]} M_i$, $\hat{m} = \bar{M}/\bar{m}$. It follows,

$$\begin{aligned}\text{PoF}_\text{O} =& \frac{M_{[K]}}{M_{[\ell]}} \cdot \frac{\sum_{i\in[\ell]} M_i}{\sum_{i\in[K]} M_i} = \left(1 + \frac{M_{1\ldots\ell+1} + \ldots + M_{1\ldots K}}{M_{[\ell]}}\right) \cdot \frac{\sum_{i\in[\ell]} M_i}{\sum_{i\in[K]} M_i} \\ &\le (1 + (K-\ell)\hat{m})\frac{\ell}{K}\hat{m}.\end{aligned}$$

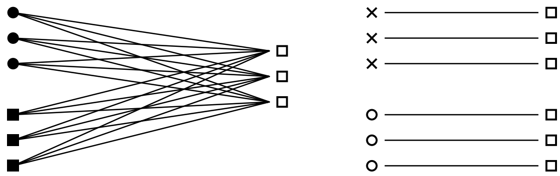


Fig. 2.12: Tight bound example for $M = 3$, $K = 4$, and white squares as jobs

Since ℓ is integral, the function $\ell \mapsto (1+(K-\ell)\hat{m})\frac{\ell}{K}\hat{m}$ reaches its maximum at $\ell^* = \frac{K}{2} + \frac{\mathbb{1}_{K\text{odd}}}{2\hat{m}}$. Plugging this in the previous bound, we get,

$$\begin{aligned}
\text{PoF}_\text{O} &\leq \left(1 + \left(\frac{K}{2} - \frac{\mathbb{1}_{K\text{odd}}}{2\hat{m}}\right)\hat{m}\right)\left(\frac{1}{2} + \frac{\mathbb{1}_{K\text{odd}}}{2K\hat{m}}\right)\hat{m} \\
&= \left(1 + \frac{K\hat{m}}{2} - \frac{\mathbb{1}_{K\text{odd}}}{2}\right)\left(\frac{\hat{m}}{2} + \frac{\mathbb{1}_{K\text{odd}}}{2K}\right) \\
&= \frac{\hat{m}}{2} + \frac{\mathbb{1}_{K\text{odd}}}{2K} + \frac{K\hat{m}^2}{4} + \frac{\mathbb{1}_{K\text{odd}}\hat{m}}{4} - \frac{\mathbb{1}_{K\text{odd}}\hat{m}}{4} - \frac{\mathbb{1}_{K\text{odd}}}{4K} \\
&= \frac{\hat{m}}{2} + \frac{K\hat{m}^2}{4} + \frac{\mathbb{1}_{K\text{odd}}}{4K},
\end{aligned}$$

which concludes the proof of the upper bound.

Now let us look at a tight counter-example. Suppose that $M_i = M, \forall i \in [K]$, for $M \in \mathbb{N}$. Consider a graph with M jobs connected to $M\lceil\frac{K}{2}\rceil$ agents divided into $\lceil\frac{K}{2}\rceil$ groups $G_1, ..., G_{\lceil K/2 \rceil}$, each group with M agents, forming a complete subgraph. In addition, consider $\lfloor\frac{K}{2}\rfloor$ subgraphs, each of them composed of M jobs connected to M agents, each subgraph corresponding to a different group $G_{\lceil K/2 \rceil+1}, ..., G_K$. Figure 2.12 illustrates such a graph. An opportunity fair matching can match at most $M/\lceil K/2 \rceil$ agents of each group due to the competition among the first $\lceil K/2 \rceil$ groups. It follows

$$\max_{x \in \mathbf{F}_\text{O}} \|x\|_1 = \frac{KM}{\lceil\frac{K}{2}\rceil} \text{ while } \max_{x \in \mathbf{M}} \|x\|_1 = M + M\lfloor\frac{K}{2}\rfloor \implies \text{PoF}_\text{O} = \frac{(1 + \lfloor K/2 \rfloor)\lceil K/2 \rceil}{K},$$

which is exactly $\frac{1}{2} + \frac{K}{4} + \frac{\mathbb{1}_{K\text{odd}}}{4K}$.

Proof of Proposition 2.18

The first part of the proposition is immediate. Suppose $\text{PoF}_\text{O} > 1$ then, as for the proof of Theorem 2.15, the point $(M_1, \ldots, M_K)/(K-1)$ is feasible, yielding $\text{PoF}_\text{O} \leq \rho(K-1) \leq 1$ as $\rho \in [1/K, 1/(K-1)]$, which is a contradiction.

Suppose next that $\rho \in [1/(K-1), 1]$ and denote $\Delta := \sum_{i \in [K]} M_i - M_{[K]}$ the additive difference between the Utopian matching and the optimum one. By assumption we get $\Delta = (1-\rho)KM$.

Now let us compute c^*. Without loss of generality (the same argument apply to other permutations) we consider the sequence of groups $([t])_{t\in[K-1]}$, and define, for $t \in [K-1]$,

$$c(t) = \frac{\sum_{i\in[t]}(M_{[i]} - M_{[i-1]})}{\sum_{i\in[t]} M_i},$$

with $M_0 = 0$. Finally, denote by $\delta_i = M_i - (M_{[i]} - M_{[i-1]})$, for $i \in [K]$. In particular, $\delta_1 = 0$. For $t \in [K-1]$, it follows,

$$\begin{aligned}
c(t) &= \frac{\sum_{i\in[t]} M_{[i]} - M_{[i-1]}}{\sum_{i\in[t]} M_i} = \frac{\sum_{i\in[t]} M_i - M_i + M_{[i]} - M_{[i-1]}}{\sum_{i\in[t]} M_i} = 1 - \frac{\sum_{i\in[t]} \delta_i}{\sum_{i\in[t]} M_i} \\
&= 1 - \frac{\Delta}{\sum_{i\in[t]} M_i} + \frac{\sum_{i=t+1}^{K} \delta_i}{\sum_{i\in[t]} M_i} = 1 - (1-\rho)\frac{K}{t} + \frac{\sum_{i=t+1}^{K} \delta_i}{tM} \\
&\geq 1 - (1-\rho)\frac{K}{t} + \max\Big(0, \frac{(1-\rho)K - (t-1)}{t}\Big),
\end{aligned}$$

where we have used that $\sum_{i=t+1}^{K} \delta_i = \Delta - \sum_{i=2}^{t} \delta_i = (1-\rho)KM - (t-1)M$. The maximum being greater than 0 is equivalent to $t \leq (1-\rho)K + 1$. As t is an integer, it holds that for all $t \leq \lfloor(1-\rho)K + 1\rfloor$, $c(t) \geq 1/t$. This function being decreasing, its value is always greater than $1/\lfloor(1-\rho)K + 1\rfloor$. Otherwise if $t \leq \lceil(1-\rho)K + 1\rceil$, the maximum is equal to 0, and $c(t) \geq 1 - (1-\rho)K/t$, which is increasing in t, thus $c(t) \geq 1 - (1-\rho)K/(\lceil(1-\rho)K+1\rceil)$. This is valid for all permutations of increasing group sets. Note that the relaxed bound can simply be obtained at this point by making t go to exactly $(1-\rho)K + 1$ in both cases. Overall, we obtain that $c^* \geq \min(1 - (1-\rho)K/(\lceil(1-\rho)K+1\rceil), 1/\lfloor(1-\rho)K+1\rfloor)$. Simplifying this expression as $\lceil(1-\rho)K+1\rceil = K + \lceil -K\rho\rceil + 1 = K - \lfloor K\rho\rfloor + 1$ and $\lfloor(1-\rho)K+1\rfloor = K + \lfloor -K\rho\rfloor + 1 = K - \lfloor K\rho\rfloor$, we get,

$$\mathrm{PoF_O} = \frac{\rho KM}{c^* KM} \leq \rho \max\left(\frac{K - \lfloor K\rho\rfloor + 1}{K\rho - \lfloor K\rho\rfloor + 1}, K - \lfloor K\rho\rfloor\right) \leq \rho((1-\rho)K + 1).$$

To end the proof, we provide and example to show the tightness of the bound, which is a continuous parametrization of the one in Section 2.9.4. Consider a graph with $\lfloor K\rho\rfloor - 1$ independent groups and that can match M agents each, one partially independent group which can match $(K\rho - \lfloor K\rho\rfloor)M$ nodes independently, and the rest of the K groups which must share M jobs between them and with $(1-\alpha)M$ agents of the partially independent group. It holds $M_i = M$ for all $i \in [K]$ and $M_{[K]} = (\lfloor K\rho\rfloor + \alpha)M = \rho KM$.

Let us compute the value of the fair optimum. There are two distinct cases.

1. $K\rho - \lfloor K\rho\rfloor \geq 1/(K - \lfloor K\rho\rfloor)$: The best identical fraction of the entitlement that the $K - \lfloor K\rho\rfloor$ competing groups (without the partially independent group) can get is $1/(K - \lfloor K\rho\rfloor)$, which is smaller than the number of matched agents that the partially independent group can get alone. Hence, it is sub-optimal to share jobs with this group, implying that the fair optimum value

is equal to $KM/(K - \lfloor K\rho \rfloor)$. In particular, we get $\text{PoF}_{\text{O}} = \rho(K - \lfloor K\rho \rfloor)$. We claim that the maximum in the upper bound is indeed this quantity. By assumption it holds $(K\rho - \lfloor K\rho \rfloor + 1) \geq (K - \lfloor K\rho \rfloor + 1)/(K - \lfloor K\rho \rfloor)$ which implies that

$$\frac{K - \lfloor K\rho \rfloor + 1}{K\rho - \lfloor K\rho \rfloor + 1} \leq K - \lfloor K\rho \rfloor,$$

concluding this first case.

2. $K\rho - \lfloor K\rho \rfloor \leq 1/(K - \lfloor K\rho \rfloor)$: In this case, some jobs need to be shared with the partially independent group. Let f_c and f_p be the optimal fraction that should be given to a competing group and to the partially independent group, respectively. We obtain the system of equations,

$$\left.\begin{array}{l} (K - \lfloor K\rho \rfloor) f_c + f_p = 1, \\ f_p + K\rho - \lfloor K\rho \rfloor = f_c. \end{array}\right\} \implies (K - \lfloor K\rho \rfloor + 1) f_c = K\rho - \lfloor K\rho \rfloor + 1,$$

This results in $\text{PoF}_{\text{O}} = \rho(K - \lfloor K\rho \rfloor + 1)/(K\rho - \lfloor K\rho \rfloor + 1)$. As done for the first case, we can show using that the maximum in the upper bound is equal to this quantity.

Proof of Proposition 2.19

Let c^* be the optimal value of Equation (2.1) and Λ^* be the subset that attains the minimum. We claim that M^σ being non-increasing for every σ implies that $\Lambda^* = [K]$ and thus, $\text{PoF}_{\text{O}} = 1$. Consider $\sigma = \text{I}_K$ the identity permutation (for the rest of permutations the argument is the same one), $x = \text{P}_\sigma$, and denote $B_i := \frac{M_{[i]}}{\sum_{j \in [i]} M_j}$. It follows,

$$B_{i+1} - B_i = \frac{M_{[i+1]}}{\sum_{j \in [i+1]} M_j} - \frac{M_{[i]}}{\sum_{j \in [i]} M_j} = \frac{\sum_{j \in [i]} (x_{i+1} M_j - M_{(i+1)} x_j)}{(\sum_{j \in [i+1]} M_j)(\sum_{j \in [i]} M_j)}.$$

Since M^σ is non-increasing, for any $j < i + 1$,

$$\frac{x_j}{M_j} \geq \frac{x_{i+1}}{M_{i+1}} \iff M_{i+1} x_j \geq x_{i+1} M_j \implies B_i \geq B_{i+1},$$

where the last implication comes from the fact that $B_{i+1} - B_i$ is the sum of only negative values. Therefore, when computing $\min_{i \in [K]} \frac{M_{[i]}}{\sum_{j \in [i]} M_j}$ the minimum is attained by $[K]$. For the rest of permutations the argument is the same. We conclude that $\Lambda^* = [K]$.

Proof of Proposition 2.20

Consider $\sigma = \text{I}_K$ the identity permutation and $x = \text{P}_\sigma$. For any other permutation the analysis is analogous. We prove that M^σ as defined in Proposition 2.19 is non-increasing by running Algorithm 5. Let x_1 be the maximum number of G_1 agents that can be matched. Match all of them and take out of the graph all the matched vertices. The resulting graph is still complete and contains $K-1$ groups. Let x_2 be the maximum number of G_2 agents that can be matched in the subgraph. Because the graph is complete there are three options:

1. $x_2 = M_2$ in which case all G_2 agents can be matched. Match them all, take all the matched vertices out of the graph, and repeat the procedure with G_3.

2. $x_2 = 0$ in which case the subgraph has $U = \varnothing$ and then $x_i = 0$ for all $i \in \{2, ..., K\}$.

3. $M_2 > x_2 > 0$. Match the x_2 possible agents and take out of the graph all matched vertices. At the following step, the only possible case is case 2.

In any case, the sequence M^σ corresponds to a sequence of only ones (every time case 1 holds), eventually an intermediary case where $x_i \in (0, 1)$ (when case 3 holds), and then all posterior entries are zero (only case 2 holds). The sequence M^σ is therefore, non-increasing. We conclude using Proposition 2.19.

Proof of Proposition 2.21

We will use the following Theorem:

Theorem (Theorem 2.2 [FK16]). *For any Erdös-Rényi random graph with $p \leq \frac{1}{\omega n^{3/2}}$, where $\omega = \omega(n) \to \infty$ arbitrarily slow as $n \to \infty$, $\mathbb{G}$ is a collection of edges and vertices with high probability.*

Denote $p_0 := \max_{i \in [K]} p_i$ and consider the non-bipartite random graph $\mathbb{G}^0$ as the graph obtained with vertex set $U \cup V$ and edge probability p_0. By the Theorem recalled in Appendix Section 2.9.4, $\mathbb{G}^0$ is a collection of edges and vertices with high probability. The random bipartite graph $\mathbb{G}$ is stochastic dominated [4] by $\mathbb{G}^0$ so, in particular, with high probability, it also corresponds to a collection of isolated vertices and simple edges. It follows that for any $i \in [K]$, $M_{[i]} = \sum_{j=1}^{i} M_j$, which implies that for any σ, the sequence M^σ as defined in Proposition 2.19, is constant equal to 1. We conclude that $\text{PoF}_{\text{O}}(\mathbb{G}) = 1$.

Proof of Proposition 2.22

We will use the following Theorem:

[4] Stochastic dominance can be proved by using a coupling technique. For more details, please check Lemma 1.1 [FK16].

Theorem (Theorem 6.1 [FK16]). *Let* $p = \frac{\log(n)+\omega}{n}$ *with* $\omega = \omega(n) \to \infty$ *arbitrarily slow as* $n \to \infty$. *Then* $\lim_{n\to\infty} \mathbb{P}(\mathbb{G}$ *has a perfect matching*$) = 1$.

We first show that $|G_i|$ concentrates. The random variable $|G_i| \sim \mathrm{Bin}(n, \alpha_i)$ corresponds to a binomial random variable of parameters n and α_i. In particular, $\mathbb{E}[|G_i|] = \alpha_i n$. By Hoeffding's inequality, as $|G_i|$ can be written as the sum of i.i.d. Bernoulli random variables,

$$\mathbb{P}\Big(||G_i| - \alpha_i n| > \sqrt{n\log(n)}\Big) \le 2\exp\Big(-2\frac{(\sqrt{n\log(n)})^2}{n}\Big) = \frac{2}{n^2} \underset{n\to\infty}{\longrightarrow} 0.$$

Regarding M_i, consider the bipartite subgraph containing only the agents in G_i and U. Since $p_i \ge \log^2(n)/n$, Section 2.9.4 implies there exists a matching of size $\min(|G_i|, n) = \min(\alpha_i n, n)$ and, therefore, $M_i = \min(\alpha_i n, n)$. Finally, consider $\sigma \in \Sigma([K])$, match the first $M_{\sigma(1)}$ agents, and remove all matched vertices from $\mathbb{G}$, leaving the set U with $[n(1-\alpha_i)]_+$ nodes. Consider the second group $\sigma(2)$. Because $p_{\sigma(2)} \ge \log^2(n)/n$, from Section 2.9.4, $M_{\sigma([2])} = \min(|G_{\sigma(2)}|, [n(1-\alpha_i)]_+)$. In other words, either $G_{\sigma(2)}$ is totally matched and we move to study the group $\sigma(3)$, it is partially matched and so no agent in $G_{\sigma(3)}$ to $G_{\sigma(K)}$ can be matched, or no agent in $G_{\sigma(2)}$ is matched. In any case, the sequence M^σ consists on a sequence of ones, eventually a value in $(0,1)$, and then a sequence of zeros. We conclude by Proposition 2.19.

3 Two-sided matching: The role of correlation of priorities

Contents

This chapter is based on the articles [CLP22] and [CLP24] by Patrick Loiseau, Bary Pradelski and myself.

3.1 Introduction

In this chapter, we study how different ranking correlation between different socio-demographic groups—*differential correlation* – affects outcome inequality and efficiency in matching markets. Our findings point to a before unacknowledged source of inequity between different groups that is specific to matching markets and should be included in future assessments of, for example, school, university, and job admissions. In particular, we find that differential correlation across groups leads to outcome inequalities even when the rankings by each college are fair, i.e., all groups are represented in each college's ranking as they are in the total applicant population. The resulting inequity is a form of systemic discrimination, i.e., discrimination that only arises through the interaction of decision-makers—via the matching mechanism—and is not due to either intentional or non-intentional discrimination by single decision-makers (cf. [Pin96; Fea13; BHI22]).

Differential correlation arises when different decision-makers—say colleges—use different information on candidates from different (socio-demographic) groups—say students—when assigning priority scores to rank and admit them. This might be the case for several reasons.

First, the cost of information acquisition may vary across groups and some information may not be available at all for some groups. Consider decentralized college admissions to competitive PhD programs and compare foreign and local candidates. Several U.S. programs virtually reserve one seat for the best foreign candidate from a given foreign school based on test scores and—due to the high cost—do not perform in-person interviews (e.g., Iran's Sharif University of Technology or India's IIT). Consequently, the priorities of foreign candidates at different programs are highly correlated. By contrast, for local candidates grades from undergraduate alone may not provide sufficient information—also due to grade inflation—and universities rely on idiosyncratic signals, such as reference letters, extracurricular activities, or interviews during campus visits (which are less costly as only local travel is required). Consequently, the priorities of local students, while correlated, are less correlated than for foreign students.

Second, differences in correlation may also arise if colleges are looking for different attributes in candidates and proxies of these attributes are more or less correlated for different groups. Consider two colleges, one admitting students for mathematics, the other admitting students for physics. Suppose that there are two groups; one group of students come from high schools where physics is taught in a theoretical manner and the other group come from high schools where physics is taught in an experimental manner and thus less mathematical. Consequently, the students from the former group will exhibit higher correlation in their high school grades in mathematics and physics than the students from the latter group.

Third, differences in correlation may also arise when colleges use selection criteria that are more or less prevalent within different groups. Concretely, such criteria could include diversity with respect to the current student body, sibling priority, or proximity. Notably, such criteria are also used in centralized school choice mechanisms, as, for example, in Chile [Mel22]. Consequently, a group for which the criteria are more prevalent will exhibit higher correlation than a group for which the criteria are less prevalent.

The latter criteria—sibling priority and proximity—along others are commonly used to break ties. This allows to make a mechanism more explainable and only use random tie-breaking for a smaller number of students (for whom all criteria are the same). Such random tie-breaking has been studied in school choice problems [ACY15; ANR19; Arn23]. When students have the same ranking at a given college a tie-breaking rule describes who should receive priority. Two natural choices are that each college breaks ties independently or colleges use a common order to break ties. Intuitively, the former leads to 0 correlation and the latter to correlation 1 (among those students for whom tie-breaking is required). As we shall see, our work also allows to extend the known theoretical results on tie-breaking to accommodate intermediate correlation levels, as often present in practice.

3.1.1 Our contribution

We study the college-admissions problem, where multiple decision-makers select a subset of applicants from an applicant pool with stability as the solution concept [GS62; AL16]. Concretely, suppose that an infinite population of students divided into groups $G_i \dots G_K$ apply to colleges A and B. Groups represent, for example, protected attributes, such as gender or race. Each college assigns a priority score to each student. A given student gets priority scores W^A at college A and W^B at college B. We propose an original model for the distributions of these scores, to study the correlation between the rankings made by the different colleges.

To formalize correlation and thus capture the vague notion of "a connection between two things in which one thing changes as the other does",[1] we leverage prior work on copulas and their relation with classical notions via *coherence*. This allows us to model correlation without a specific functional form and, in particular, nest classical notions as special cases, e.g., Spearman's and Kendall's correlation indices. With this at hand, we assume that the correlation between priority scores at different colleges depends on a candidate's group—we call this feature *differential correlation*.

How does differential correlation impact a stable matching's efficiency*, i.e., the number of students getting their first choice and* inequality*, i.e., the difference between groups in their probability to remain unmatched?* To answer this question, we first consider comparative statics. We show that efficiency is increasing in each group's correlation level, i.e., increasing the correlation level of any group increases the amount of students getting their first choice in all groups (Theorem 3.7). Moreover, the proportion of students from a given group remaining unmatched is increasing in

[1] Oxford Advanced Learner's Dictionary, 2023

its own correlation level and decreasing in the correlation level of all other groups (Corollary 3.9). This implies that it is advantageous to belong to the low-correlation group. We then show that a given efficiency level can be reached by a continuum of different correlation vectors, yielding different levels of inequality, thus showing that efficiency differences cannot explain inequality (Proposition 3.10). Finally, our results imply extensions of known results on tie-breaking (cf. [ANR19], [AN20], and [Arn23]), in particular to multiple priority classes and intermediate levels of correlation (Proposition 3.11).

3.1.2 Further related literature

Matching. The college admission problem, i.e, how to assign prospective students to colleges given each student's preferences and colleges' priorities over students and capacities such that the outcome is stable, was introduced by Gale and Shapley [GS62]. A variant of this model where colleges do not have priorities over students is commonly called the school choice problem, and has been investigated in ([BS99], [AS03], [Abd05], [ES06], [Yen13]). The idea of considering a continuum of students and a finite number of colleges has previously been exploited due to its analytical tractability (cf., [CLS14], [ACY15], [AL16], [Arn22]). We follow [AL16] who develop a supply and demand framework allowing to easily analyze the quality of a matching and deriving comparative statics.

Matching with correlated types. We study matching in the presence of correlation between the priority scores given by each college to a given student. A special case of this problem has been studied for centralized school choice problems, where a lot of students have the same priority and therefore ties are broken at random.[2] [ANR19; AN20], and [Arn23] compare the welfare of students in two settings: either one common lottery is used by all colleges, or all colleges draw independent lotteries. In our model, this corresponds to correlation 1 or 0, respectively, and our results nest elements of these prior papers. Another line of work has considered correlation between other features, e.g., correlation between students' preferences and colleges' rankings. In this context, [BH22] focus on one-to-one matching and identify a matching that does not favor one side over the other, while [CT19] and [LL20] study the stability-efficiency trade-off by comparing Deferred Acceptance and Top Trading Cycles (see [SS74]), and how the magnitude of this trade-off depends on the correlation between agents preferences and priorities. Considering a transferable utility model, [Gol21] studies how workers sort into two competing sectors (such that their wages are maximized) and the impact of technological change. While their model and analysis is quite distant to ours, it shares the use of copulas to model correlation and the necessary restriction to two sectors, respectively colleges.

Algorithmic monoculture. Finally, our work also contributes to a recent literature on algorithmic monoculture, i.e., the fact that recommendations, choices, and preferences become

[2] The implications of this feature on students' welfare have been studied by [EE08; APR09] and [ACY15].

homogeneous with the rise of algorithmic curation and analysis. [KR21] study the utility of multiple decision-makers who use algorithms to evaluate candidates. They show that decision-makers are sometimes better off each using a different, low-precision algorithm than all using the same high-precision one. In empirical work, [Bom+22] find that outcomes are more homogeneous when models and training data sets are shared between decision-makers. Through our theoretical analysis we thus shed light on the impact of algorithmic monoculture from the candidates' viewpoint.

3.1.3 Outline

This chapter is organized as follows. Section 3.2 introduces the model and the concept of differential correlation. Section 3.3 introduces our welfare metrics and presents preliminary results. Our main results are in Section 3.4. In Section 3.5, we discuss the extension of our results when there are more than two colleges, and Section 3.6 details some special cases where additional results can be provided. Finally, Section 3.7 concludes with a discussion on the generality of our findings and future avenues of research.

3.2 Setup

We here introduce the college admission problem with a continuum of students, then formalize the notion of *correlation* in Section 3.2.2, and introduce the supply and demand framework to identify stable matching in Section 3.2.3. A summary of the main notation used in this chapter is provided in Table 3.1.

3.2.1 Model

Let A and B be two colleges to which a continuum unit mass of students, S, is to be matched. The mass of a subset of S is measured with a function η.[3] Colleges have maximum capacities of the mass of students they can admit, $(\alpha^A, \alpha^B) := \alpha \in (0,1]^2$. The students are divided into K groups $G_1, \dots, G_K$, with a fraction $\gamma_j \in [0,1]$ of students belonging to G_j, with $\sum_{j=1}^K \gamma_j = 1$. Define the vector $\gamma := (\gamma_j)_{j \in [K]}$, using the notation $[K] := \{1, 2, \dots, K\}$. We denote the group to which a student $s \in S$ belongs by $G(s)$.

Students have strict preferences over colleges, and the amount of students preferring college A might differ between groups: among group G_j, a share $\beta_j \in [0,1]$ prefer college A to college B, the remaining $1 - \beta_j$ prefer B. When student s prefers college A to college B, we write $A \succ_s B$, and vice versa. Note that β_j is a share that is conditional on the group, and not a mass, for instance,

[3] The formal definition of this measure is deferred to Appendix 3.8.2.

$\eta(\{s \in G_j : A \succ_s B\}) = \gamma_j \beta_j$. We assume that all students prefer attending some college to remaining unmatched. We write $\beta := (\beta_j)_{j \in [K]}$ as we did for γ.

Each college assigns a priority score to each student, the higher the better. Each student s thus is assigned a vector of priority scores (W_s^A, W_s^B). This means that college C prefers $s \in S$ to s' if and only if $W_s^C > W_{s'}^C$. The (marginal) distribution of scores W_s^C given by college C to students in G_j is described by a probability density function (pdf) f_j^C defined over the support $I_j^C \subseteq \mathbb{R}$, assumed to be an interval. Let $I_j = I_j^A \times I_j^B$. We denote by $\underline{I}_j^C$ and $\bar{I}_j^C$ the lower and upper bounds of I_j^C. These bounds might be finite or not. We define $\mathbf{f} := (f_1^A, \dots, f_K^A, f_1^B, \dots, f_K^B)$, and denote by F_j^C the cumulative distribution function (cdf) associated to f_j^C.

Differential correlation Consider the joint distribution of the vectors (W^A, W^B). For each group G_j, the grade vectors of students $s \in G_j$ follow some distribution with pdf $f_j : I_j \to \mathbb{R}$ and cdf F_j. A joint distribution can be characterized by its marginals, i.e., the distribution of each component of the vector, and the shape of the joint distribution, captured by a coupling function, called *copula*. A copula is a cdf over $[0, 1]^n$, for some n, with uniform marginals. [Skl59]'s theorem states that any joint distribution can be decomposed into (independent) marginals and a unique copula:

Sklar [Skl59, Theorems 1, 2, and 3] *Let F be a n-dimensional cdf with marginal cdfs $F^1, ..., F^n$. Then there exists a unique n-dimensional copula $H : [0, 1]^n \to [0, 1]$ such that*

$$F(x^1, ..., x^n) = H(F^1(x^1), ..., F^n(x^n)).$$

Conversely, for any n-dimensional copula H and for any set of n 1-dimensional cdfs $F^1, ..., F^n$, $F(x^1, ..., x^n) := H(F^1(x^1), ..., F^n(x^n))$ is a n-dimensional cdf with marginals $F^1, ..., F^n$.

Each group G_j then has a joint distribution with joint pdf f_j and cdf F_j, that can be represented by its marginals F_j^A, F_j^B and a (unique) copula H_j. We assume that there exists a family of 2-dimensional copulas $(H_\theta)_{\theta \in \Theta}$ (Θ being an interval $[\underline{\Theta}, \bar{\Theta}]$ of $\mathbb{R}$) and, for all $j \in [K]$, there exists a parameter $\theta_j \in \Theta$ such that H_{θ_j} is the copula associated to G_j's distribution, i.e., $H_{\theta_j} = H_j$. This assumption is made without loss of generality, but some of our results will require additional assumptions regarding the copula family. Denote by $\theta := (\theta_j)_{j \in [K]}$ the vector containing each group's parameter. Notice that each groups has a different θ_j, and thus a different joint distribution. With some foresight to the explanations provided in Section 3.2.2, we call this feature of the model *differential correlation*.[4] Finally, we assume that all the copulas in $(H_\theta)_{\theta \in \Theta}$ have full support over $[0, 1]^2$. We write f_{j,θ_j} and F_{j,θ_j} instead of f_j and F_j as we will consider them as functions of θ.

Given a family of copulas $(H_\theta)_{\theta \in \Theta}$, we then refer to the tuple $(\gamma, \beta, \alpha, \mathbf{f}, \theta)$ as the *college admission problem*. Note that we only assume that distributions admit a density and have full support,

[4]This is in spirit of the notion of differential variance studied in [Eme+22] and [GLM21].

that the distribution family is parameterized by a scalar, and that the marginals remain the same for any θ. In Appendix 3.8.2 we discuss implications of these assumptions and provide details of distributions that satisfy them (i.e., Gaussian copulas, Archimedean copulas and examples thereof).

This model allows for each group to have different grade distributions at each college. Notice that even though inside a given group the students preferring A have the same grade distribution as students preferring B, this does not cause any loss of generality. Indeed, we can always split a given group into two groups with different distributions.

3.2.2 Correlation and the coherence assumption

The proxy for correlation in our model will be the parameter θ, rather than some specific functional form. We use a condition, namely coherence, on the family of distributions: whenever $(f_{j,\theta})_{\theta\in\Theta}$ is coherent, there exists a bijection between θ_j and classical measures of correlation. For details on classical correlation measures, namely Pearson's, Spearman's, and Kendall's correlation, see Appendix 3.8.2.

Assumption 3.1 (Coherence)**.** *We say that a family of copulas $(H_\theta)_{\theta\in\Theta}$ is* coherent *if for all $(x, y) \in (0, 1)^2$, $H_\theta(x, y)$ is strictly increasing in θ on Θ.*

The following lemma states that under coherence θ is naturally interpreted as a measure of correlation.

Lemma 3.1. *If a family of copulas $(H_\theta)_{\theta\in\Theta}$ is coherent, and (X, Y) is a random couple drawn according to H_θ, then $\forall(x, y) \in (0, 1)^2$, $\mathbb{P}(X < x, Y < y)$ and $\mathbb{P}(X > x, Y > y)$ are increasing in θ, while $\mathbb{P}(X < x, Y > y)$ and $\mathbb{P}(X > x, Y < y)$ are decreasing in θ.*

Proof. Since $H_\theta(x, y) = \mathbb{P}(X < x, Y < y)$ by definition, then the first part of the lemma is just a rewriting of the definition of coherence. For the second part, we have

$$\begin{aligned}\mathbb{P}(X > x, Y > y) &= \mathbb{P}(Y > y) - \mathbb{P}(X < x, Y > y)\\ &= \mathbb{P}(Y > y) - \mathbb{P}(X < x) + \mathbb{P}(X < x, Y < y)\end{aligned}$$

and $\mathbb{P}(Y > y), \mathbb{P}(X < x)$ are constant in θ (H_θ are copulas therefore they all have uniform marginals) while $\mathbb{P}(X < x, Y < y)$ is increasing, so $\mathbb{P}(X > x, Y > y)$ is also increasing. Finally, we also get that

$$\mathbb{P}(X > x, Y < y) = \mathbb{P}(Y < y) - \mathbb{P}(X < x, Y < y)$$

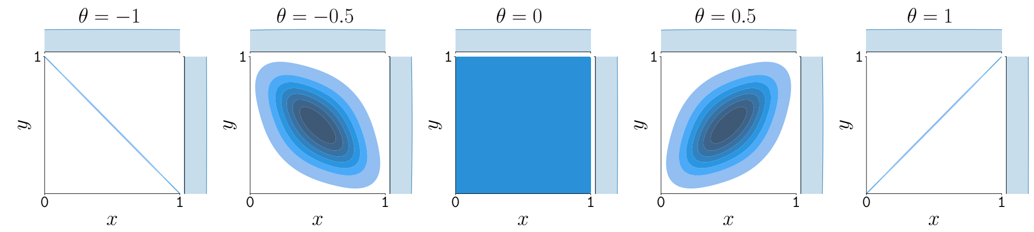


Fig. 3.1: Gaussian copula (i.e., bivariate Gaussian with marginals renormalized to uniform distributions, cf. Appendix 3.8.2) for five different correlation levels, θ. The shades of blue represent the distribution density (darker means higher).

and

$$\mathbb{P}(X < x, Y > y) = \mathbb{P}(X < x) - \mathbb{P}(X < x, Y < y)$$

and both are thus decreasing. ■

Intuitively, when X and Y are highly correlated, then if, for example, X is small then Y is likely also small.

Further supporting the choice of θ as our measure of correlation, we note that there is an equivalence between θ and two classical, ordinal measures of correlation, namely Spearman's correlation, denoted ρ, and Kendall's correlation, denoted τ.

Scarsini [Sca84, Theorems 4 and 5] *If $(H_\theta)_{\theta\in\Theta}$ is coherent, and (X_θ, Y_θ) are random variables drawn according to H_θ, then Spearman's and Kendall's correlation coefficients $\rho(X_\theta, Y_\theta)$ and $\tau(X_\theta, Y_\theta)$ are strictly increasing functions of θ.*

The above results show that, under the coherence assumption, θ is a bijection of classical measures of correlation. To illustrate the effect of θ on a coherent distribution family, Figure 3.1 shows draw from Gaussian copulas, which are coherent, with different values of the covariance used as θ. When $\theta = 0$ the variables are independent, when θ is positive the joint distribution gets closer to the diagonal $X = Y$, and when θ is negative the joint distribution gets closer to the diagonal $Y = -X$, which corresponds to the common idea of correlation as well as Spearman's and Kendall's correlations. Since most of our results will be qualitative, they are stated using θ but would still be true if θ was replaced by ρ or τ in the statement.

We finally introduce a technical assumption that will be required for some of our results, especially when considering comparative statics in θ.

Assumption 3.2 (Differentiability). *We say that $(H_\theta)_{\theta\in\Theta}$ is* differentiable *if for all $(x, y) \in \mathring{I}$ and for all $\theta \in \mathring{\Theta}$, $h_\theta(x, y)$ is differentiable in θ.*

The coherence and differentiability assumptions are not particularly restrictive, for instance the Gaussian copula with covariance as the parameter verifies them, as well as Frenkel's copula and almost all example copulas mentioned in Appendix 3.8.2.

3.2.3 The supply and demand framework

In the Introduction (Section 1.1), we defined matchings as explicit lists of couples of agents, and stability through the notions of justified envy and waste. A matching problem can be alternatively viewed through a supply and demand lens, where a stable matching is a Walrasian equilibrium. This characterization was first shown by Balinski and Sonmez [BS99], and was applied to the continuum model by Azevedo and Leshno [AL16].

Definition 3.1 (Cutoffs and demand)**.** If μ is a stable matching, define the *cutoff* at $C \in \{A, B\}$ as $P^C := \inf\{W_s^C : \mu(s) = C\}$. Given $\mathbf{P} = (P^A, P^B)$, we call the *demand* of student s, denoted $D_s(\mathbf{P}) \in \{A, B\} \cup \{s\}$, the college they prefer among those where their score is above the cutoff, or themselves if their score does not exceed the cutoff at any college. The *aggregate demand* at college C is the mass of students demanding it: $D^C(\mathbf{P}) = \eta(\{s : D_s(\mathbf{P}) = C\}$. We denote by $\mathbf{D}$ the vector (D^A, D^B).

The cutoff of a college represents the score above which a student who applies gets admitted. Recall that $I_j = I_j^A \times I_j^B$ is the support of $f_{j,\theta}$, $\underline{I}_j^A$ and $\bar{I}_j^A$ the lower and upper bounds of I_j^A,and same for B. If $P^C = \min_{j\in[K]} \underline{I}_j^C$, then college C rejects no one, if $P^C = \max_{j\in[K]} \bar{I}_j^C$ it accepts no one. The supply associated to this demand is simply the capacity of each college.

Consider the equilibria of this problem:

Definition 3.2 (Market clearing)**.** The cutoff vector $\mathbf{P}$ is *market clearing* if for $C \in \{A, B\}$, $D^C(\mathbf{P}) \leq \alpha^C$, with equality if $P^C > \min_{j\in[K]} \underline{I}_j^C$.

A cutoff vector is therefore market-clearing if it induces a demand that is equal to colleges' capacities when they reach their capacity constraint, and lower for colleges that are not full. When the constraint is reached at both colleges, i.e., when $\alpha^A + \alpha^B < 1$, the system

$$\mathbf{D}(\mathbf{P}) = \alpha \tag{3.1}$$

is called the *market-clearing equation*, and the market-clearing cutoffs P^A and P^B can be computed by solving the system.

The following result from [AL16] establishes the link between market-clearing cutoffs and stable matchings:

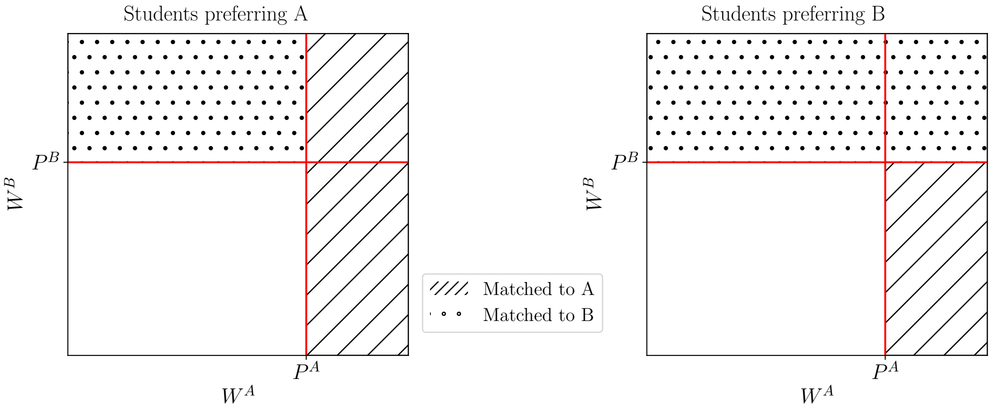


Fig. 3.2: Cutoff representation of stable matchings. Students in the hashed area are matched to college A, those in the dotted area to college B, and those in the white area remain unmatched.

Azevedo and Leshno [AL16, Lemma 1]

1. *If μ is a stable matching, the associated cutoff vector $\boldsymbol{P}$ is market-clearing;*
2. *If $\boldsymbol{P}$ is market-clearing, we define μ such that for all $s \in S$, $\mu(s) = D_s(\boldsymbol{P})$. Then μ is stable.*

This allows to analyze stable matchings by studying the cutoffs of each college. Figure 3.2 illustrates the link between the cutoffs and the matching: students who prefer A get admitted there if and only if their score W^A is higher than the cutoff P^A. Otherwise, they get admitted to B if their score W^B is higher than P^B, and stay unmatched if it is not. The situation is symmetric for students preferring college B.

In the continuous college admissions problem the same authors show that there is a unique stable matching.

Lemma 3.2 (Special case of [AL16, Theorem 1])**.** *For any college admission problem* $(\gamma, \beta, \alpha, \mathbf{f}, \theta)$*, there exists a unique stable matching.*

Note that the original theorem specifies conditions on the distribution of students' types, such as being continuous and having full support, which hold in our definition of a college admission problem. Unlike the finite case where typically several stable matchings exist, in the continuum model the stable matching is unique and therefore no considerations regarding selection among the set of stable matchings are necessary. From now on, we will therefore consider the cutoff vector **P** as the one uniquely determined by the parameters of the problem and the market-clearing equation. We shall say *student s goes to college C* to mean that they are matched to college C in the unique stable matching.

[AL16] further show that the stable matching varies continuously in the parameters of the problem and that the set of stable matchings from a college admission problem with a finite number of students converges to the unique stable matching of the continuum problem with the same parameters. The latter result justifies the approximation of large finite instances by their limit.[5]

3.3 Welfare metrics and preliminary results

In selection problems, inequalities between groups are measured by the proportion of admitted candidates in each group. In a matching setting, the situation is more complex: on the one hand, one group might have a higher proportion of unmatched students than the other, but on the other hand, the proportion of students getting their first choice might also differ. If all students in a group get their first choice and all students in the other get their second choice, the matching may be deemed unfair. In this section, we define metrics that allow us to quantify the satisfaction of students from each group.

Consider an individual's likelihood of getting their first choice, second choice, or being rejected from both colleges in a stable matching as a function of differential correlation.

Definition 3.3 (Welfare metrics)**.** Under a stable matching μ_θ induced by differential correlation parameters θ, define $V_1^{G_j,A}(\theta)$ and $V_1^{G_j,B}(\theta)$, as the proportion of students from each group-preference profile who get their first choice. Formally,

$$\begin{aligned} V_1^{G_j,A}(\theta) &:= \frac{1}{\gamma_j \beta_j}\eta(\{s \in G_j : A \succ_s B, \mu_\theta(s) = A\}), \\ V_2^{G_j,A}(\theta) &:= \frac{1}{\gamma_j \beta_j}\eta(\{s \in G_j : A \succ_s B, \mu_\theta(s) = B\}), \\ V_\emptyset^{G_j,A}(\theta) &:= \frac{1}{\gamma_j \beta_j}\eta(\{s \in G_j : A \succ_s B, \mu_\theta(s) = \emptyset\}). \end{aligned}$$

The metrics for students preferring college B are defined similarly, by inverting the roles of A and B and replacing β_j by $1 - \beta_j$ in the equations.

Those metrics can be thought of in two ways: $V_1^{G_j,A}(\theta)$, for instance, is the relative mass of students getting their first choice among those in group G_j who prefer college A, or equivalently, it is the probability of a randomly drawn student to get their first choice conditionally on belonging to G_j and preferring A.[6]

[5] For a better approximation for instances with a small number of students, [Arn22] proposes a related framework.

[6] We condition over preferences because of the following observation: assume there are two groups, if students from group G_1 all prefer college A, but only half of the students from group G_2 prefer A, and A has a low capacity and B a large one. Then very few students from G_1 will get their first choice while half of the students from G_2 are likely to get their first choice since it is a less demanded college. Thus students' satisfaction might differ across groups only due to their own preferences, and not because of differential correlation.

We next provide expressions for these metrics for the unique (cf. Lemma 3.2) stable matching μ_θ via its cutoffs, $\mathbf{P}$.

Lemma 3.3. *Let $C \in \{A, B\}$ be a college, $\overline{C}$ be the other college, and G_j be a group. Let $\boldsymbol{P}$ be the cutoffs associated to μ_θ, we have:*

$$V_1^{G_j,C}(\theta) = \mathbb{P}_j(W^C \geq P^C(\theta)), \tag{3.2}$$

$$V_2^{G_j,C}(\theta) = \mathbb{P}_{j,\theta_j}(W^C < P^C(\theta), W^{\overline{C}} \geq P^{\overline{C}}(\theta)), \tag{3.3}$$

$$V_\emptyset^{G_j,C}(\theta) = \mathbb{P}_{j,\theta_j}(W^C < P^C(\theta), W^{\overline{C}} < P^{\overline{C}}(\theta)). \tag{3.4}$$

Proof. Consider student $s \in G_j$ who prefers college A. By Lemma 1 from [AL16] (cf. Section 3.2.3), s is admitted to A if and only if $s \in D_A(\mathbf{P})$, i.e., if and only if their score at A is greater than P^A. Then by definition of η,

$$\begin{aligned} V_1^{G_j,A}(\theta) &= \frac{\eta(\{s \in G_j : A \succ_s B, \mu(s) = A\})}{\gamma_j \beta_j} \\ &= \mathbb{P}_{j,\theta_j}((W^A, W^B) \in [P^A, +\infty) \times \mathbb{R}) \\ &= \mathbb{P}_j(W^A > P^A). \end{aligned}$$

The same reasoning applies to $V_1^{G_j,B}$ if s prefers B, which proves (3.2).

The same student s is admitted to B if and only if $s \in D_B(P^A, P^B)$, i.e., if and only if $W_s^B \geq P^B$ and $W_s^A < P^A$. Then we have

$$\begin{aligned} V_2^{G_j,A}(\theta) &= \frac{\eta(\{s \in G_j : A \succ_s B, \mu(s) = B\})}{\gamma_j \beta_j} \\ &= \mathbb{P}_{j,\theta_j}((W^A, W^B) \in (-\infty, P^A) \times [P^B, +\infty)). \end{aligned}$$

The same reasoning applies to $V_2^{G_j,B}$, which proves (3.3).

Student s remains unmatched if and only if $W_s^A < P^A$ and $W_s^B < P^B$. Then we have

$$\begin{aligned} V_\emptyset^{G_j,A}(\theta) &= \frac{\eta(\{s \in G_j : A \succ_s B, \mu(s) = s\})}{\gamma_j \beta_j} \\ &= \mathbb{P}_{j,\theta_j}((W^A, W^B) \in (-\infty, P^A) \times (-\infty, P^B)). \end{aligned}$$

which proves (3.4). ∎

The notation $\mathbb{P}_{j,\theta_j}$ is used as shorthand for $\mathbb{P}_{(W^A,W^B)\sim f_{j,\theta_j}}$. Lemma 3.3 allows to compare probabilities of admission of different types of students, and derive comparative statics with respect to differential correlation.

Regarding the probability of staying unmatched, we can derive a simple yet important result.

Lemma 3.4. *The probability that a student remains unmatched depends only on their group and is independent of their preference. Moreover, the total mass of unmatched students is constant in any group's correlation level. Formally, let* **P** *be associated to* μ_θ*. Then, for* $j \in [K]$, $V_\emptyset^{G_j,A}(\theta) = V_\emptyset^{G_j,B}(\theta)$*; and* $\eta(\{s \in S : \mu_\theta(s) = \emptyset\}) = \max(0,\ 1 - \alpha^A - \alpha^B)$ *(which does not depend on* θ*).*

Proof. It is sufficient to notice that Equation (3.4) is symmetric in C and $\bar{C}$ to obtain the first part of the lemma. The second one follows from the fact that either there is excess capacity and everyone is matched, or both colleges are full and the mass of matched students is the sum of the capacities. ■

Remark 3.1. The first part of Lemma 3.4 could also be derived from the strategy-proofness for students of the student-proposing deferred acceptance algorithm [Rot85]. Indeed, the fact that students cannot improve their outcome by modifying the order of their preferences implies that them being unmatched or not does not depend on which college they reported to be their first choice.

With Lemma 3.4 at hand, we will use the notation $V_\emptyset^{G_j}$, since these quantities do not depend on the preference of students. For all the metrics we defined, when there is no ambiguity, we also omit the dependence on θ and write $V_i^{G_j,C}$ instead.

We now define two global metrics, i.e., metrics that are not conditioned on the groups and preferences of student, namely efficiency and inequality:

Definition 3.4 (Efficiency and Inequality)**.** Define the *efficiency* $E(\theta)$ of a matching as the proportion of students getting their first choice, and the *inequality* $L^{G_i,G_j}(\theta)$ between two groups, $i, j \in [K]$ as the difference of the probability of staying unmatched between those two groups:

$$\begin{aligned} E(\theta) &= \eta\left(\{s \in S : \mu_\theta(s) = C \text{ and } C \succ_s \bar{C}\}\right) \\ &= \sum_{j\in[K]} \gamma_j \beta_j V_1^{G_j,A}(\theta) + \sum_{j\in[K]} \gamma_j (1-\beta_j) V_1^{G_j,B}(\theta) \qquad (3.5) \\ L^{G_i,G_j}(\theta) &= |V_\emptyset^{G_i}(\theta) - V_\emptyset^{G_j}(\theta)|. \qquad (3.6) \end{aligned}$$

By Lemma 3.4 the mass of unmatched, and therefore matched, students is constant, and thus matched students get either their first or second choice. Therefore, ceteris paribus, it is desirable to maximize the mass of students getting their first choice E. Regarding inequality, we measure the inequality between two groups by the difference in their proportions of unassigned students.

Proposition 3.5. *If two groups have the same marginal distributions at some college C, then for students whose first choice is college C the probability of getting this college is the same for students of both groups. Formally, if $f_j^C = f_\ell^C$, then $V_1^{G_j,C} = V_1^{G_\ell,C}$.*

Proof. The result follows directly by applying (3.2) to both groups, and by observing that the integral in (3.2) depends on θ only through the cutoff vector $\mathbf{P}$. Therefore, if groups G_i and G_j have the same marginal F_j^C at college C, then $V_1^{G_i,C} = V_1^{G_j,C}$. ■

Proposition 3.5, albeit simple, is an important property of the model. If two students prefer the same college, their probabilities of getting it only depend on their respective groups' marginals, and not on their correlation levels — so differential correlation has no effect on this metric. Proposition 3.5 also justifies the choice of L as a measure of inequality: the proportion of students getting their first choice is the same for two groups as long as they have the same marginals, and differences only emerge in second choice admittance versus remaining unmatched. Consequently, when the proportion of unmatched students is higher in one group than the other, then the matching is unequal.[7]

The following result shows that if there is capacity excess, differential correlation does not affect the stable matching.

Proposition 3.6. *If capacity is not constrained, i.e., $\alpha^A + \alpha^B \geq 1$, then correlation has no effect on the stable matching. The cutoffs P^A and P^B are constant in θ, on then so are $V_1^{G_j,C}$ and $V_2^{G_j,C}$ for all j and C. Moreover, $V_\emptyset^{G_j} = 0$, therefore $\forall i, j \in [K]$, $L^{G_i,G_j}(\theta) = 0$.*

Proof. The value of V_1 comes from Lemma 3.3. If $\alpha^A + \alpha^B \geq 1$, then all students are admitted to some college, therefore $V_\emptyset^{G_j} = 0$ for all j. Moreover, either $P^A = \min_j \underline{I}_j^A$ or $P^B = \min_j \underline{I}_j^B$, or both. Let us suppose the former holds. Then

$$V_1^{G_j,A} = 1 - F_j^A(\min_j \underline{I}_j^A) = 1$$
$$V_2^{G_j,A} = 0$$

[7] Another choice to measure inequality could be to compare the proportions of students getting their second choice in each group, however Proposition 3.5 implies that this quantity is equal to L.

and

$$V_1^{G_j,B} = 1 - F_j^B(P^B)$$
$$V_2^{G_j,B} = F_j^B(P^B).$$

It remains to show that P^B is constant in θ. Define

$$T : x \in \mathbb{R} \mapsto \sum_j \gamma_j(1-\beta_j)\left(1 - F_j^B(x)\right).$$

Note that since all F_j^B are invertible T is invertible too. Since all students preferring A get it, the market clearing equation for B becomes $T(P^B) = \alpha^B$, i.e., $P^B = T^{-1}(\alpha^B)$. From this relation it appears that P^B is indeed constant in θ. The same reasoning applies if $P^A \neq \min_j \underline{I}_j^A$ and $P^B = \min_j \underline{I}_j^B$. ■

More detail about this case can be found in Section 3.6.1.

3.4 Main results

This section contains our main results on the impact of differential correlation on the properties of stable matchings. We consider college admission problems where γ, β and α are assumed constant, and study the influence of differential correlation, θ, on the stable matching. Section 3.4.1 contains general comparative statics and Section 3.4.2 studies tie-breaking.

3.4.1 Comparative statics

We first consider how the efficiency of the matching, i.e., the probability of getting one's first choice, varies when changing the correlation for one group.

Theorem 3.7. *Suppose that $(H_\theta)_{\theta\in\Theta}$ is coherent and differentiable, and that $\alpha^A + \alpha^B < 1$. Then for all groups and all preferences the proportion of students getting their first choice is increasing in all correlation parameters θ_j, $j \in [K]$, and consequently so is the global efficiency $E(\theta)$. Formally, suppose that $\theta \in (\mathring{\Theta})^K$. Then for any $C \in \{A, B\}$, for any $j, \ell \in [K]$, $V_1^{G_j,C}(\theta)$ is differentiable and*

$$\frac{dV_1^{G_j,C}(\theta)}{d\theta_\ell} > 0.$$

The immediate consequence is that $E(\theta)$ is differentiable and for any $j \in [K]$,

$$\frac{dE(\theta)}{d\theta_j} > 0.$$

The proof relies on the following lemma:

Lemma 3.8. *Suppose that $\theta \in (\mathring{\Theta})^K$. Then for any $C \in \{A, B\}$, $P^C(\theta)$ is differentiable and*

$$\frac{dP^C(\theta)}{d\theta_j} < 0 \, \forall j \in [K].$$

Proof sketch. The proof follows several steps. First, we rewrite the market-clearing Equation (3.1) using Lemma 3.3. We obtain a system of two equations, where the variables are the cutoffs P^A and P^B, parameterized by θ. We then apply the implicit function theorem to a mapping whose roots are the solution of this system of equations. We next compute the partial derivatives. To characterize the sign of the derivatives with respect to θ, we use the coherence assumption. Through analytical derivations, we can conclude. The proof is provided in Appendix 3.8.3. ■

Proof of Theorem 3.7. From Lemma 3.8, the cutoffs are decreasing in each θ_j. We can then conclude that for any $j \in [K]$ and for $C \in \{A, B\}$,

$$\frac{dV_1^{G_j,C}}{d\theta_j} = \frac{d\int_{P^C}^{\infty} f_j^C(x)\,\mathrm{d}x}{d\theta_j} = \frac{d\int_{P^C}^{\infty} f_j^C(x)\,\mathrm{d}x}{dP^C} \cdot \frac{dP^C}{d\theta_j} > 0. \quad ■$$

Theorem 3.7 implies that, if the correlation decreases for one of the groups, then all groups suffer from a decrease in first-choice admittance. Conversely, increasing the correlation for one group leads to an increase in first-choice admittance for all groups.

Intuitively, when the correlation increases, students' score vectors accumulate close to the diagonal, and therefore in the lower-left and upper-right quadrants, while the other two quadrants are increasingly empty. This phenomenon is illustrated in Figure 3.3 with a bivariate Gaussian distribution. If the cutoffs did not change, then the amount of unmatched students would increase. As the capacities are assumed constant this would renter the resulting matching unstable. Therefore, at least one of the cutoffs decreases and, in fact, Lemma 3.8 implies that both decrease. As a consequence, the mass of matched students remains the same but more students get their first choice.

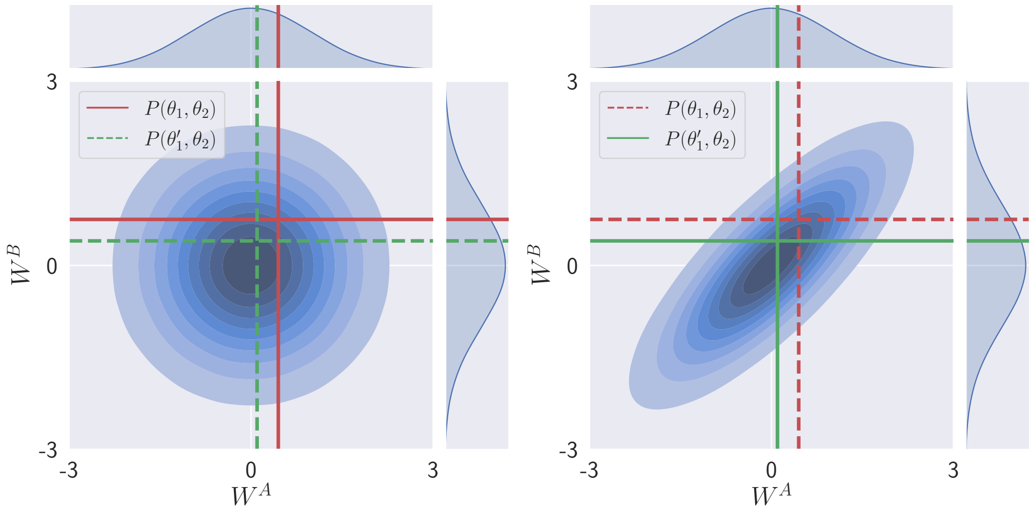


Fig. 3.3: Illustration the cutoff shift described in Lemma 3.8. The distribution of group G_i is a bivariate Gaussian with θ equal to the covariance: $\theta_1 = 0$ in the left-hand figure and $\theta_1' = 0.8$ in the right-hand figure. P^A is represented as a vertical line and P^B as an horizontal line. Both cutoffs decrease as θ_1 increases. In each sub-figure, the cutoffs corresponding to the current value of θ_1 are represented as full lines and the cutoffs corresponding to the other value of θ_1 as dashed lines.

Remark 3.2. The formal statement of Theorem 3.7 excludes the extremities of Θ. This assumption is made only to avoid the case where rankings are fully correlated, which would mean that $\mathbf{f}$ does not have full support. However, since V_1^A and V_1^B are continuous in θ, they are increasing on the whole interval Θ.

Theorem 3.7 allows us to derive the following corollary regarding a student's probability of remaining unmatched.

Corollary 3.9. *Suppose that $(H_\theta)_{\theta\in\Theta}$ is coherent and differentiable, assume $\theta \in \mathring{\Theta}^K$ and $\alpha^A + \alpha^B < 1$. Then the proportion of students from a given group remaining unmatched is increasing in its own correlation level and decreasing in the correlation level of all other groups: for $i, j \in [K]$, $i \neq j$,*

$$\frac{dV_\emptyset^{G_i}(\theta)}{d\theta_i} > 0 \quad \textit{and} \quad \frac{dV_\emptyset^{G_i}(\theta)}{d\theta_j} < 0.$$

Moreover, the inequality between two groups is decreasing in the correlation level of the group with lowest rate of unmatched students and increasing in the other group's correlation level. Formally, assume $V_\emptyset^{G_i}(\theta) < V_\emptyset^{G_j}(\theta)$. Then

$$\frac{dL^{G_i,G_j}(\theta)}{d\theta_i} < 0 \quad \textit{and} \quad \frac{dL^{G_i,G_j}(\theta)}{d\theta_j} > 0.$$

Proof. By Lemma 3.8, P^A and P^B are decreasing in both θ_1 and θ_2, thus for $i \neq j \in [K]$:

$$\begin{aligned}
\frac{dV_\emptyset^{G_j}}{\partial\theta_i} &= \frac{d\mathbb{P}_{j,\theta_j}(W^A < P^A, W^B < P^B)}{d\theta_i} \\
&= \left(\frac{\partial\,\mathbb{P}_{j,\theta_j}(W^A < P^A, W^B < P^B)}{dP^A}, \frac{\partial\,\mathbb{P}_{j,\theta_j}(W^A < P^A, W^B < P^B)}{dP^B}\right) \cdot \left(\frac{dP^A}{d\theta_i}, \frac{dP^B}{d\theta_i}\right)^T \\
&< 0.
\end{aligned}$$

Since the total capacity (of the two colleges) is constant, the mass of unmatched student must also be constant. Therefore, we have

$$\gamma_i V_\emptyset^{G_i} + \sum_{j\neq i} \gamma_j V_\emptyset^{G_j} = 1 - \alpha^A - \alpha^B.$$

By differentiating this equation we get

$$\begin{aligned}
&\gamma_i \frac{dV_\emptyset^{G_i}}{d\theta_i} + \sum_{j\neq i} \gamma_j \frac{dV_\emptyset^{G_j}}{d\theta_i} = 0 \\
\Leftrightarrow &\frac{dV_\emptyset^{G_i}}{d\theta_i} = -\frac{1}{\gamma_i} \sum_{j\neq i} \gamma_j \frac{dV_\emptyset^{G_j}}{d\theta_i} \\
\Rightarrow &\frac{dV_\emptyset^{G_i}}{d\theta_i} > 0
\end{aligned}$$

which proves the first part of Corollary 3.9. Moreover, since $L(G_i, G_j) = |V_\emptyset^{G_i} - V_\emptyset^{G_j}|$, assume that $V_\emptyset^{G_i} < V_\emptyset^{G_j}$, then $L(G_i, G_j) = V_\emptyset^{G_j} - V_\emptyset^{G_i}$, and using the first part of the result we get that $\frac{dL(G_i,G_j)}{d\theta_i} < 0$ and $\frac{dL(G_i,G_j)}{d\theta_j} > 0$, which proves the second part. ■

Remark 3.3. As a consequence, for any two groups G_i, G_j, if $\theta_i \neq \theta_j$ then the matching almost surely exhibits inequality for those groups ($L^{G_i,G_j}(\theta) > 0$). In particular, this is true even when those groups have the same marginals.

The probability of staying unmatched is different for students from different groups, even with identical marginals. This is in contrast to Proposition 3.5. Different levels of correlation lead to an unequal matching. This is the case as with identical marginals the proportion of students above some cutoff is the same in every group, but for a group with high correlation, the set of students above the cutoff is almost the same at each college, while for a group with low correlation those sets are quite different at each college. Therefore, the set of matched students in group G_j, which is $\{s \in G_j | W_s^A \geq P^A\} \bigcup \{s \in G_j | W_s^B \geq P^B\}$, is larger for groups with lower correlation. This result is quite counter-intuitive: consider the point of view of some college C, which has identical marginals for all groups. From C's point of view, there is no difference between the groups, and the proportion of students with $W_s^C \geq P^C$ is the same across all groups. However, Corollary 3.9 implies that the groups with the lowest correlation levels are going to be overrepresented at C, and the groups with the highest correlation underrepresented. College C then ends up with a set of student that could be deemed "unfair" regarding demographic parity, while C's ranking was in fact perfectly fair.

Corollary 3.9 helps understand the influence of correlation on inequality. When correlation levels of two groups are equal and marginals are identical, there is no inequality. If marginals are different, there is some "baseline" inequality that can be increased or decreased by changing the correlation levels: to decrease the inequality one would need to increase the correlation of the group with the lowest proportion of unassigned student (therefore the better-off group) and/or decrease the correlation of the worse-off group. Overall, even when colleges have fair rankings (identical marginals across all groups), the matching might still exhibit inequality.

Theorem 3.7 and Corollary 3.9 study efficiency and inequality separately. The following proposition describes their interaction. Concretely, different correlation vectors, θ, can lead to the same efficiency E, while inducing different inequality levels between groups. This shows that the effect of differential correlation cannot be solely explained via differences in efficiency.

Proposition 3.10. *Suppose that $(H_\theta)_{\theta\in\Theta}$ is coherent and differentiable, assume Θ, and $\alpha^A+\alpha^B < 1$. Let $\theta = (\theta_1, \ldots, \theta_K)$, and $\hat{E} = E(\theta)$.*

1. *There exist infinitely many correlation vectors achieving a given efficiency. Formally, the set of vectors θ' such that $E(\theta') = \hat{E}$ is a connected hypersurface of dimension K-1 (unless $\theta = (\underline{\Theta} \ldots, \underline{\Theta})$ or $(\bar{\Theta} \ldots, \bar{\Theta})$, in which case it is a singleton).*

2. *Fixing efficiency, correlation levels are substitutes. Formally, for any two groups G_i, G_j, there exists an interval $U := [\underline{\theta}, \bar{\theta}] \subseteq \Theta$ and a differentiable and decreasing function $\phi : U \to \Theta$ such that ($\theta_i \in U$ and $\theta_j = \phi(\theta_i)) \implies E(\theta) = \hat{E}$. The boundaries of U are optimal/pessimal with respect to the mass of unassigned students ($V_\emptyset^{G_i}, V_\emptyset^{G_j}$) for G_i respectively G_j and there is a unique $\hat{\theta} \in U$ such that $\theta_1 = \hat{\theta}, \theta_j = \phi(\hat{\theta})$ minimizes inequality $L^{G_i,G_j}(\theta)$.*

Proof. V_1 is a convex combination of the first choice functions that are increasing in all θ_j. Moreover it is continuous, and we assumed Θ to be an interval, so the set of possible values for V_1 is an interval, say $[V_1^{\min}, V_1^{\max}]$. Fix $V \in (V_1^{\min}, V_1^{\max})$, and consider the solutions of the equation $V(\theta) = V$. By continuity, this equation has a solution. The implicit function theorem applied to express some θ_j (the choice of j does not matter) as a function ϕ of all the other elements of θ shows that the solutions of $V(\theta) = V$ is a connected subset of Θ^K, and also a hypersurface because the function ϕ is monotonous in all θ_i (this comes from the fact that V_1 is itself monotonous). This proves the first part of the proposition.

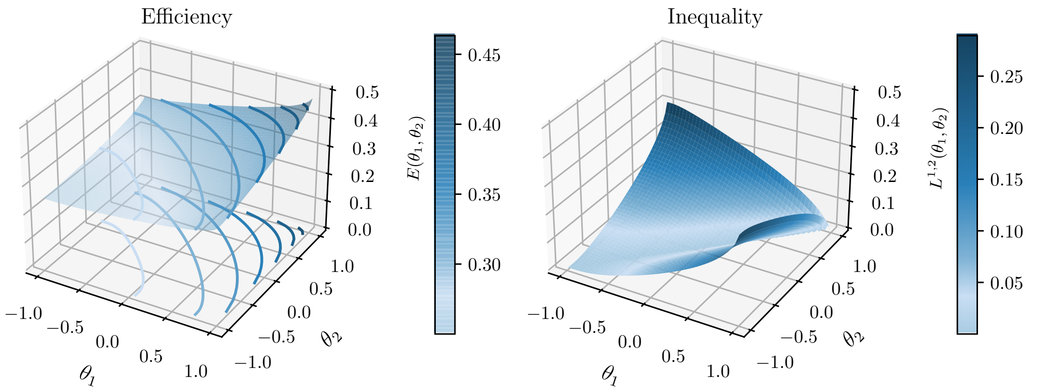


Fig. 3.4: Variations of efficiency (E) and (L) for two groups (G_1, G_2) as a function of their respective correlation parameters (θ_1, θ_2); with standard Gaussian marginals and copula with θ equal to the covariance. Other parameters: $\alpha^A = \alpha^B = 0.25, \gamma_1 = \gamma_2 = 0.5, \beta_1 = \beta_2 = 0.5$. Left:The surface represents the efficiency and the level lines indicate constant efficiency (also projected to the bottom of the figure). Right: The surface represents the inequality.

Let us choose two groups G_i, G_j, and fix all θ_ℓ for $\ell \neq i, j$. We apply the implicit function theorem to express θ_j as a function of θ_i, which shows that there exists an interval $U := [\underline{\theta}, \bar{\theta}] \subseteq \Theta$ and a differentiable function $\phi : U \to \Theta$ such that ($\theta_i \in U$ and $\theta_j = \phi(\theta_i)) \implies V_1(\theta) = V$. Since $V(\theta)$ is increasing in all arguments, ϕ is necessarily decreasing. If we keep $\theta_j = \phi(\theta_i)$, then $\frac{dV_\emptyset^{G_i}}{d\theta_i} = \frac{\partial V_\emptyset^{G_i}}{\partial \theta_i} + \frac{\partial V_\emptyset^{G_i}}{\partial \theta_j}\phi'(\theta_i)$, which is positive by Corollary 3.9, so $(\theta_i, \theta_j) = (\underline{\theta}, \phi(\underline{\theta})$ minimizes $V_\emptyset^{G_i}$, and $(\theta_i, \theta_j) = (\bar{\theta}, \phi(\bar{\theta})$ maximizes it. The same reasoning shows that those two points respectively maximize and minimize $V_\emptyset^{G_j}$. Finally, since $V_\emptyset^{G_i}$ is increasing and $V_\emptyset^{G_j}$ decreasing, $L(G_i, G_j) = |V_\emptyset^{G_i} - V_\emptyset^{G_j}|$ has a unique local (and therefore global) minimum. ■

Beyond the intuition that correlation favors efficiency, Proposition 3.10 provides a precise insight to the relation between efficiency and inequality and the trade-off between the two. The first part states that there is, in general, a continuum of correlation vectors achieving the same level of efficiency. The second part considers the comparative statics between two groups. Fixing the efficiency, correlation parameters behave as rival goods. As the correlation increases for one groups it necessarily decreases for the other group.

Figure 3.4 illustrates this for two groups, G_1, G_2, with standard Gaussian marginals and the parameter of the copula, θ, equal to the covariance. The left panel shows the efficiency and the right panel shows the inequality as functions of θ_1, θ_2. In the left panel, the level lines show the decreasing relation between θ_1 and θ_2 when E is kept constant. In the right panel, the inequality is minimized along the diagonal where $\theta_1 = \theta_2$ and increases as the parameters become more disparate.

3.4.2 Tie-Breaking

Some recent papers have studied the impact of tie-breaking rules on school choice problems, which has a strong link with correlation. In this section, we extend some of the prior results and discuss the relation to the literature.

Assume there is only one group, and each school C has n_C priority classes, i.e., there exists a partition of $S = Q_1^C \bigsqcup \cdots \bigsqcup Q_{n_C}^C$ such that for $i, j \in \{1, \ldots, n_C\}$, $s \in Q_i^C$, $s' \in Q_j^C$, s has higher priority than s' at C if $i < j$. Students belonging to the same priority class at a school are assumed to have the same priority at this school, but due to limited capacity the school might need to choose between them. To achieve this, schools use a random ranking of students to which they refer each time they need to choose between students from the same priority class; this random ranking is called a *tie-breaker*.

A natural question that has been actively studied in recent years is whether there is a difference in students' welfare if schools use the same tie-breaker (called single tie-breaker, or STB), instead of each producing an independent one (multiple tie-breakers, or MTB). [ANR19; AN20; Arn23] show — with slightly different models and assumptions (and among other results) — that when the total capacity of schools is lower than the number of students, then students are better off under STB than MTB. To ease the comparison, we restate their results here in a simplified form. Given n students and m schools:

- *[ANR19, Main Theorem]*: Suppose that there is capacity shortage, students' preferences are drawn uniformly at random and there is only one priority class (the whole ranking is random), then for any $k < m$, as the number of students and total capacity of colleges both grow to infinity the fraction of students matched to one of their top k choices approaches 0 under MTB but approaches a positive constant under STB.

- *[AN20, Theorem 3.2]*: Suppose there is one slot per school, only one priority class, and schools are divided into two tiers (top and bottom) with students' preferences inside a given tier drawn uniformly at random and a capacity shortage at top schools, then, with high probability, STB stochastically Pareto-dominates MTB.

- *[Arn23, Theorem 2]*: Suppose there is only one priority class and students only list $l < m$ schools in a uniform random order, then the number of students matched to their first choice is greater under STB than under MTB.

Our model, compared to prior work on tie-breaking, allows to for any number of priority classes, intermediate levels of correlation or even negative correlations, and several groups of students with different tie-breaking rules. To this end, let $(H_\theta)_{\theta \in \Theta}$ be a coherent and differentiable family of copulas such that $\theta = 0$ gives independent random variables, $\theta = 1$ fully correlated variables and $\theta = -1$ full negative correlation. Define the θ-TB as the tie-breaker drawn according to H_θ. Thus, MTB corresponds to $\theta = 0$ and STB to $\theta = 1$. Moreover, we can assume the existence of several groups with different θ.

Intermediate correlation levels can arise in tie-breaking if, for example, student characteristics are introduced into rankings to break ties, e.g., sibling priority or distance to from home to school [Cor+22]. This is commonly done to render algorithms more deterministic and thus understandable. Consider priority for students with lower distance to a school and suppose that there are two villages with one school each. Then, ceteris paribus, a student who lives in one village exhibits negative correlation between the grades at each of the two schools. On the other hand, a student living in neither village may exhibit any level of correlation. Note that this example also illustrates how negative correlation naturally arises.

Proposition 3.11. *Let there be a continuum mass of students and assume that students prefer any school over being unmatched. Let A, B be two schools with n_A, n_B priority classes, and constrained capacities $\alpha^A + \alpha^B < 1$. Further suppose that students are divided into K groups, such that the θ_j-TB is used for group G_j. Then:*

1. *The mass of students getting their first choice*
    - *is non-decreasing in each θ_j,*
    - *is almost surely strictly increasing in all θ_j, if all products of priority classes $Q_i^A \times Q_j^B$ contain a positive mass of students of each group,[8] and*
    - *is strictly increasing in each θ_j, if there is only one priority class.*
2. *The inequality between two groups, $L^{G_i,G_j}(\theta)$, is non-decreasing in the correlation θ of the group with the lowest $V_\emptyset$, and non-increasing in the other group's θ.*

Proof sketch. We build a distribution family that encompasses the priorities of students at each school taking into account priority classes as well as tie-breakers, such that MTB and STB correspond to values of $\theta = 0$ and $\theta = 1$. The obtained distribution, while being complex, still satisfies most of the assumptions required by our model, and with some adjustments we are able to apply Theorem 3.7 and Corollary 3.9 and conclude. The proof is provided in Appendix 3.8.3. ■

This result shows that increasing the correlation of tie-breakers, for one or several groups, increases the amount of students getting their first choice. Moreover, it also shows that a policy maker able

[8] More precisely, the set of vectors (γ, β, α) such that the mass of students getting their first choice is constant in some θ_j has Lebesgue measure 0.

to change the correlation of tie-breakers for some groups can use it to mitigate the inequalities between groups.

Proposition 3.11 is in some regards more restrictive than the results from the literature presented above, because it only applies to two schools and assumes students prefer either school over being unmatched. On the other hand, it is more general in that it applies to cases with several priority classes and does not require students preferences to be uniform (in our model, we can have any fraction β of students preferring school A). It also allows to have several groups with different tie-breaking rules. Finally, Proposition 3.11 allows for intermediate tie-breaking rules interpolating between MTB and STB, and also for negatively correlated tie-breaking rules.

Remark 3.4. The reason for Theorem 3.7 not completely applying, which explains the use of "almost surely" in Proposition 3.11, is that in the case where a cutoff falls exactly in between two priority classes, then there is no need to break ties at this school, and the correlation of the tie-breaker naturally does not play a role anymore. We consider that this is almost surely not the case in the sense that the set of capacities at each school and preferences of students that would lead to this situation has zero mass in the space of all possible values for those parameters.

3.5 Extension to more than two colleges

In this section, we discuss the extension of our results to more than two colleges.

3.5.1 Model

We start by proposing an extension of the model for more than two colleges, since it is not straightforward and choices have to be made.

The set of colleges becomes $\mathbf{C} = \{C_1, \ldots, C_m\}$. Colleges have respective capacities $\alpha_1, \ldots, \alpha_m \in (0,1]$, we call α the vector containing colleges capacities. Each student has strict preferences over colleges: when student s prefers college C to college C', we write $C \succ_s C'$. This preference list can also be represented by a permutation $\sigma_s \in \Sigma([m])$, where $\sigma(1)$ is the favorite college and $\sigma(m)$ the least favorite. For any σ, the amount of students with preference list σ among group G_j is $\beta^{\sigma}_{G_j} \in (0,1)$. We call $\boldsymbol{\beta}_{G_j}$ the vector $(\beta^{\sigma}_{G_j})_{\sigma \in \Sigma([m])}$. Each college C_i assigns a priority W^i score to each student, the higher the better. Each student s then has vector of priority scores $(W^1_s, \ldots, W^m_s)$. This means that college C_i prefers $s \in S$ to s' if and only if $W^i_s > W^i_{s'}$.

Regarding the distribution of grades, the model has a natural extension. The marginal distribution of scores W^i_s given by college C_i to students s in G_j is described by a pdf f^i_j defined over the support $I_j^i \subseteq \mathbb{R}$, assumed to be an interval. Let $I_j = \prod_{i\in[m]} I_j^i$. We denote by $\underline{I}_j^i$ and $\bar{I}_j^i$ the lower and upper bounds of I_j^i, and by F_j^i the cdf associated to f_j^i.

As before, each group G_j then has a joint distribution with joint pdf f_j and cdf F_j, that can be represented by its marginals $(F_j^i)_{i\in[m]}$ and a (unique) copula H_j. We assume that there exists a family of m-dimensional copulas $(H_\theta)_{\theta\in\Theta}$ and, for all $j \in [K]$, there exists a parameter $\theta_j \in \Theta$ such that H_{θ_j} is the copula associated to G_j's distribution, i.e., $H_{\theta_j} = H_j$. The coherence assumption remains the same: we say that $(H_\theta)_{\theta\in\Theta}$ is coherent if for all $x \in (0,1)^m$, $H_\theta(x)$ is strictly increasing in θ.

The supply and demand framework does not change. Given cutoffs $(P^1, \ldots, P^m) = \mathbf{P}$, student s demand $D_s(\mathbf{P})$ is the college they prefer among those where they pass the cutoff. The demand $D^i(\mathbf{P})$ at college C_i is the mass of students who demand C_i. The unique vector of market clearing cutoffs is the solution of $\mathbf{D}(\mathbf{P}) = \alpha$ when $\sum_{i\in[m]} \alpha^i < 1$. When $\sum_{i\in[m]} \alpha^i \geq 1$, the market clearing cutoffs are given bu $D^i(\mathbf{P}) = \alpha^i$ for full colleges and $D^i(\mathbf{P}) = \underline{I}_j^i$ for colleges where empty seats remain.

3.5.2 Results

Turning to our results, Proposition 3.5 remains true. That is, whenever two groups have the same marginal at some college, they have the same probability of getting this college as a first choice. On the other hand, Proposition 3.6 does not hold anymore: for any pair of colleges whose joint capacity is less than 1, correlation between their rankings will have an effect on the matching, even if the total capacity is more than 1. We can still derive a weaker version.

Proposition 3.12. *If for all pair of colleges $i, j \in [m]$ we have $\alpha^i + \alpha^j \geq 1$, then correlation has no effect on the stable matching. The cutoffs P^A and P^B are constant in θ, on then so are $V_1^{G_j,C}$ and $V_2^{G_j,C}$ for all j and C. Moreover, $V_\emptyset^{G_j} = 0$, therefore $\forall i,j \in [K]$, $L^{G_i,G_j}(\theta) = 0$.*

Proof. With this assumption, students get either their first or second choice, i.e., they cannot get a worst choice or remain unassigned. From there, the reasoning used in the original proof applies. ■

The other results all rely on Lemma 3.8. We found, through numerical experiments, that Lemma 3.8 generally does not extend beyond two colleges. See Figure 3.5 for a counterexample with four colleges, one group, standard Gaussian marginals at every college, a Gaussian copula with the same covariance θ at each pair of colleges, and $\alpha = (0.05, 0.05, 0.2, 0.5)$. The cutoff P^C of the third college is increasing for high values of θ.

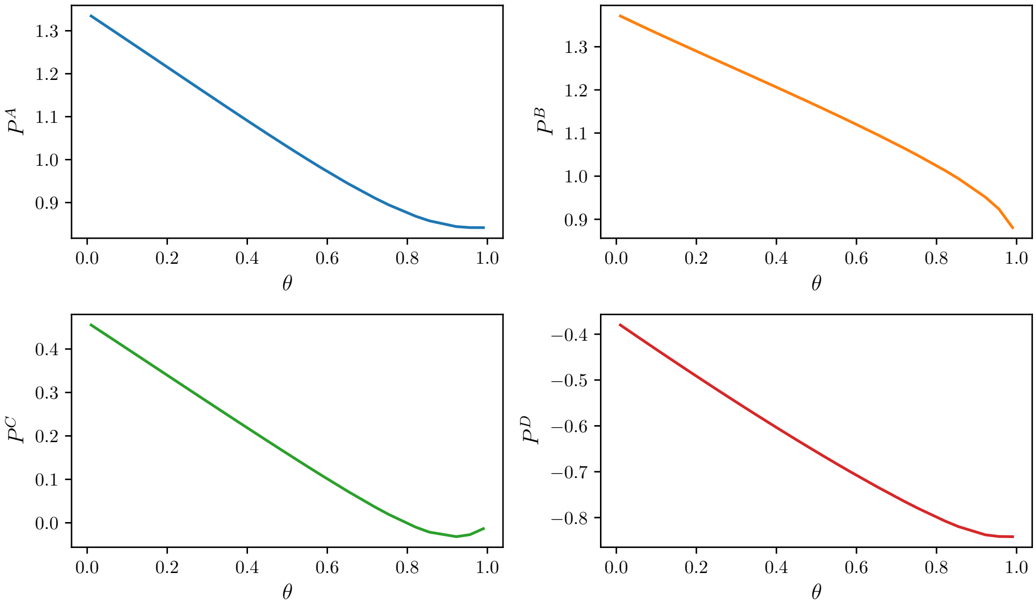


Fig. 3.5: Counterexample to the extension of Lemma 3.8 to more than two colleges: Four colleges, one group, multivariate Gaussian distribution with variance 1 at each college and covariance θ on all pairs of colleges, and $\alpha = (0.05, 0.05, 0.2, 0.5)$. Note that P^C is not monotone decreasing.

To understand more clearly the reason for Lemma 3.8 not being true with more than two colleges, we can try to prove it the same way we did in the two colleges case. The details are provided in Appendix 3.8.4. Every part of the proof works except for one point: with two colleges, terms of the form $\mathbb{P}(W^A \geq P^A, W^B < P^B)$ were decreasing in θ due to the coherence assumption, but for instance with three colleges terms of the type $\mathbb{P}(W^1 \geq P^1, W^2 < P^2, W^3 < P^3)$ are not necessarily monotonic in θ even with the coherence assumption. To make the proof work we could add the assumption that all quantities of the form presented above must be monotonic in θ, but it would be an extremely restrictive assumption, as we do not know any classical copula family that satisfies it.

Under some additional assumptions, those results could be recovered, however they are quite restrictive too. For instance, if we assume that students preferences are drawn uniformly at random and all colleges have the same capacity, as in Peng and Garg [PG23], then all cutoffs are equal (or at least increasing functions of each other, if colleges use different marginals) and we recover Lemma 3.8 and all subsequent results.

3.6 Special Cases

In our model, the exact solutions of the market-clearing equation, and thus the metrics, usually do not admit closed-form expressions. In this section, we focus on some notable special cases for which these calculations are possible and allow us to have a quantitative view of the effects of correlation. In particular, since Theorem 3.7 and 3.9 state that the metrics are monotonic, computing them for correlation levels of -1, 0 and 1 provides bounds for all correlation values in between.

3.6.1 Excess of capacity

In the case where $\alpha^A + \alpha^B \geq 1$, the market clearing equation no longer allows to compute the stable matching. In fact, there might even exist several stable matchings. We can find one by following the steps of the deferred acceptance algorithm (see Algorithm 7 in the Appendix). We consider three (partitioning) cases:

(1) *There is not enough room in college A for all students preferring it to college B, i.e.,* $\sum_j \gamma_j \beta_j \geq \alpha^A$. In this case, there is necessarily enough room in college B for all students preferring it, since $\alpha^A + \alpha^B \geq 1$. Therefore, following the steps of DA, we find:

(i) At step one, $\sum_j \gamma_j \beta_j$ students preferring A apply there and the best α^A are temporarily admitted, and $\sum_j \gamma_j (1-\beta_j)$ students preferring B apply there and are all temporarily admitted.

(ii) At step two, the $\sum_j \gamma_j \beta_j - \alpha^A$ students rejected from A apply to B, and are admitted since there is enough room for them (considering the students previously admitted).

This results in the following probabilities of a student to get their first or second choice:

$$V_1^{G_j,A} = 1 - F_j^A(P^A), \qquad V_1^{G_j,B} = 1,$$
$$V_2^{G_j,A} = F_j^A(P^A), \qquad V_2^{G_j,B} = 0,$$
$$P^A = \left(\sum_j \gamma_j \beta_j \left(1 - F_j^A\right)\right)^{-1} (\alpha^A).$$

Finally, as every student is admitted somewhere, $V_\emptyset^{G_j} = 0$.

(2) *There is not enough room in college B for all students preferring it to A, i.e.,* $\sum_j \gamma_j (1-\beta_j) \geq \alpha^B$. Symmetrically we get

$$V_1^{G_j,A} = 1, \qquad V_1^{G_j,B} = 1 - F_j^B(P^B),$$
$$V_2^{G_j,A} = 0, \qquad V_2^{G_j,B} = F_j^B(P^B),$$
$$P^B = \left(\sum_j \gamma_j (1-\beta_j) \left(1 - F_j^B\right)\right)^{-1} (\alpha^B), \qquad V_\emptyset^{G_j} = 0.$$

(3) *There is enough room in each college to admit all students who prefer attending it, i.e.,* $\sum_j \gamma_j \beta_j \leq \alpha^A$ *and* $\sum_j \gamma_j (1-\beta_j) \leq \alpha^B$. It follows that everyone gets their first choice: for $j \in [K]$ and $C \in \{A, B\}$,

$$V_1^{G_j,C} = 1,$$
$$V_2^{G_j,C} = V_\emptyset^{G_j} = 0.$$

3.6.2 One group

Suppose that there is only one group of students (i.e., $\gamma_1 = 1$) and therefore all students have the same correlation parameter θ. Since there is only one group, there is only one parameter β for the proportion of students preferring A, and the metrics V_1, V_2 and $V_\emptyset$ do not depend on the group. In this section we first treat the case of capacity excess in more detail using this additional assumption, then we consider three specific cases, where the colleges have either full correlation (they use the same ranking of students), no correlation at all (their rankings are statistically independent), or anti-correlation (the ranking at one college is the prefect opposite of the other). These special cases will allow us to understand the matching's dependencies on the capacities and preferences of students.

Excess of capacity

First, when $\alpha^A + \alpha^B \geq 1$, we can directly apply the results found in the previous section. First, as before we have $V_\emptyset = 0$. Then,

- if $\beta \geq \alpha^A$ (college A is over-demanded):

$$V_1^A = \frac{\alpha^A}{\beta}, \qquad V_1^B = 1,$$
$$V_2^A = 1 - \frac{\alpha^A}{\beta}, \qquad V_2^B = 0.$$

- if $\beta \leq 1 - \alpha^B$ (college B is over-demanded):

$$V_1^A = 1, \qquad V_1^B = \frac{\alpha^B}{1-\beta},$$
$$V_2^A = 0, \qquad V_2^B = 1 - \frac{\alpha^B}{1-\beta}.$$

- if $1-\alpha^B \leq \beta \leq \alpha^A$ (both colleges are under-demanded):

$$V_1^A = 1, \qquad V_1^B = 1,$$
$$V_2^A = 0, \qquad V_2^B = 0.$$

For the three following cases (full correlation, full independence and anti-correlation) we assume that $\alpha^A + \alpha^B < 1$.

Full correlation

We first study the case where students have the same rank in both colleges.

Proposition 3.13. *When both colleges use the same ranking, the metrics $V_1, V_2, V_\emptyset$ can be computed exactly, and their expressions are:*

- *if $\beta \leq \frac{\alpha^A}{\alpha^A+\alpha^B}$ (college A is under-demanded):*

$$V_1^A = \alpha^A + \alpha^B, \qquad V_1^B = \frac{\alpha^B}{1-\beta},$$
$$V_2^A = 0, \qquad V_2^B = \alpha^A - \frac{\beta}{1-\beta}\alpha^B,$$
$$V_\emptyset = 1 - \alpha^A - \alpha^B$$

- *if $\beta \geq \frac{\alpha^A}{\alpha^A+\alpha^B}$ (college A is over-demanded):*

$$V_1^A = \frac{\alpha^A}{\beta}, \qquad V_1^B = \alpha^A + \alpha^B,$$
$$V_2^A = \alpha^B - \frac{1-\beta}{\beta}\alpha^A, \qquad V_2^B = 0,$$
$$V_\emptyset = 1 - \alpha^A - \alpha^B$$

Proof sketch. The proof amounts to solving the market-clearing equation, the details are provided in Appendix 3.8.3. ■

This result is illustrated by the blue and orange lines in Figure 3.6. For the top row, $\beta = 0.3$, and V_1^A and V_1^B are computed for $\alpha^A = \alpha^B := \alpha$ varying from 0 to 1. The probability of getting one's first choice is increasing in the capacity. Students preferring college A get either their first choice or nothing, as shown in Proposition 3.13. Indeed, college A is easier to get in than college B since $\beta \leq \frac{\alpha^A}{\alpha^A+\alpha^B}$, so a student rejected from college A is necessarily rejected from college B. For $\alpha > 0.5$, Proposition 3.13 does not apply anymore because there is capacity excess, and we need to refer to Section 3.6.1. For the bottom row, $\alpha^A = \alpha^B = 0.25$ and V_1^A and V_1^B are shown for β varying from 0 to 1. The two figures are mirrored images of each other, which is natural as the problem is symmetric in A and B. Observe that students who prefer the least popular college have a higher probability of getting their first choice (the blue line is higher on the left plot than on the right for $\beta < 0.5$, and lower for $\beta > 0.5$).

Remark 3.5. The full correlation case does not fit our model's assumptions, since the distribution does not have full support. The consequence is that uniqueness of the stable matching is not guaranteed. However, by solving the market clearing equation, we showed that there is indeed only one solution, so the stable matching is still unique.

Full independence

Now, consider the case where the ranks of a student at A and B are not correlated at all.

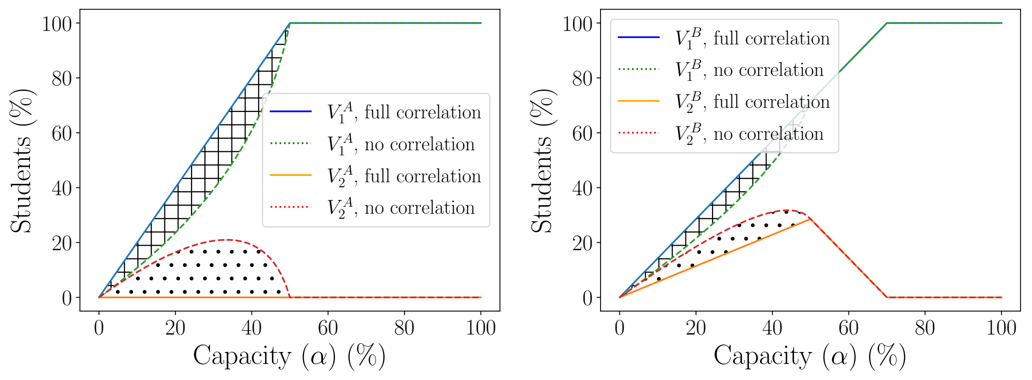
Proposition 3.14. *When colleges' rankings are independent and $\alpha^A + \alpha^B < 1$, the metrics $V_1, V_2, V_\emptyset$ can be computed exactly, and their expressions are:*

$$
\begin{aligned}
V_1^A &= 1 - \frac{1-\beta}{2\beta}(\Delta - \zeta), & V_1^B &= 1 - \frac{1}{2}(\Delta + \zeta), \\
V_2^A &= \frac{1-\beta}{2\beta}(\Delta - \zeta) - \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), & V_2^B &= \frac{1}{2}(\Delta + \zeta) - \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), \\
V_\emptyset &= 1 - \alpha^A - \alpha^B
\end{aligned}
$$

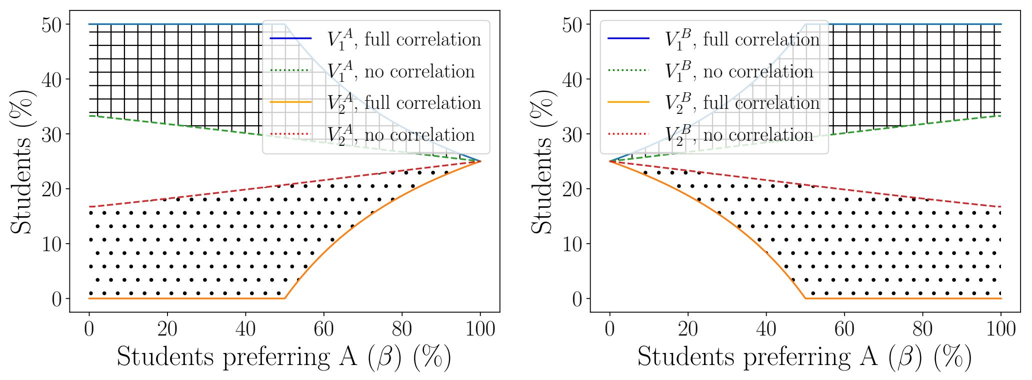
with $\zeta = \frac{1-2\beta}{1-\beta} + \frac{\beta}{1-\beta}\alpha^A - \alpha^B$ *and* $\Delta = \sqrt{\zeta^2 + \frac{4\beta}{1-\beta}(1 - \alpha^A - \alpha^B)}$.

Proof sketch. Again, the proof amounts to solving the market-clearing equation, the details are provided in Appendix 3.8.3. ■

These results are illustrated by the dashed green and red lines in Figure 3.6. For the top row, once again, Proposition 3.14 applies only for $\alpha < 0.5$. Note that V_2^A and V_2^B are both strictly positive, because even though college A is easier to get in than college B, students rejected from college A still have an independent second chance at B.



(a) V_1^A and V_2^A as a function of α, $\beta = 0.3$ **(b)** V_1^B and V_2^B as a function of α, $\beta = 0.3$



(c) V_1^A and V_2^A as a function of β, $\alpha = 0.25$ **(d)** V_1^B and V_2^B as a function of β, $\alpha = 0.25$

Fig. 3.6: Proportion of students getting their first and second choice with one group, and $\alpha^A = \alpha^B = \alpha$. The solid blue and orange lines are for the full-correlation case, the dashed green and red ones for the no-correlation case. As the blue and green lines represent the same metric in two different settings, the gap between them is hashed to highlight the welfare increase when switching from one to the other. The same applies to the orange and red lines, where the gap is dotted.

Comparing the blue and green lines, note that the amount of students getting their first choice is always larger with full correlation. This follows from Theorem 3.7 as V_1 is increasing in the correlation, for any values of the parameters. Correspondingly, the amount of students getting their second choice is always lower with full correlation. For values of θ for which the correlation is between 0 and 1, it is generally not possible to obtain closed-form expressions for these metrics. However, Theorem 3.7 implies that their graphs have to be contained between the two extreme cases, i.e., in the colored area highlighted in Figure 3.6. These areas show the extent of the influence of correlation on students' welfare. It appears that for intermediate values of the problem's fundamentals, the gap is substantial. For instance, looking at the bottom row, we see that for $\beta = 0.5$ and $\alpha^A = \alpha^B = 0.25$, increasing the correlation can make the number of students getting their first choice grow from 30% to 50%.

Anti-correlation

We study the case where the ranking of one college is the exact opposite of the other. Once again, we can compute explicitly the metrics.

Proposition 3.15. *When colleges' rankings are opposed and* $\alpha^A + \alpha^B < 1$*, the metrics* $V_1, V_2, V_\emptyset$ *can be computed exactly, and their expressions are:*

$$V_1^A = \alpha^A, \qquad V_1^B = \alpha^B,$$
$$V_2^A = \alpha^B, \qquad V_2^B = \alpha^A,$$
$$V_\emptyset = 1 - \alpha^A - \alpha^B$$

Proof. We assume WLOG that the marginals are uniform on $[0, 1]$. Then, since correlation is -1, we have for every student $W^A = 1 - W^B$. Then, the market clearing equation immediately gives $P^A = 1 - \alpha^A$ and $P^B = 1 - \alpha^B$, which gives $V_1^A = V_2^B = \alpha^A$ and $V_1^B = VDA = \alpha^B$. ■

As in the full correlation case, the distribution of grades does not have full support, and therefore the stable matching might not be unique, but by solving the market clearing equation we prove that it in fact is.

Noisy priorities

We can use our results to study a model where the priorities are composed of each student's latent quality and some additive noise, as we presented in Section 1.2.3. This model is strictly more complex than our base model, so it will not make the computation of solutions easier in the general case. However, we can use the results obtained above to study the special case where the noise has infinite variance. This case is actually equivalent to the case with no correlation that we considered above, but the fact priorities are noisy gives rise to new interesting metrics, such as the number of students having justified envy with respect to the true priorities (i.e., the latent qualities), or the number who got a worse (or better) outcome because of the noise.

We start by computing the amount of students having justified envy. A student s has justified envy if, in the matching obtained with noisy priorities, some student with a lower latent quality was admitted to a college that s would have preferred to their current college.

First, notice that all students that get their first choice cannot have justified envy since they do not envy anyone. Second, consider a student s who did not get their first choice, college C. If some student with a lower latent quality than them is admitted to C, then s has justified envy. However, in the continuum model, an infinity of students have a lower latent quality than s, and the probability that one of them is admitted to C is 1. Therefore, the number of student with justified envy is the number of students who do not get their first choice $1 - E$.

Proposition 3.16. *The amount of students with justified envy is:*

- *if* $\alpha^A + \alpha^B \geq 1$*:*
  - $1 - \beta - \alpha^B$ *for* $\beta \leq 1 - \alpha^B$,
  - 0 *for* $1 - \beta \leq \beta \leq \alpha^A$,
  - $\beta - \alpha^A$ *for* $\beta \geq \alpha^A$.
- *if* $\alpha^A + \alpha^B < 1$*:* $(1 - \beta)\Delta$,

where $\Delta = \sqrt{\zeta^2 + \frac{4\beta}{1-\beta}(1 - \alpha^A - \alpha^B)}$ *as before.*

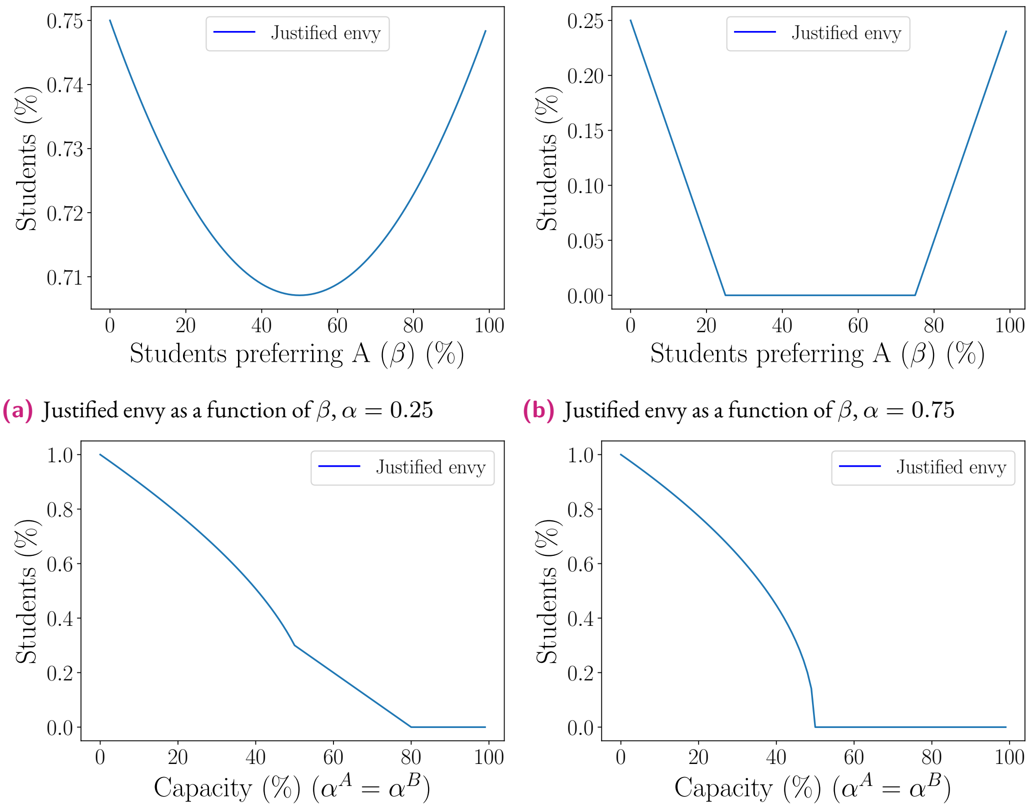
Proof. The amount of students with justified envy is $1-E = 1-\beta V_1^A-(1-\beta)V_1^B$. From there, we only have to input the expressions for V_1^A and VB found above (in the case with correlation 0, since infinite noise is equivalent to no correlation). ■

Those expressions are plotted on Figure 3.7. We can notice on the top row that when capacity is low justified envy is not very sensible to students' preferences, but it is much more sensible to it when capacity increases. Bottom row confirms the natural intuition that when capacity increases, justified envy decreases as more students can have their first choice independently of the noise.

We now turn to computing the amount of students benefiting or losing from the noise. We consider that a student is losing (resp. benefiting) from the noise if they get a worse (resp. better) outcome in the matching with noise than they would have gotten with the latent qualities. With infinite noise, the outcome of a student are totally independent of their latent quality, which makes the computation of the computation of the mass of students in each situation easier.

Proposition 3.17. *When* $\alpha^A + \alpha^B < 1$*, there are always more students losing from the noise than benefiting from it.*

We explain below how to compute the two quantities, which provides an immediate proof for the proposition. A student is losing from the noise if they get their second choice when they



(a) Justified envy as a function of β, $\alpha = 0.25$

(b) Justified envy as a function of β, $\alpha = 0.75$

(c) Justified envy as a function of α, $\beta = 0.2$

(d) Justified envy as a function of α, $\beta = 0.5$

Fig. 3.7: Proportion of students having justified envy with one group, infinite noise on the priorities, and $\alpha^A = \alpha^B = \alpha$. On the top row capacity is fixed and β varies, on the bottom row it is the opposite.

should have gotten their first, or if they remain unassigned when they should have been admitted somewhere. This quantity is then equal to

$$\beta V_1^A(1)V_2^A(0) + (1-\beta)V_1^B(1)V_2^B(0) + V_\emptyset(1 - V_\emptyset),$$

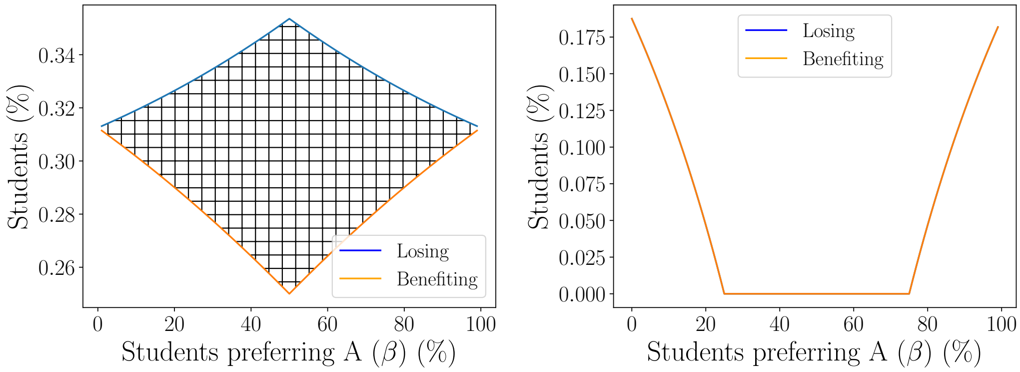
where (0) or (1) as argument of the metrics refers to the value of those metrics with no correlation of with full correlation respectively, and is omitted for $V_\emptyset$ since they are equal. Conversely, the mass of students benefiting from the noise is

$$\beta V_1^A(0)V_2^A(1) + (1-\beta)V_1^B(0)V_2^B(1) + V_\emptyset(1 - V_\emptyset).$$

If we subtract the second expression from the first, we get that

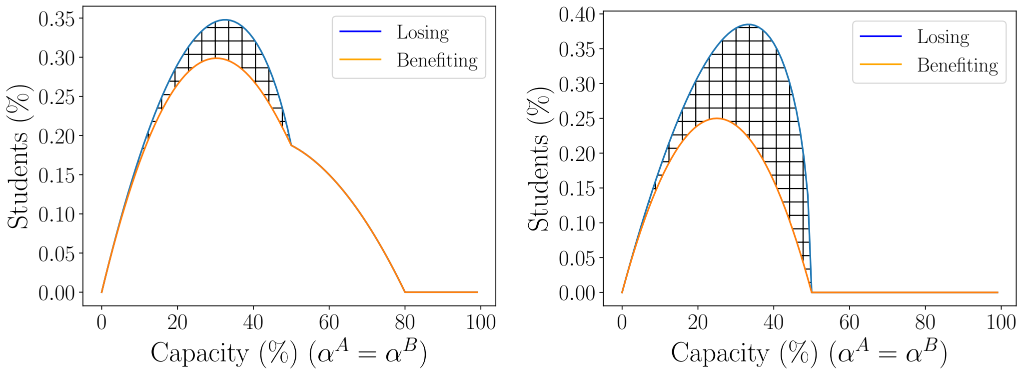
$$\beta\left(V_1^A(1)V_2^A(0) - V_1^A(0)V_2^A(1)\right) + (1-\beta)\left(V_1^B(1)V_2^B(0) - V_1^B(0)V_2^B(1)\right)$$

more students are losing from the noise than benefiting. Since $V_1^A(1) \geq V_1^A(0)$ and $V_2^A(1) \leq V_2^A(0)$, and same for B, this quantity is always positive. We can notice from the expression that when $\alpha^A + \alpha^B \geq 1$, both metrics are equal and thus the difference is zero, since V_1^A, V_1^B, V_2^A and V_2^B are independent from correlation. While it was expected that making priorities random would help some students and hurt others, it is quite interesting and unexpected that it always hurts more students than it helps.



(a) Students losing and benefiting from the noise as a function of β, $\alpha = 0.25$

(b) Students losing and benefiting from the noise as a function of β, $\alpha = 0.75$



(c) Students losing and benefiting from the noise as a function of α, $\beta = 0.2$

(d) Students losing and benefiting from the noise as a function of α, $\beta = 0.5$

Fig. 3.8: Proportion of students losing and benefiting from the noise with one group, infinite noise on the priorities, and $\alpha^A = \alpha^B = \alpha$. On the top row capacity is fixed and β varies, on the bottom row it is the opposite. The hashed areas highlight the difference between the two metrics.

Figure 3.8 illustrates the amount of students losing and benefiting from the noise. We notice that, as stated in Proposition 3.17, when $\alpha^A + \alpha^B < 1$ there are always more students losing from the noise than there are benefiting from it. This difference is highlighted by the hashed areas. The difference is largest when $\beta = \frac{1}{2}$. Here is an interpretation: when all students prefer the same college, if noise hurts a student, it means that some other student gets their seat and benefits from it, so the two metrics are equal. When students have different preferences, if noise hurts a student, the student who gets their seat might also be losing from the noise, which explains that noise hurts more students than it helps, and that this effect is maximized when $\beta = \frac{1}{2}$. As stated above, when $\alpha^A + \alpha^B \geq 1$, both metrics are equal.

3.7 Discussion

We have introduced a tractable model to study the impact of differential correlation between different groups and studied its effect on on outcome inequality and efficiency in matching markets. Our framework is general in that it accommodates almost any grade distribution, any number of groups with different distributions and different student preferences, and colleges of any capacity.

However, a limitation is our focus on two colleges. We showed in Section 3.5 that Lemma 3.8 does not generally stand with more than two colleges. It follows that Theorem 3.7 also does not extend beyond two colleges as it entirely relies on Lemma 3.8. The other results (Corollary 3.9, and Propositions 3.10 and 3.11) may remain true, but different proof techniques would be required. Our experiments suggest that the statement of Lemma 3.8 may still be true for large classes of correlation, thus suggesting the possibility to extend our results under some additional assumptions.

To conclude, we believe that there is ample scope to study themes that have already been considered in the single decision-maker settings in the matching context. Our analysis suggests that in matching new phenomena arise and it is important to further understand them. In models where decision-makers use noisy estimates of applicants' latent quality, existing results about algorithmic monoculture could be extended, providing a theoretical foundation to experimental results such as those of Bommasani et al. [Bom+22]. Using the noise structure, interesting variations could be allowing applicants to invest in accurate assessment, e.g., via acquiring certifications or doing in-person interviews, or considering the effects of risk aversion. Other possible directions could include making applications costly, or allowing applicants to not list all colleges in their preferences, and a more thorough study of colleges' utility.

3.8 Appendices

3.8.1 Notation

Table 3.1 provides a summary of the notation used throughout this chapter.

3.8.2 Definitions and technical details

Tab. 3.1: Notation for Chapter 3

Agents:	
A, B	Colleges (generic: C)
s	An arbitrary student
S	Students set
$G_1, \dots G_K$	Groups of students, partition of S
η	Measure for student masses
Agents' features:	
α^A, α^B	Colleges' capacities ($\in (0,1)$)
γ_j	Mass of students in group G_j ($\in [0,1]$)
β_j	Share of students in group G_j preferring college A ($\in [0,1]$)
Priority scores:	
W_s^C	Score at C of student s (generic: W)
f_j^C, F_j^C	Marginal pdf and cdf of college C for group G_j
$(H_\theta)_{\theta\in\Theta}, (h_\theta)_{\theta\in\Theta}$	Copula family and associated pdfs, indexed by θ
θ	Parameter for a copula family
$f_{j,\theta_j}, F_{j,\theta_j}$	Group G_j's score vectors' joint pdf and cdf, $F_{j,\theta_j} = H_{\theta_j}(F_j^A, F_j^B)$
Θ	Set of possible values for θ
I_j^C, I_j	Support of f_j^C and f_{j,θ_j} respectively. $I_j = I_j^A \times I_j^B$
$\underline{I}_j^C, \bar{I}_j^C$	Lower and upper bounds of I_j^C
Correlation:	
r	Pearson's correlation
ρ	Spearman's correlation
τ	Kendall's correlation
Matching:	
μ	Matching
$V_1^{G_j,C}$	Share of students of group G_j and preferring C who get their first choice
$V_2^{G_j,C}$	Share of students of group G_j and preferring C who get their second choice
$V_\emptyset^{G_j}$	Share of students of group G who are unassigned
V_1	Total mass of students getting their first choice
$L^{G_i,G_j}(\theta)$	Inequality between G_i and G_j, equal to $\lvert V_\emptyset^{G_i}(\theta) - V_\emptyset^{G_j}(\theta)\rvert$

Formal definition of the mass η

Here we formally define the notion of mass for a subset of students. This section is not necessary to understand the results of the chapter; the notations introduced here are not used elsewhere. We identify S to $\Sigma := \mathbb{R}^2 \times \{G1, \dots, G_K\} \times \{A, B\}$. We partition Σ into several subsets: $\Sigma_{G_j,C} := \{s \in \Sigma : s = ((x,y), G_j, C), x, y \in \mathbb{R}\}$ is the subset of students belonging to group G and preferring college C. Given a vector of parameters θ and priorities W^A, W^B distributed according to f_{j,θ_j} for G_j students, we say that a subset $J \subseteq \Sigma$ is measurable if and only if $\{(W_s^A, W_s^B) : s \in J\}$ is Borel-measurable in $\mathbb{R}^2$. We can partition J into subsets $J_{G_j,C} := J \cap \Sigma_{G_j,C}$. On each $\Sigma_{G_j,C}$ we define a measure $\eta_{G_j,C}$ as follows: for $J \subseteq \Sigma$ measurable,

$$\begin{aligned} \eta_{G_j,A}(J_{G_j,A}) &= \gamma_j \beta_j \mathbb{P}_{\theta_j}((W^A, W^B) \in \{(W_s^A, W_s^B) : s \in J_{G_j,A}\}), \\ \eta_{G_j,B}(J_{G_j,B}) &= \gamma_j (1-\beta_j) \mathbb{P}_{\theta_j}((W^A, W^B) \in \{(W_s^A, W_s^B) : s \in J_{G_j,B}\}), \end{aligned} \tag{3.7}$$

Let $\mathcal{B}(S)$ be the set of measurable subsets of S. We define over $\mathcal{B}(S)$ the probability measure $\eta : \mathcal{B}(S) \to [0, 1]$ such that for any measurable subset J of S,

$$\eta(J) = \sum_{j \in [K]} \eta_{G_j,A}(J_{G_j,A}) + \eta_{G_j,B}(J_{G_j,B}). \tag{3.8}$$

This definition is consistent with the definition of the parameters, as it verifies $\eta(G_j) = \gamma_j$, $\eta(\{s \in G_j : A \succ_s B\}) = \gamma_j \beta_j$ and so on.

Discussion on distributional assumptions

We assume that distributions admit a density and have full support, and that they can be represented using a copula family and marginals that remain the same for any θ, and that this copula is coherent and differentiable. We here explain why these assumptions are not very restrictive by presenting canonical examples of classical copulas satisfying our assumptions.

1. *Gaussian copula*: The Gaussian copula is obtained by composing the cdf Φ_θ of a bivariate Gaussian with covariance matrix $\begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$ and the univariate cdf ϕ of the standard Gaussian: $H_\theta(x, y) = \Phi_\theta(\phi(x), \phi(y))$. Here, the parameter θ controls the covariance.

2. *Archimedean copulas*: Archimedean copulas are a broad family of copulas, each element of this family being itself a parametric family of copulas with parameter θ. The general formula is
$$H_\theta(x, y) = \psi_\theta^{-1}\left(\psi_\theta(x) + \psi_\theta(y)\right)$$
where $\psi_\theta : [0, 1] \to \mathbb{R}_+$ is a continuous strictly decreasing and convex function such that $\psi_\theta(1) = 0$. Examples include:

   - Clayton: $H_\theta(x, y) = \left(\max\{x^{-\theta} + y^{-\theta} - 1; 0\}\right)^{-1/\theta}$
   - Frank: $H_\theta(x, y) = -\frac{1}{\theta} \log\left(1 + \frac{(\exp(-\theta x)-1)(\exp(-\theta y)-1)}{\exp(-\theta)-1}\right)$
   - Gumbel: $H_\theta(x, y) = \exp\left(-((-\log(x))^\theta + (-\log(y))^\theta)^{1/\theta}\right)$

The Gaussian copula, as well as Clayton's, Frank's, Gumbel's and other Archimedean copulas, all satisfy our coherence and differentiability assumptions.

The only assumption our model makes on the marginals is that they are continuous. This is not particularly restrictive as long as there are no ties (see Section 3.4.2 for a treatment of ties).

Elements of correlation theory

In this section, we present common measures of correlation used in the literature, and some of their properties.

Definition 3.5 (Common measures of correlation)**.** Let (X, Y) be two random variables with respective cdfs F_X, F_Y. Define:

1. Pearson's correlation: assume X, Y have finite standard deviations σ_X and σ_Y. Then $r_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$.

2. Spearman's correlation: let $rk_X = F_X(X)$ and $rk_Y = F_Y(Y)$. We can think of rk_X as describing the ranking of X inside a sample. Then Spearman's correlation is $\rho_{X,Y} = r_{rk_X, rk_Y}$.

3. Kendall's correlation: let (X_1, Y_1) and (X_2, Y_2) be two independent pairs of random variables with the same joint distribution as (X, Y). Then Kendall's correlation is

$$\begin{aligned}\tau_{X,Y} = \quad & \mathbb{P}\left[(X_1 > X_2 \cap Y_1 > Y_2) \cup (X_1 < X_2 \cap Y_1 < Y_2)\right] - \\ & \mathbb{P}\left[(X_1 > X_2 \cap Y_1 < Y_2) \cup (X_1 < X_2 \cap Y_1 > Y_2)\right].\end{aligned}$$

We use the same letter r for the covariance of the standard bivariate Gaussian and for Pearson's correlation as they are equal. Moreover, for this distribution simple expressions exist for the two other correlation coefficients:

$$\rho = \frac{6}{\pi} \arcsin(r/2), \tau = \frac{2}{\pi} \arcsin(r).$$

A correlation measure should be zero when variables are independent, and reach its maximum when the variables are totally dependent on each other. The following lemma provides these properties for the measures we just introduced.

Lemma 3.18 ([Sca84, Theorems 1, 4, and 5])**.** *Let X, Y be two real random variables.*

1. $r_{X,Y}, \rho_{X,Y}, \tau_{X,Y} \in [-1, 1]$.

2. $\rho_{X,Y} = 1$ *if and only if* $Y = g(X)$ *with* $g : \mathbb{R} \to \mathbb{R}$ *increasing. The same holds for* $\tau_{X,Y}$. $r_{X,Y} = 1$ *if and only if the relation is affine.*

3. *If* X *and* Y *are independent, then* $r_{X,Y} = \rho_{X,Y} = \tau_{X,Y} = 0$.

Complements on matching with a continuum of students

To define matching in a continuum context, we follow [AL16].

Definition 3.6. A *matching* is an assignment of students to colleges, described by a mapping $\mu : S \cup \{A, B\} \to 2^S \cup C \cup S$, with the following properties:

1. for all $s \in S$, $\mu(s) \in \{A, B\} \cup \{s\}$;

2. for $C \in \{A, B\}$, $\mu(C) \subseteq S$ is measurable and $\eta(\mu(C)) \leq \alpha_C$;

3. $C = \mu(s)$ if and only if $s \in \mu(C)$;

4. for $C \in \{A, B\}$, the set $\{s \in S : \mu(s) \preceq_s C\}$ is open.

The first three conditions are common to almost all definitions of matching in discrete or continuous models. Condition (1) ensures that a student is either matched to a college or to themselves, which means that they remain unmatched. Condition (2) ensures that colleges are assigned to a subset of students that respects the capacity constraints. Condition (3) ensures that the matching is consistent, i.e., if a student is matched to a college, then this college is also matched to the student. Condition (4) was introduced by Azevedo and Leshno [AL16] and is necessary to ensure that there do not exist several stable matchings that only differ by a set of students of measure 0.

To produce a stable matching, one can extend the classic deferred acceptance algorithm by [GS62] to the continuum model. This algorithm is described in Algorithm 7.

If the algorithm stops, the matching it outputs is stable; [ACY15] show that even when the number of steps is infinite, the algorithm converges to a stable matching.

Remark 3.6. Note that stable matchings do not only result from centralized algorithms but are often the result of a decentralized process (see, e.g., [RV90]).

Algorithm 7: Deferred acceptance algorithm for a continuum of students

First step: All students apply to their favorite college, they are temporarily accepted. If the mass of students applying to college C is greater than its capacity α_C, then C only keeps the α_C best

while *A positive mass of students are unmatched and have not yet been rejected from every college* **do**

Each student who has been rejected at the previous step proposes to her preferred college among those which have not rejected them yet

Each college C keeps the best α_C mass of students among those it had temporarily accepted and those who just applied, and rejects the others

End: If the mass of students that are either matched or rejected from every college is 1, the algorithm stops. However it could happen that it takes an infinite number of steps to converge.

3.8.3 Omitted proofs

Proof of Lemma 3.8.

Assume $\alpha^A + \alpha^B < 1$, and $\theta \in \mathring{\Theta}^K$. Let $P^A, P^B \in \mathbb{R}$ be the cutoffs of colleges A and B.

By definition of the quantities V_1 and V_2, the market-clearing equation (3.1) can be written as

$$\begin{cases} \sum_{j\in[K]} \left(\gamma_j\beta_j V_1^{G_j,A} + \gamma_j(1-\beta_j)V_2^{G_j,B}\right) & = \alpha^A, \\ \sum_{j\in[K]} \left(\gamma_j\beta_j V_2^{G_j,A} + \gamma_j(1-\beta_j)V_1^{G_j,B}\right) & = \alpha^B. \end{cases}$$

Then, using Lemma 3.3, we can rewrite it as

$$\begin{cases} \sum_{j\in[K]} \left(\gamma_j\beta_j\mathbb{P}_j(W^A \geq P^A) + \gamma_j(1-\beta_j)\mathbb{P}_{j,\theta_j}(W^A \geq P^A, W^B < P^B)\right) & = \alpha^A, \\ \sum_{j\in[K]} \left(\gamma_j\beta_j\mathbb{P}_{j,\theta_j}(W^A < P^A, W^B \geq P^B) + \gamma_j(1-\beta_j)\mathbb{P}_j(W^B \geq P^B)\right) & = \alpha^B, \end{cases}$$

which is finally equivalent to

$$\begin{cases} \sum_{j\in[K]} \left(\gamma_j\beta_j \int_{P^A}^{\infty} f_j^A(x)\,\mathrm{d}x + \gamma_j(1-\beta_j)\int_{P^A}^{\infty}\int_{-\infty}^{P^B} f_{j,\theta_j}(x,y)\,\mathrm{d}x\mathrm{d}y\right) & = \alpha^A, \\ \sum_{j\in[K]} \left(\gamma_j\beta_j \int_{-\infty}^{P^A}\int_{P^B}^{\infty} f_{j,\theta_j}(x,y)\,\mathrm{d}x\mathrm{d}y + \gamma_j(1-\beta_j)\int_{P^B}^{\infty} f_j^B(x)\,\mathrm{d}x\right) & = \alpha^B. \end{cases} \tag{3.9}$$

We fix θ, and we want to study how the solution (P^A, P^B) of the above equation varies as a function of θ_j for some $j \in [K]$. Let us define $Z : \mathbb{R}^2 \times \Theta \to \mathbb{R}^2$, $(P^A, P^B, \theta_j) \mapsto$

$(D_A(P^A, P^B) - \alpha^A, D_B(P^A, P^B) - \alpha^B)$. (We will denote by Z_1 and Z_2 its two components.)

$$\begin{pmatrix} Z_1(P^A, P^B, \theta_j) \\ Z_2(P^A, P^B, \theta_j) \end{pmatrix} = \begin{pmatrix} \sum_{j\in[K]} \left(\gamma_j\beta_j \int_{P^A}^{\infty} f_j^A(x)\,\mathrm{d}x + \gamma_j(1-\beta_j) \int_{P^A}^{\infty}\int_{-\infty}^{P^B} f_{j,\theta_j}(x,y)\,\mathrm{d}x\mathrm{d}y \right) - \alpha^A \\ \sum_{j\in[K]} \left(\gamma_j\beta_j \int_{-\infty}^{P^A}\int_{P^B}^{\infty} f_{j,\theta_j}(x,y)\,\mathrm{d}x\mathrm{d}y + \gamma_j(1-\beta_j) \int_{P^B}^{\infty} f_j^B(x)\,\mathrm{d}x \right) - \alpha^B \end{pmatrix} \tag{3.10}$$

Then for each $\theta_j \in \Theta$, (P^A, P^B) is the solution of the equation $Z(P^A, P^B, \theta_j) = (0, 0)$. In order to show that P^A and P^B are decreasing in θ_j, we wish to apply the implicit function theorem. Let $P^A, P^B \in \mathbb{R}$ and $\theta_j \in \Theta$ such that $Z(P^A, P^B, \theta_j) = 0$. Function Z is of class C^1. We first verify that the partial Jacobian $J_{Z,(P^A,P^B)}(P^A, P^B, \theta_j)$ is invertible, where

$$J_{Z,(P^A,P^B)}(P^A, P^B, \theta_j) = \begin{pmatrix} \frac{\partial Z_1}{\partial P^A} & \frac{\partial Z_1}{\partial P^B} \\ \frac{\partial Z_2}{\partial P^A} & \frac{\partial Z_2}{\partial P^B} \end{pmatrix}. \tag{3.11}$$

To show that the determinant $\frac{\partial Z_1}{\partial P^A}\frac{\partial Z_2}{\partial P^B} - \frac{\partial Z_1}{\partial P^B}\frac{\partial Z_2}{\partial P^A} \neq 0$, we will show that it is in fact strictly positive. From (3.9), it is clear that Z_1 is decreasing in P^A and increasing in P^B, and that Z_2 is increasing in P^A and decreasing in P^B. Therefore, to prove that $\frac{\partial Z_1}{\partial P^A}\frac{\partial Z_2}{\partial P^B} - \frac{\partial Z_1}{\partial P^B}\frac{\partial Z_2}{\partial P^A} > 0$, we only need to prove that $\left|\frac{\partial Z_1}{\partial P^A}\right| > \frac{\partial Z_2}{\partial P^A}$ and $\left|\frac{\partial Z_2}{\partial P^B}\right| > \frac{\partial Z_1}{\partial P^B}$.

By symmetry, we will only prove the first one. We can compute each term separately:

$$\frac{\partial Z_1}{\partial P^A} = \sum_{j\in[K]} \left(\gamma_j\beta_j \frac{\partial\,\mathbb{P}_j(W^A \geq P^A)}{\partial P^A} + \gamma_j(1-\beta_j) \frac{\partial\,\mathbb{P}_{j,\theta_j}(W^A \geq P^A, W^B < P^B)}{\partial P^A} \right),$$
$$\frac{\partial Z_2}{\partial P^A} = \sum_{j\in[K]} \gamma_j\beta_j \frac{\partial\,\mathbb{P}_{j,\theta_j}(W^A < P^A, W^B \geq P^B)}{\partial P^A}.$$

All terms of Z_1 are decreasing in P^A and all terms of Z_2 are increasing in P^A, therefore we can proceed term by term:

$$\begin{aligned} \left|\gamma_j\beta_j \frac{\partial\,\mathbb{P}_j(W^A \geq P^A)}{\partial P^A}\right| &= \gamma_j\beta_j \quad \frac{\partial\,\mathbb{P}_j(W^A < P^A)}{\partial P^A}, \\ &= \gamma_j\beta_j \quad \left(\frac{\partial\,\mathbb{P}_{j,\theta_j}(W^A < P^A, W^B < P^B)}{\partial P^A} \right. && (3.12) \\ &\qquad \left. + \frac{\partial\,\mathbb{P}_{j,\theta_j}(W^A < P^A, W^B \geq P^B)}{\partial P^A} \right), && (3.13) \\ &> \gamma_j\beta_j \quad \frac{\partial\,\mathbb{P}_{j,\theta_j}(W^A < P^A, W^B \geq P^B)}{\partial P^A}. \end{aligned}$$

We conclude that $\left|\frac{\partial Z_1}{\partial P^A}\right| > \frac{\partial Z_2}{\partial P^A}$, and similarly $\left|\frac{\partial Z_2}{\partial P^B}\right| > \frac{\partial Z_1}{\partial P^B}$. Therefore the Jacobian in (3.11) has positive determinant and is invertible.

By the implicit function theorem, there exists a neighborhood U of (P^A, P^B), a neighborhood V of θ_j, and a function $\psi : V \to U$ such that for all $(x, y) \in \mathbb{R}^2, \theta \in \Theta$,

$$(\,(x, y, \theta) \in U \times V \text{ and } Z(x, y, \theta) = 0\,) \Leftrightarrow (\,\theta \in V \text{ and } (x, y) = \psi(\theta)\,).$$

In particular, $(P^A, P^B) = \psi(\theta_j)$, and we can compute the derivative of ψ:

$$\begin{aligned} J_\psi(\theta_j) &= -J_{Z,(P^A,P^B)}(P^A, P^B, \theta_j)^{-1}\, J_{Z,\theta_j}(P^A, P^B, \theta_j), \\ &= \frac{-1}{\frac{\partial Z_1}{\partial P^A}\frac{\partial Z_2}{\partial P^B} - \frac{\partial Z_1}{\partial P^B}\frac{\partial Z_2}{\partial P^A}} \begin{pmatrix} \dfrac{\partial Z_2}{\partial P^B} & -\dfrac{\partial Z_1}{\partial P^B} \\ -\dfrac{\partial Z_2}{\partial P^A} & \dfrac{\partial Z_1}{\partial P^A} \end{pmatrix} \begin{pmatrix} \dfrac{\partial Z_1}{\partial \theta_j} \\ \dfrac{\partial Z_2}{\partial \theta_j} \end{pmatrix}, \\ &= \frac{-1}{\frac{\partial Z_1}{\partial P^A}\frac{\partial Z_2}{\partial P^B} - \frac{\partial Z_1}{\partial P^B}\frac{\partial Z_2}{\partial P^A}} \begin{pmatrix} \dfrac{\partial Z_2}{\partial P^B}\dfrac{\partial Z_1}{\partial \theta_j} - \dfrac{\partial Z_1}{\partial P^B}\dfrac{\partial Z_2}{\partial \theta_j} \\ -\dfrac{\partial Z_2}{\partial P^A}\dfrac{\partial Z_1}{\partial \theta_j} + \dfrac{\partial Z_1}{\partial P^A}\dfrac{\partial Z_2}{\partial \theta_j} \end{pmatrix}. \end{aligned} \tag{3.14}$$

We only need to know the sign of each term to conclude about the variations of ψ. We already know the sign of the derivatives in P^A and P^B, so we only need those in θ_j. The terms of Z_1 that depend on θ_j are $\sum_{j\in[K]} \gamma_j(1-\beta_j)\mathbb{P}_{j,\theta_j}(W^A \geq P^A, W^B < P^B)$. By Lemma 3.1, $\mathbb{P}_{\theta_j}(W^A \geq P^A, W^B < P^B)$ is decreasing in θ_j, and thus $\frac{\partial Z_1}{\partial \theta_j} < 0$. By the same argument, $\frac{\partial Z_2}{\partial \theta_j}$ is also negative. (Note that here the implicit functions theorem requires that we compute the partial derivatives of Z as if P^A and P^B were not functions of θ_j.)

If we replace each term of the last line of Equation (3.14) by its signs, we get

$$-\frac{1}{+}\begin{pmatrix} (-\times-) - (+\times-) \\ -(+\times-) + (-\times-) \end{pmatrix} = \begin{pmatrix} - \\ - \end{pmatrix}.$$

We conclude that ψ and therefore P^A and P^B are decreasing in θ_j.

Note that we require $\theta_j \in \mathring{\Theta}$ because if one of the θ_j is maximal, i.e., the distribution is fully correlated (W^B is a deterministic function of W^A), then the V_1 metrics are not differentiable at this point. However, they are continuous, therefore they are increasing on the whole interval Θ. Moreover, if the distribution is not fully correlated when θ is maximal, then we can replace $\mathring{\Theta}$ by Θ in the statement of the lemma. ■

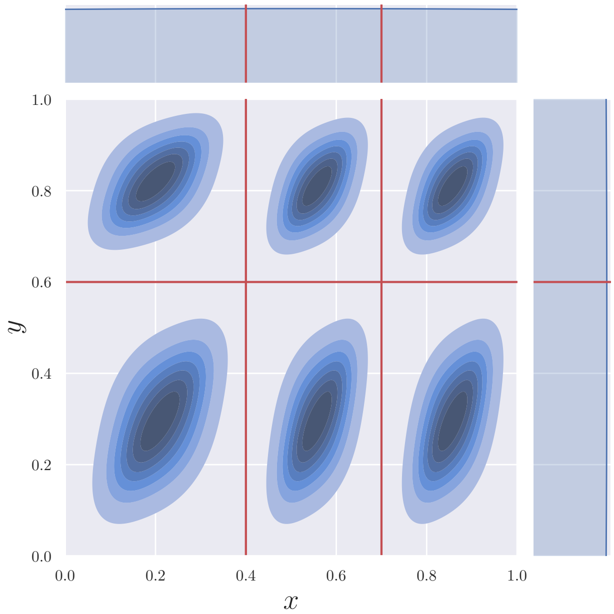


Fig. 3.9: Illustration of the distribution f_θ with three priority classes at A (30% of applicants in the first class, 30% in the second, 40% in the third), two priority classes at B (40% in the first class, 60% in the second), and correlation $\theta = 0.5$.

Proof of Proposition 3.11.

We start by building a distribution family that can represent both STB and MTB for two values of the parameter. The priority classes are $Q_1^A, \ldots Q_{n_A}^A$ and $Q_1^B, \ldots Q_{n_B}^B$, and we denote by $\kappa_j^C = \eta(Q_j^C)$ the mass of students inside class j of college C. Let $a_0 = 0, a_1 = \kappa_1^A, a_2 = \kappa_1^A + \kappa_2^A, \ldots, a_{n_A} = 1$, such that they form a partition of $[0,1]$ with the j-th segment having length κ_j^A. Define $b_0, \ldots, b_{n_B}$ similarly. Finally, for any $i \leq n_A, j \leq n_B$, let $\kappa_{i,j} = \eta(Q_i^A \times Q_j^B)$ be the mass of students belonging to class i for A and class j for B.

Let ϕ_θ be the pdf of the Gaussian copula with uniform marginals on $[0,1]^2$ and covariance θ. For any $\theta \in [-1,1]$, let $f_\theta : [-1,1]^2 \to \mathbb{R}$ be defined as:

$$f_\theta(x,y) = \kappa^{i,j} \phi_\theta\left(\frac{x - a_{i-1}}{\kappa_i^A}, \frac{y - b_{j-1}}{\kappa_j^B}\right) \text{ with } a_{i-1} \leq x \leq a_i, b_{j-1} \leq y \leq b_j$$

Defined this way, f_r is a pdf since it is non-negative and has integral 1. The marginals are uniform and do not depend on θ. Moreover, the integral of f_r over each rectangle $Q_i^A \times Q_j^B$ is $\kappa^{i,j}$, and each rectangle contains a "copy" of the Gaussian copula adjusted to its dimensions. There is no "spill" between classes: if student s is in a higher priority class at college C than student s', then s will have a higher score with probability 1. If for all i, j, $\kappa^{i,j} > 0$, and $\theta \notin \{-1, 1\}$, then f_θ has full support. This distribution is depicted in Figure 3.9.

We can verify that this definition recovers MTB and STB: if $\theta = 0$, if two students are in the same priority class for a college, they have the same ex-ante probability of getting a seat there, and if they also are in the same priority class for the other college (i.e., they are in the same rectangle $Q_i^A \times Q_j^B$), the result of this second tie-breaking is independent from the first one. When $\theta = 1$, if two students are in the same priority class for a college, have the same ex-ante probability of getting a seat there, but if they also are in the same priority class for the other college (i.e., they are in the same rectangle $Q_i^A \times Q_j^B$), the winner of the tie-breaking is the same as for the first college since scores inside the rectangle are perfectly correlated. In that case, the distribution does not have full support but this is not an issue as explained in Remark 3.2. Therefore MTB is the case $\theta = 0$ and STB $\theta = 1$.

Let us now prove the proposition:

1. - The family $(f_\theta)_{\theta \in [-1,1]}$ is differentiable by differentiability of the Gaussian copula. It is also coherent (except for the (x, y) such that $x = a_i$ or $y = b_j$, i.e., on the sides of rectangles, in which case the cdf is constant and not increasing). Therefore by applying Theorem 3.7, E is either increasing or constant.
   - Moreover, the case where it could be constant can only happen if there are several priority classes, so if there is only one it is strictly increasing.
   - Let us look into the multiple priority classes case.Suppose that $\exists \theta \in \Theta$ such that $P^A(\theta) \neq a_i$ and $P^B(\theta) \neq b_j$ for all i, j. We can then apply Theorem 3.7, and deduce that V_1^A and V_1^B are increasing in all θ_j. If there exists no such θ, it implies that P^A, P^B are constant in θ and so are V_1^A and V_1^B. However, as any perturbation of either $\gamma, \beta or \alpha$ would change the cutoffs, and resolve the issue, the set of problematic values of (γ, β, α) has Lebesgue measure 0.
2. Finally, Corollary 3.9 can be applied with the same adjustments, which gives the fourth point.

Proof of Proposition 3.13

For this proof, to simplify computations, assume without loss of generality that marginals follow a uniform distribution on $[0, 1]$. Since the Deferred Acceptance algorithm only depends on ordinal comparisons, this assumption is indeed not restrictive and switching to a uniform distribution will greatly help solving the market-clearing equation 3.9. The students' score vectors are therefore uniformly distributed along the diagonal of the square $[0,1]^2$. The cutoffs P^A and P^B belong to $[0,1]$, and the metrics are given by:

$$\begin{aligned} &V_1^A = 1 - P^A, && V_1^B = 1 - P^B, \\ &V_2^A = \max(P^A - P^B, 0), && V_2^B = \max(P^B - P^A, 0), \\ &V_\emptyset = \min(P^A, P^B). \end{aligned} \tag{3.15}$$

Therefore, the market-clearing equation is

$$\begin{cases} \beta(1 - P^A) + (1-\beta)\max(P^B - P^A, 0) &= \alpha^A, \\ \beta \max(P^A - P^B, 0) + (1-\beta)(1 - P^B) &= \alpha^B. \end{cases}$$

Assume that $P^B \geq P^A$. Then we have

$$\begin{cases} \beta(1 - P^A) + (1-\beta)(P^B - P^A) &= \alpha^A, \\ (1-\beta)(1 - P^B) &= \alpha^B, \end{cases} \tag{3.16}$$

which is equivalent to

$$\begin{cases} P^A = 1 - \alpha^A - \alpha^B, \\ P^B = 1 - \frac{\alpha^B}{1-\beta}. \end{cases}$$

Moreover, the assumption $P^B \geq P^A$ implies that $\beta \leq \frac{\alpha^A}{\alpha^A + \alpha^B}$. P^A and P^B are well-defined, that is, they are in $[0,1]$. For P^A, this follows from the assumption $\alpha^A + \alpha^B < 1$, and for P^B it is implied by the relation $\beta \leq \frac{\alpha^A}{\alpha^A + \alpha^B}$. If $P^A \geq P^B$ instead, we have:

$$\begin{cases} P^A = 1 - \frac{\alpha^A}{\beta}, \\ P^B = 1 - \alpha^A - \alpha^B. \end{cases} \tag{3.17}$$

Similarly, this implies that $\beta \geq \frac{\alpha^A}{\alpha^A + \alpha^B}$, and using this we can verify that $P^A, P^B \in [0,1]$.

We can then conclude that if $\beta \leq \frac{\alpha^A}{\alpha^A + \alpha^B}$, then

$$\begin{aligned} &V_1^A = \alpha^A + \alpha^B, && V_1^B = \frac{\alpha^B}{1-\beta}, \\ &V_2^A = 0, && V_2^B = \alpha^A - \frac{\beta}{1-\beta}\alpha^B, \\ &V_\emptyset = 1 - \alpha^A - \alpha^B; \end{aligned}$$

and if $\beta \geq \frac{\alpha^A}{\alpha^A + \alpha^B}$, then

$$\begin{aligned} &V_1^A = \frac{\alpha^A}{\beta}, && V_1^B = \alpha^A + \alpha^B, \\ &V_2^A = \alpha^B - \frac{1-\beta}{\beta}\alpha^A, && V_2^B = 0, \\ &V_\emptyset = 1 - \alpha^A - \alpha^B. \end{aligned}$$

This is obtained by replacing in (3.15) the values of P^A and P^B found in (3.16) and (3.17). ■

Proof of Proposition 3.14

As in the proof of Proposition 3.13, we assume without loss of generality for this proof that the marginals are uniform over $[0, 1]$. Then the grades at colleges A and B are independent random variables with a uniform distribution over [0, 1]. Students' score vectors are thus uniformly distributed on the whole area of the square $[0, 1]^2$. Therefore the metrics as functions of P^A and P^B are:

$$\begin{array}{ll} V_1^A = 1 - P^A, & V_1^B = 1 - P^B, \\ V_2^A = P^A(1 - P^B), & V_2^B = P^B(1 - P^A), \\ V_\emptyset = P^A P^B. & \end{array} \tag{3.18}$$

The market-clearing equation is:

$$\begin{aligned} & \begin{cases} \beta(1 - P^A) + (1-\beta)P^B(1 - P^A) = \alpha^A, \\ \beta P^A(1 - P^B) + (1-\beta)(1 - P^B) = \alpha^B, \end{cases} \\ \Leftrightarrow \quad & \begin{cases} P^B = 1 - \alpha^B - \frac{\beta}{1-\beta}(1 - P^A - \alpha^A), \\ P^A P^B = 1 - \alpha^A - \alpha^B, \end{cases} \\ \Leftrightarrow \quad & \begin{cases} P^B = 1 - \alpha^B - \frac{\beta}{1-\beta}(1 - P^A - \alpha^A), \\ \frac{\beta}{1-\beta}(P^A)^2 + (\frac{1-2\beta}{1-\beta} + \frac{\beta}{1-\beta}\alpha^A - \alpha^B)P^A - (1 - \alpha^A - \alpha^B) = 0. \end{cases} \end{aligned} \tag{3.19}$$

Let $\zeta = \frac{1-2\beta}{1-\beta} + \frac{\beta}{1-\beta}\alpha^A - \alpha^B$ and $\Delta = \sqrt{\zeta^2 + \frac{4\beta}{1-\beta}(1 - \alpha^A - \alpha^B)}$. From (3.19) and the fact that $P^A \geq 0$ we deduce that

$$\begin{cases} P^A = \frac{1-\beta}{2\beta}(\Delta - \zeta), \\ P^B = \frac{1}{2}(\Delta + \zeta). \end{cases}$$

Injecting this in Equation (3.18) finally gives

$$\begin{array}{ll} V_1^A = 1 - \frac{1-\beta}{2\beta}(\Delta - \zeta), & V_1^B = 1 - \frac{1}{2}(\Delta + \zeta), \\ V_2^A = \frac{1-\beta}{2\beta}(\Delta - \zeta) - \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), & V_2^B = \frac{1}{2}(\Delta + \zeta) - \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), \\ V_\emptyset = \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), & \end{array}$$

which concludes the proof of the proposition. ■

3.8.4 Extension to more than two colleges: proof attempt

In this section we try to prove Lemma 3.8 for more than two colleges, using the model introduced in Section 3.5. The market-clearing equation (3.1) can be written as

$$\begin{cases} \sum_{j\in[K]} \gamma_j \sum_{i=1}^{m} \sum_{\substack{\sigma\in\Sigma([m])\\ \sigma(i)=1}} \beta^{\sigma}_{G_j} \mathbb{P}_{\theta_j}\left(\bigcap_{n<i} (W^{\sigma(n)} < P^{\sigma(n)}) \cap W^{\sigma(i)} \geq P^{\sigma(i)}\right) & = \alpha^1 \\ \vdots & \\ \sum_{j\in[K]} \gamma_j \sum_{i=1}^{m} \sum_{\substack{\sigma\in\Sigma([m])\\ \sigma(i)=m}} \beta^{\sigma}_{G_j} \mathbb{P}_{\theta_j}\left(\bigcap_{n<i} (W^{\sigma(n)} < P^{\sigma(n)}) \cap W^{\sigma(i)} \geq P^{\sigma(i)}\right) & = \alpha_m. \end{cases} \tag{3.20}$$

We fix all θ_i except for θ_j for some j, and we want to study how the solution $\mathbf{P}(\theta)$ of the above equation varies as a function of θ_j. Let us define $T : \mathbb{R}^m \times \Theta \to \mathbb{R}^m$, $(\mathbf{P}, \theta_j) \mapsto (D_{C_1}(\mathbf{P}(\theta)) - \alpha_1, \ldots, D_{C_m}(\mathbf{P}(\theta)) - \alpha_m)$. We will denote by $T^1, \ldots, T^m$ the components of T. Then for any $\theta_j \in \Theta$, the solution $\mathbf{P}$ of Equation (3.20) is also the solution of the equation $T(\mathbf{P}, \theta_j) = 0_{\mathbb{R}^m}$. In order to show that $P^1, \ldots, P^m$ are all decreasing in θ_j, we wish to apply the implicit function theorem. Let $\mathbf{P} \in \mathbb{R}^m$ and $\theta_j \in \Theta$ such that $T(\mathbf{P}, \theta_1) = 0$. Function T is of class $\mathcal{C}^1$ because $(H_\theta)_{\theta\in\Theta}$ is differentiable. We first verify that the partial Jacobian $J_{T,\mathbf{P}}(\mathbf{P}, \theta_j)$ is invertible, where

$$J_{T,\mathbf{P}}(\mathbf{P}, \theta_j) = \begin{pmatrix} \frac{\partial T^1}{\partial P^1} & \cdots & \frac{\partial T^1}{\partial P^m} \\ \vdots & & \vdots \\ \frac{\partial T^m}{\partial P^1} & \cdots & \frac{\partial T^m}{\partial P^m} \end{pmatrix}. \tag{3.21}$$

To prove this, we will show that no linear combination of the rows of $J_{T,\mathbf{P}}(\mathbf{P}, \theta_j)$ can be equal to zero. We start by proving the following inequality:

$$\forall i \in \{1, \ldots, m\}, \sum_{k\in\{1,\ldots,m\}} \frac{\partial T^k}{\partial P^i} < 0 \tag{3.22}$$

We start by noticing that for any $i \neq k$, $\frac{\partial T^i}{\partial P^i} < 0$ and $\frac{\partial T^k}{\partial P^i} > 0$. This is due to T^k being a sum of probabilities of intersections containing terms of the type $W^\ell < P_\ell$ for all $\ell \neq k$, and $W^k \geq P^k$. Therefore, T^k is decreasing in P^i and increasing in all the other cutoffs P^ℓ. So proving the relation (3.22) amounts to proving that $-\frac{\partial T^i}{\partial P^i} > \sum_{k\neq i} \frac{\partial T^k}{\partial P^i}$. The first term can be written:

$$-\frac{\partial T^i}{\partial P^i} = -\sum_{\ell\in[K]} \gamma_\ell \sum_{k=1}^{m} \sum_{\substack{\sigma\in\Sigma([m])\\ \sigma(k)=i}} \beta^{\sigma}_{\ell} \frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n<k} (W^{\sigma(n)} < P^{\sigma(n)}) \cap W^{\sigma(i)} \geq P^{\sigma(i)}\right)}{\partial P^i}. \tag{3.23}$$

Notice that

$$\begin{aligned}
&\mathbb{P}_{\theta_\ell}\left(\bigcap_{n<k}(W^{\sigma(n)} < P^{\sigma(n)}) \cap W^{\sigma(i)} \geq P^{\sigma(i)}\right)\\
=&\mathbb{P}_{\theta_\ell}\left(\bigcap_{n<k}(W^{\sigma(n)} < P^{\sigma(n)})\right) - \mathbb{P}_{\theta_\ell}\left(\bigcap_{n\leq k}(W^{\sigma(n)} < P^{\sigma(n)})\right)
\end{aligned}$$

where the first term is constant in $P^i = P^{\sigma(k)}$. We then deduce that

$$\begin{aligned}
&\frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n<k}(W^{\sigma(n)} < P^{\sigma(n)}) \cap W^{\sigma(i)} \geq P^{\sigma(i)}\right)}{\partial P^i}\\
=&-\frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n\leq k}(W^{\sigma(n)} < P^{\sigma(n)})\right)}{\partial P^i}.
\end{aligned}$$

Injecting this relation in (3.23) gives:

$$-\frac{\partial T^i}{\partial P^i} = \sum_{\ell\in[K]}\gamma_\ell \sum_{k=1}^{m} \sum_{\substack{\sigma\in\Sigma([m])\\ \sigma(k)=i}} \beta_\ell^\sigma \frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n\leq k}(W^{\sigma(n)} < P^{\sigma(n)})\right)}{\partial P^i}. \tag{3.24}$$

Now consider the terms in T^k for $k \neq i$:

$$\frac{\partial T^k}{\partial P^i} = \sum_{\ell\in[K]}\gamma_\ell \sum_{k=1}^{m} \sum_{\substack{\sigma\in\Sigma([m])\\ \sigma(k)=C_k\\ \sigma^{-1}(C_i)<k}} \beta_\ell^\sigma \frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n<k}(W^{\sigma(n)} < P^{\sigma(n)}) \cap W^{\sigma(i)} \geq P^{\sigma(i)}\right)}{\partial P^i}.$$

We now want to prove that $\sum_{k\neq i}\frac{\partial T^k}{\partial P^i} < -\frac{\partial T^i}{\partial P^i}$. We can do so by comparing them term by term. Let $\sigma \in \Sigma([m])$, and consider the terms in each side of the inequality that have $\beta^\sigma\ell$ as a factor for some ℓ. On the right side, it gives

$$\sum_{\ell\in[K]}\beta_\ell^\sigma \frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n\leq k}(W^{\sigma(n)} < P^{\sigma(n)})\right)}{\partial P^i}.$$

On the left side, this term exists only for the ks such that $\sigma^{-1}(i) < \sigma^{-1}(k)$:

$$\sum_{\ell\in[K]}\beta_\ell^\sigma \sum_{\{k:\sigma^{-1}(i)<\sigma^{-1}(k)\}} \frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n<k}(W^{\sigma(n)} < P^{\sigma(n)}) \cap W^{\sigma(i)} \geq P^{\sigma(i)}\right)}{\partial P^i}. \tag{3.25}$$

Fix a group ℓ and consider the last term of the sum in (3.25)

$$\frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n<m}(W^{\sigma(n)} < P^{\sigma(n)}) \cap W^{\sigma(m)} \geq P^{\sigma(m)}\right)}{\partial P^i}$$

that can be upper-bounded by

$$\frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n<m}(W^{\sigma(n)} < P^{\sigma(n)})\right)}{\partial P^i}.$$

The penultimate term being

$$\frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n<m-1}(W^{\sigma(n)} < P^{\sigma(n)}) \cap W^{\sigma(m-1)} \geq P^{\sigma(m-1)}\right)}{\partial P^i},$$

we can add those two to obtain

$$\frac{\partial \mathbb{P}_{\theta_\ell}\left(\bigcap_{n<m-1}(W^{\sigma(n)} < P^{\sigma(n)})\right)}{\partial P^i}.$$

and so on, we can continue packing the terms together, and we finally obtain

$$\sum_{\{k:\sigma^{-1}(i)<\sigma^{-1}(k)\}} \frac{\partial \mathbb{P}_{\theta_\ell}(W^{\sigma(1)} < P^{\sigma(1)} \cap \dots \cap W^i < P^i \cap \dots \cap W^k \geq P^k)}{\partial P^i}$$
$$< \frac{\partial \mathbb{P}_{\theta_\ell}(W^{\sigma(1)} < P^{\sigma(1)} \cap \dots \cap W^i < P^i)}{\partial P^i}$$

which is the term associated to β_ℓ^σ in $-\frac{\partial T^i}{\partial P^i}$ (cf. Equation (3.24)). We finally conclude that $\forall i \in \{1, \dots, m\}, \sum_{k\in\{1,\dots,m\}} \frac{\partial T^k}{\partial P^i} < 0$.

We can now use this result to prove that $J_{T,\mathbf{P}}(\mathbf{P}, \theta_j)$ is invertible. Let us call $R^i := (\frac{\partial T^i}{\partial P^k})_k$ the i-th row of $J_{T,\mathbf{P}}(\mathbf{P}, \theta_j)$. Assume that there exists $\lambda^1, \dots, \lambda^m$ such that $\sum \lambda^i R^i = 0_{\mathbb{R}^m}$. Let $i_0 \in \arg\max_i \lambda^i$. Then on the i_0-th component we have

$$\sum_i \lambda^i \frac{\partial T^i}{\partial P^{i_0}} \leq \lambda^{i_0} \sum_i \frac{\partial T^i}{\partial P^{i_0}}$$
$$< 0$$

which contradicts $\sum \lambda^i R^i = 0_{\mathbb{R}^m}$. We conclude that no linear combination of the rows of $J_{T,\mathbf{P}}(\mathbf{P}, \theta_j)$ can be zero, therefore it is invertible.

The assumptions of the implicit function theorem are verified, therefore there exists a neighborhood $U \subseteq \mathbb{R}^m \times \Theta$ of $(\mathbf{P}, \theta_j)$, a neighborhood $V \subseteq \Theta$ of θ_j, and a function $\psi : V \to \mathbb{R}^m$ such that for all $(x, \theta) \in \mathbb{R}^m \times \Theta$,

$$(\,(x,\theta) \in U \text{ and } T(x,\theta) = 0\,) \Leftrightarrow (\,\theta \in V \text{ and } x = \psi(\theta)\,).$$

In particular, $\mathbf{P}(\theta) = \psi(\theta_j)$, and we can compute the derivative of ψ:

$$\begin{aligned} J_\psi(\theta_1) &= -J_{T,\mathbf{P}}(\mathbf{P}, \theta_1)^{-1}\, J_{T,\theta_1}(\mathbf{P}, \theta_1), \\ &= \frac{-1}{|J_{T,\mathbf{P}}(\mathbf{P}, \theta_1)|} C^T \begin{pmatrix} \frac{\partial T^1}{\partial \theta_j} \\ \vdots \\ \frac{\partial T^m}{\partial \theta_j} \end{pmatrix}, \end{aligned} \tag{3.26}$$

where C is the comatrix of $J_{T,\mathbf{P}}(\mathbf{P}, \theta_1)$. We want to show that ψ is decreasing in all its components, i.e., all P^i are decreasing in θ_j. However, the $\frac{\partial T^i}{\partial \theta_j}$ do not have constant sign as they did in the two college case, therefore we cannot conclude about the variations of ψ. While this is not a proof that the Lemma is flase, the counter example provided in Section 3.5 is. By using the calculations above, we might find particular values of the parameters for which the Lemma is true, but we would need to compute explicitly the expression above, that in most cases cannot be expressed in terms of usual functions.

Conclusion

Global discussion

In this thesis, I studied two classical and very generic matching models, without and with preferences. In both models I discovered explicit relations between the structure of the input (respectively, the shape of the graph and the correlation of preferences) and the (group) fairness of the matchings that are commonly chosen as solution concepts (respectively, maximum matchings and stable matchings).

The main conclusion that can be drawn from those results is that inequalities between groups will almost always arise if the central authority choosing the matching does not pay attention to them, as the common solution concepts do not include fairness constraints and very often lead to the selection of an unfair matching. The attention of the authority to group fairness is made even more important by the fact that inequalities arise even when all agents individually treat all groups fairly, only because of the global structure of the problem (that could be interpreted as a lack of coordination). Inequalities are then unintentional and agents either are oblivious to them, or notice them but have no way to understand their origin without looking at the problem in its entirety. In many cases of centralized matching, only the central authority has access to enough information to detect the sources of unfairness highlighted in this thesis and potentially mitigate them.

On the positive side, I showed that in many cases fairness can be achieved without any decrease in efficiency. Even when that is not the case, my results indicate that fairness and efficiency are not orthogonal and in many cases both can be improved simultaneously. In bipartite matching without preferences, there always exists a matching that is both maximum and fair when there are two groups, which is the case in many applications. When there are more groups, the price of fairness is bounded, and when the graph is drawn randomly it is very close to 1, a setting closer to real applications than the very specific cases where the PoF reaches the bound. In two sided matching, I showed that the main factor acting on both fairness and efficiency is the correlation of priorities. While it seems complex to impose an arbitrary degree of correlation between colleges rankings, a simple solution to theoretically achieve both fairness and maximal efficiency is to have a common exam, or standardized test, or to use a single tie-breaker when there are ties, ensuring a correlation of 1 for all groups. The fairness in this setting is theoretical because it implies that the exam or test is itself fair, in the sense that the grade distribution for each group is the same and all groups have equal access and preparation to it, which is a very strong assumption.

Regarding two-sided matching, we considered the most common solution concept, i.e., stability. However, there exist other mechanisms that Deferred Acceptance, such as Top Trading Cycles (TTC) or Random Serial Dictatorship (RSD), that output a matching that is not necessarily stable but that are more efficient than DA. RSD is equivalent to running student proposing DA with a common ranking for all colleges that is drawn randomly (equivalent to a single tie breaker with only one priority class), and therefore ensures maximal efficiency and perfect fairness, however it is very likely to feature justified envy. On the other hand, TTC outputs a matching that is Pareto efficient (there is no matching that is preferred by all students), which is not the case of DA, and is the Pareto-efficient mechanism that minimizes the number of blocking pairs, offering a compromise between RSD and DA. Two questions then arise. First, it would be interesting to perform on TTC the same analysis that I performed on DA in this thesis, in order to study its group-fairness properties and compare it to DA. Secondly, since RSD is efficient and fair, and much simpler than DA, choosing DA over it is a strong choice as it implies sacrificing some efficiency and fairness for stability.

Perspectives

To improve the understanding of group fairness issues in matching many avenues lie before us. The most direct ones were already discussed in the discussion section of each chapter, here I present some more general directions.

- The fairness literature has not yet explored in much detail intersectional fairness, i.e., the fact that a person might belong to several protected groups. Molina and Loiseau [ML22] showed that ensuring fairness towards each group is not sufficient to ensure fairness towards each intersection of groups (e.g., being fair between men and women, and between rich and poor people, does not ensure that poor women are treated fairly). A way to circumvent this problem is to consider each intersection of groups as a separate group. However, for bipartite matching, the bounds on the price of fairness increase with the number of groups, so the price of fairness could become very large with this approach. Using the particular structure of intersectional fairness could lead to better bounds that those obtained for disjoint groups.

- As mentioned above, TTC is a popular mechanism for two-sided matching that would be worth studying as we did with DA. It even has a cutoff characterization as shown by Leshno and Lo [LL20]. However, my analysis does not directly extend, because even with only two colleges the coherence assumption is not sufficient to know how the cutoffs vary with correlation. This is due to the fact that the cutoff structure is different: to get their first choice, students only need to pass the cutoff at some school, not necessarily at their preferred one. Conversely, they get their second choice (say college B) if they are below both those cutoffs but above a secondary cutoff at B. The variations in θ of the corresponding masses of students cannot be derived from the coherence assumption alone as I did for stable matchings. Studying the effect of correlation on TTC will then require additional work.

- Both models (with and without preferences) could be made online, by having agents/students arrive one by one, or by batches. The decision-maker would then need to decide whether to match them and to whom. Without preferences, the objective would be both to match as many agents as possible (see for instance [BE98]), while keeping the matching fair. In on line algorithms, an algorithm is evaluated by its competitive ratio, i.e., the ratio between its performance and the performance of the best offline algorithm on the same instance. For matching, classically we compare the size of the matching found by the online algorithm to the size of maximum matchings on the whole revealed graph. Similarly, we could compare the size of the fair matching found by the online algorithm to the largest fair matching in the whole graph as we now know how to find it. In two-sided matching, the objective classically is to minimize the number of blocking pairs. We could extend our study to existing algorithms to determine if correlation of priorities still plays a role, and if matchings found by online algorithms are more or less unfair than those found offline with DA.

Closing reflection

Matching theory and algorithms are widely used in applications that have a huge impact on people's lives, such as school choice, college admission, refugee resettlement, and many others. The questions regarding their fairness and efficiency properties are therefore highly sensitive, and impact society as a whole beyond the scientific communities that usually study them as abstract objects. We saw that choosing a matching in the set of all those that are feasible is a non trivial choice, that implies to make judgments about values: we can prioritize group fairness, efficiency, stability, or other objectives, but not all of them at once, at least not always. For each of those choices, the role of mathematicians, computer scientists and economists is to provide the best possible understanding of their consequences, and algorithms or mechanisms that give the best solution according to each choice. However, making the choice in the first place of the values to prioritize involves society as a whole.

The most prevalent cause of inequalities in matching problems stems from the input itself, i.e., the graph in bipartite matching or the marginal distributions in two-sided matching. If some groups have very few nodes connected to them, or marginal distribution of grades worse than the other groups, they will always be disadvantaged and the levers that act on inequalities presented in this thesis can only slightly mitigate that. Preventing such situations requires the attention of central authorities as well as transparency in the way edges of the graph are built or grades are given.

Another question that has to be answered before talking about group fairness, in any instance, is to define the groups. Which subsets of the population need to be treated equally, and which do not? What are the relevant attributes to define groups? Those questions requires once again societal choices to be made, enlightened by scientific work from other fields such as sociology and applied economics.

Regarding two-sided matching, we can also question the necessity of ranking people. While expressing preferences towards objects, services or education programs seems quite natural and necessary to each person's well being, the effects of institutions ranking people, sometimes with important consequences on their lives, are less obviously positive. First, most rankings are based on some notion of merit, that is a very ill-defined concept that is context dependent (a person can have some merit as a mathematician but much less as a football player) and cannot be directly observed. As stated by Hume [Hum40], "the external performance has no merit. We must look within to find the moral quality. This we cannot do directly; and therefore fix our attention on actions, as on external signs". We measure merit based on performance in some task, but the performance can be affected by many outside factors. The Stanford Encyclopedia of Philosophy, in its entry about meritocracy [Mul23], gives the following example: "suppose that Daryl and John are applying for jobs at the widget factory. Daryl is more skilled than John, works faster, is more careful, more collegial, and so on. But this factory is filled with racists and Daryl belongs to the disfavored race. As a result, Daryl would perform worse (make fewer widgets per hour) than John would. Observe that, conceptually, Daryl remains more meritorious than John even though hiring John would generate better outcomes". Further, we must consider the impact of ranking on people and society, in if merit was well defined. By Goodhart's law, "when a measure becomes a target, it ceases to be a good measure", meaning that agents will do anything they can to improve their value according to the chosen metric, and thus the metric will no longer measure what it was intended to. Moreover, the strategies employed by the agents to improve their value might have adversarial effects on society. For instance, "if a government sets a target for reducing crime rates, police officers may be incentivized to focus on low-level offenses instead of tackling more serious crimes"[9]. Finally, having competitions at every level of society, school choice, college admission, job market, and many others, is not the only possible social organization. When the term meritocracy emerged in the 1950's, in writings by Young, Floud, Fox or Lamartine, it had a very negative connotation, even though it is now used by many political figures as a positive value. As Meurice [Meu23] stated, "what is a winner, if not a creator of losers? What meaning can there be in enjoying someone else's failure as a sign of our own power?[...] It is this structure that legitimates all iniquities.", highlighting the link between competition and unfairness. In the same spirit, for Jacquard, "the current collective morality makes us believe that what is important is dominating others, struggling, winning. We are in a society of competition. But a winner is a producer of losers. We need to rebuild a human society where competition will be eliminated".

[9]Example taken from Dinker Charak's blog: `https://www.dinker.in/goodharts-law-when-a-measure-becomes-a-target-it-ceases-to-be-a-good-measure/`

Bibliography

[Abd05] Atila Abdulkadiroğlu. “College admissions with affirmative action”. In: *International Journal of Game Theory* 33.4 (2005), pp. 535–549 (cit. on pp. 20, 66).

[ACY15] Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda. “Expanding “Choice" in School Choice”. In: *American Economic Journal: Microeconomics* 7.1 (2015), pp. 1–42 (cit. on pp. 65, 66, 101).

[Aha+21] Narges Ahani, Paul Gölz, Ariel D. Procaccia, Alexander Teytelboym, and Andrew C. Trapp. “Dynamic Placement in Refugee Resettlement”. In: *ACM Conference on Economics and Computation*. 2021 (cit. on p. 2).

[AKR22] Peter Arcidiacono, Josh Kinsler, and Tyler Ransom. “Asian American Discrimination in Harvard Admissions”. In: *European Economic Review* 144 (2022), p. 104079 (cit. on p. 1).

[AL16] Eduardo M. Azevedo and Jacob D. Leshno. “A Supply and Demand Framework for Two-Sided Matching Markets”. In: *Journal of Political Economy* 124.5 (2016), pp. 1235–1268 (cit. on pp. 65, 66, 71–74, 101).

[AN20] Itai Ashlagi and Afshin Nikzad. “What matters in school choice tie-breaking? How competition guides design”. In: *Journal of Economic Theory* 190 (2020), p. 105120 (cit. on pp. 66, 83).

[ANR19] Itai Ashlagi, Afshin Nikzad, and Assaf Romm. “Assigning more students to their top choices: A comparison of tie-breaking rules”. In: *Games and Economic Behavior* 115 (2019), pp. 167–187 (cit. on pp. 65, 66, 83).

[APR09] Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth. “Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match”. In: *American Economic Review* 99.5 (2009), pp. 1954–1978 (cit. on pp. 3, 66).

[Arn22] Nick Arnosti. “A Continuum Model of Stable Matching with Finite Capacities”. In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. Boulder, CO, USA, 2022, p. 960 (cit. on pp. 66, 73).

[Arn23] Nick Arnosti. “Lottery Design for School Choice”. In: *Management Science* 69.1 (2023), pp. 244–259 (cit. on pp. 65, 66, 83).

[Arr73] Kenneth J. Arrow. “The theory of discrimination”. In: *Discrimination in Labor Markets*. Ed. by Orley Ashenfelter and Albert Rees. Princeton University Press, 1973, pp. 3–33 (cit. on p. 18).

[AS03] Atila Abdulkadiroğlu and Tayfun Sönmez. “School Choice: A Mechanism Design Approach”. In: *American Economic Review* 93.3 (2003), pp. 729–747 (cit. on p. 66).

[Ban+23] Sayan Bandyapadhyay, Fedor V. Fomin, Tanmay Inamdar, and Kirill Simonov. "Proportionally Fair Matching with Multiple Groups". In: *Graph-Theoretic Concepts in Computer Science*. Ed. by Daniël Paulusma and Bernard" Ries. 2023 (cit. on pp. 19, 39).

[BCT85] Robert E Bixby, William H Cunningham, and Donald M Topkis. "The partial order of a polymatroid extreme point". In: *Mathematics of Operations Research* 10 (1985), pp. 367–378 (cit. on pp. 32, 54).

[BE98] Allan Borodin and Ran El-Yaniv. *Online computation and competitive analysis*. USA: Cambridge University Press, 1998 (cit. on p. 115).

[BEF21] Moshe Babaioff, Tomer Ezra, and Uriel Feige. "Fair and truthful mechanisms for dichotomous valuations". In: *AAAI*. 2021 (cit. on pp. 26, 36).

[Beh77] Fred Alois Behringer. "Lexicographic quasiconcave multiobjective programming". In: *Zeitschrift für Operations Research* 21 (1977), pp. 103–116 (cit. on p. 34).

[Bei+19] Xiaohui Bei, Xinhang Lu, Pasin Manurangsi, and Warut Suksompong. "The Price of Fairness for Indivisible Goods". In: *Theory of Computing Systems* (2019) (cit. on p. 24).

[Ben+21] Nawal Benabbou, Mithun Chakraborty, Ayumi Igarashi, and Yair Zick. "Finding fair and efficient allocations for matroid rank valuations". In: *ACM Conference on Economics and Computation*. 2021 (cit. on pp. 26, 36).

[Ber57] Claude Berge. "Two Theorems in Graph Theory". In: *Proceedings of the National Academy of Sciences of the United States of America* 43.9 (Sept. 1957), pp. 842–844 (cit. on p. 6).

[BFT11] Dimitris Bertsimas, Vivek F. Farias, and Nikolaos Trichakis. "The Price of Fairness". In: *Oper. Res.* 59 (2011), pp. 17–31 (cit. on p. 24).

[BH22] Angelina Brilliantova and Hadi Hosseini. "Fair Stable Matching Meets Correlated Preferences". In: *CoRR* abs/2201.12484 (2022) (cit. on p. 66).

[BHI22] J Aislinn Bohren, Peter Hull, and Alex Imas. *Systemic discrimination: Theory and measurement*. National Bureau of Economic Research working paper. 2022 (cit. on p. 64).

[BHN19] Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning*. 2019 (cit. on p. 17).

[Bla84] Charles E. Blair. "Every Finite Distributive Lattice Is a Set of Stable Matchings". In: *J. Comb. Theory A* 37.3 (1984), pp. 353–356 (cit. on p. 13).

[Blu90] Norbert Blum. "A New Approach to Maximum Matching in General Graphs". In: *International Colloquium on Automata, Languages and Programming*. 1990 (cit. on p. 2).

[BM01] Anna Bogomolnaia and Hervé Moulin. "A new solution to the random assignment problem". In: *Journal of Economic theory* 100 (2001), pp. 295–328 (cit. on p. 34).

[BM04] Marianne Bertrand and Sendhil Mullainathan. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination". In: *American Economic Review* 94.4 (2004), pp. 991–1013 (cit. on p. 1).

[Bom+22] Rishi Bommasani, Kathleen A. Creel, Ananya Kumar, Dan Jurafsky, and Percy S Liang. "Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization?" In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh. Vol. 35. Curran Associates, Inc., 2022, pp. 3663–3678 (cit. on pp. 21, 67, 97).

[BP03] Lawrence Bodin and Aaron Panken. “High Tech for a Higher Authority: The Placement of Graduating Rabbis from Hebrew Union College—Jewish Institute of Religion”. In: *Interfaces* 33.3 (June 2003), pp. 1–11 (cit. on p. 3).

[BS20] Avrim Blum and Kevin Stangl. “Recovering from Biased Data: Can Fairness Constraints Improve Accuracy?” In: *Proceedings of the Symposium on Foundations of Responsible Computing (FORC 2020)*. 2020 (cit. on p. 17).

[BS99] Michel Balinski and Tayfun Sonmez. “A Tale of Two Mechanisms: Student Placement”. In: *Journal of Economic Theory* 84.1 (1999), pp. 73–94 (cit. on pp. 66, 71).

[Bud11] Eric Budish. “The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes”. In: *Journal of Political Economy* 119 (2011), pp. 1061–1103 (cit. on p. 25).

[BV22] Siddharth Barman and Paritosh Verma. “Truthful and fair mechanisms for matroid-rank valuations”. In: *AAAI*. 2022 (cit. on p. 26).

[Car+19] Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D Procaccia, Nisarg Shah, and Junxing Wang. “The unreasonable fairness of maximum Nash welfare”. In: *ACM Conference on Economics and Computation*. 2019 (cit. on p. 25).

[Cas+24] Rémi Castera, Felipe Garrido-Lucero, Mathieu Molina, Simon Mauras, Patrick Loiseau, and Vianney Perchet. “The Price of Fairness in Bipartite Matching”. In: *arXiv preprint n°2403.00397* (2024) (cit. on p. 24).

[CH11] Bala Chandran and Dorit Hochbaum. “Practical and theoretical improvements for bipartite matching using the pseudoflow algorithm”. In: *Computing Research Repository - CORR* (May 2011) (cit. on p. 8).

[Chi+19] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. “Matroids, Matchings, and Fairness”. In: *International Conference on Artificial Intelligence and Statistics*. 2019 (cit. on pp. 19, 21, 39, 47).

[Cho17] Alexandra Chouldechova. “Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments”. In: *Big Data* 5.2 (June 2017), pp. 153–163 (cit. on p. 17).

[CLP22] Rémi Castera, Patrick Loiseau, and Bary S.R. Pradelski. “Statistical Discrimination in Stable Matchings”. In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. EC ’22. Boulder, CO, USA: Association for Computing Machinery, 2022, pp. 373–374 (cit. on p. 63).

[CLP24] Rémi Castera, Patrick Loiseau, and Bary Pradelski. “Correlation of Rankings in Matching Markets”. In: *Hal preprint hal-03672270v6* (Feb. 2024) (cit. on p. 63).

[CLS14] H. Chade, G. Lewis, and L. Smith. “Student Portfolios and the College Admissions Problem”. In: *The Review of Economic Studies* 81.3 (2014), pp. 971–1002 (cit. on p. 66).

[Cor+22] José Correa, Natalie Epstein, Rafael Epstein, et al. “School Choice in Chile”. In: *Operations Research* 70.2 (2022), pp. 1066–1087 (cit. on p. 84).

[CS98] Melvyn G. Coles and Eric Smith. “Marketplaces and Matching”. In: *International Economic Review* 39 (1998), pp. 239–254 (cit. on p. 1).

[CT19] Yeon-Koo Che and Olivier Tercieux. “Efficiency and Stability in Large Matching Markets”. In: *Journal of Political Economy* 127.5 (2019), pp. 2301–2342 (cit. on p. 66).

[Der88] Ulrich Derigs. “Solving non-bipartite matching problems via shortest path techniques”. In: *Annals of Operations Research* 13 (1988), pp. 225–261 (cit. on p. 1).

[Dev+23] Siddartha Devic, David Kempe, Vatsal Sharan, and Aleksandra Korolova. "Fairness in matching under uncertainty". In: *Proceedings of the 40th International Conference on Machine Learning*. ICML'23. Honolulu, Hawaii, USA, 2023 (cit. on pp. 15, 21).

[DF81] L. E. Dubins and D. A. Freedman. "Machiavelli and the Gale-Shapley Algorithm". In: *The American Mathematical Monthly* 88.7 (1981) (cit. on p. 12).

[dG77] Claude d'Aspremont and Louis Gevers. "Equity and the Informational Basis of Collective Choice". In: *The Review of Economic Studies* 44.2 (1977), pp. 199–209 (cit. on p. 33).

[DG78] Robert Deschamps and Louis Gevers. "Leximin and utilitarian rules: A joint characterization". In: *Journal of Economic Theory* 17 (1978), pp. 143–163 (cit. on p. 33).

[DKT23a] David Delacrétaz, Scott Duke Kominers, and Alexander Teytelboym. "Matching Mechanisms for Refugee Resettlement". In: *American Economic Review* (2023) (cit. on p. 2).

[DKT23b] David Delacrétaz, Scott Duke Kominers, and Alexander Teytelboym. "Matching Mechanisms for Refugee Resettlement". In: *American Economic Review* 113.10 (Oct. 2023), pp. 2689–2717 (cit. on p. 20).

[DPS14] John P. Dickerson, Ariel D. Procaccia, and Tuomas Sandholm. "Price of fairness in kidney exchange". In: *Adaptive Agents and Multi-Agent Systems*. 2014 (cit. on p. 24).

[DPS20] Umut Dur, Parag A. Pathak, and Tayfun Sönmez. "Explicit vs. statistical targeting in affirmative action: Theory and evidence from Chicago's exam schools". In: *Journal of Economic Theory* 187 (2020), p. 104996 (cit. on p. 20).

[Dwo+12] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. "Fairness through awareness". In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ACM, 2012 (cit. on p. 15).

[Edm65] Jack Edmonds. "Paths, Trees, and Flowers". In: *Canadian Journal of Mathematics* 17 (1965), pp. 449–467 (cit. on p. 5).

[Edm70] Jack Edmonds. "Submodular functions, matroids, and certain polyhedra". In: *Combinatorial Structures and Their Applications* (1970), pp. 69–87 (cit. on pp. 25, 31).

[Edm79] Jack Edmonds. "Matroid intersection". In: *Annals of discrete Mathematics*. Vol. 4. Elsevier, 1979, pp. 39–49 (cit. on pp. 26, 31).

[EE08] Aytek Erdil and Haluk Ergin. "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice". In: *American Economic Review* 98.3 (June 2008), pp. 669–89 (cit. on p. 66).

[EF65] Jack Edmonds and Delbert Ray Fulkerson. "Transversals and Matroid Partition". In: *Journal of Research of the National Bureau of Standards Section B Mathematics and Mathematical Physics* (1965), p. 147 (cit. on p. 25).

[EIV23] Federico Echenique, Nicole Immorlica, and Vijay Vazirani. *Online and Matching-Based Market Design*. Cambridge University Press, 2023 (cit. on p. 1).

[Eme+20] Vitalii Emelianov, Nicolas Gast, Krishna P. Gummadi, and Patrick Loiseau. "On Fair Selection in the Presence of Implicit Variance". In: *Proceedings of the 21st ACM Conference on Economics and Computation (EC)*. 2020 (cit. on p. 18).

[Eme+22] Vitalii Emelianov, Nicolas Gast, Krishna P. Gummadi, and Patrick Loiseau. "On fair selection in the presence of implicit and differential variance". In: *Artificial Intelligence* 302 (2022), p. 103609 (cit. on pp. 17, 18, 68).

[ES06] Haluk Ergin and Tayfun Sönmez. "Games of school choice under the Boston mechanism". In: *Journal of Public Economics* 90.1-2 (2006), pp. 215–237 (cit. on p. 66).

[Esm+22] Seyed-Alireza Esmaeili, Sharmila Duppala, Vedant Nanda, Aravind Srinivasan, and John P. Dickerson. "Rawlsian Fairness in Online Bipartite Matching: Two-sided, Group, and Individual". In: *Adaptive Agents and Multi-Agent Systems*. 2022 (cit. on p. 19).

[Fea13] Joe Feagin. *Systemic racism: A theory of oppression*. Routledge, 2013 (cit. on p. 64).

[FF56] Lester Randolph Ford and Delbert R Fulkerson. "Maximal flow through a network". In: *Canadian journal of Mathematics* 8 (1956), pp. 399–404 (cit. on pp. 7, 30, 133).

[FK16] Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, 2016 (cit. on pp. 45, 60, 61).

[Fle21] Will Fleisher. "What's Fair about Individual Fairness?" In: AIES '21. Association for Computing Machinery, 2021, pp. 480–490 (cit. on p. 15).

[Fol66] Duncan Karl Foley. *Resource allocation and the public sector*. Yale University, 1966 (cit. on p. 25).

[Fre+23] Danielle N. Freund, Thodoris Lykouris, Elisabeth Paulson, Bradley Sturt, and Wen-Yu Weng. "Group fairness in dynamic refugee assignment". In: *ACM Conference on Economics and Computation*. 2023 (cit. on p. 2).

[GLM21] Nikhil Garg, Hannah Li, and Faidra Monachou. "Standardized Tests and Affirmative Action: The Role of Bias and Variance". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. 2021 (cit. on pp. 18, 68).

[Gol21] Paweł Gola. "Supply and Demand in a Two-Sector Matching Model". In: *Journal of Political Economy* 129.3 (2021), pp. 940–978 (cit. on p. 66).

[GS13] David Gale and Lloyd S. Shapley. "College Admissions and the Stability of Marriage". In: *The American Mathematical Monthly* 120 (2013), pp. 386–391 (cit. on p. 1).

[GS62] David Gale and Lloyd. S. Shapley. "College Admissions and the Stability of Marriage". In: *The American Mathematical Monthly* 69.1 (1962), pp. 9–15 (cit. on pp. 3, 9, 10, 13, 20, 65, 66, 101, 133).

[GT91] Harold N. Gabow and Robert E. Tarjan. "Faster scaling algorithms for general graph matching problems". In: *J. ACM* 38.4 (Oct. 1991), pp. 815–853 (cit. on p. 2).

[Hal35] P. Hall. "On Representatives of Subsets". In: *Journal of the London Mathematical Society* s1-10.1 (1935), pp. 26–30 (cit. on p. 2).

[Har+16] Moritz Hardt, Eric Price, Eric Price, and Nati Srebro. "Equality of Opportunity in Supervised Learning". In: *Advances in Neural Information Processing Systems*. Ed. by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett. Vol. 29. Curran Associates, Inc., 2016 (cit. on p. 17).

[HH02] Jürgen Herzog and Takayuki Hibi. "Discrete polymatroids". In: *Journal of Algebraic Combinatorics* 16 (2002), pp. 239–268 (cit. on pp. 25, 28, 32).

[HK73] John E. Hopcroft and Richard M. Karp. "An $n^{5/2}$ Algorithm for Maximum Matchings in Bipartite Graphs". In: *SIAM Journal on Computing* 2.4 (1973), pp. 225–231 (cit. on pp. 8, 133).

[Hos+23] Hadi Hosseini, Zhiyi Huang, Ayumi Igarashi, and Nisarg Shah. "Class fairness in online matching". In: *AAAI*. 2023 (cit. on p. 19).

[HRW00] Wouter J. den Haan, Garey Ramey, and Joel Watson. "Liquidity Flows and Fragility of Business Enterprises". In: *Macroeconomics eJournal* (2000) (cit. on p. 1).

[Hum40] David Hume. *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. London, 1740 (cit. on p. 116).

[Kar73] Alexander Karzanov. “Tochnaya otsenka algoritma nakhozhdeniya maksimal’nogo potoka, primenennogo k zadache o predstavitelyakh”. In: *Seminar on Combinatorial Mathematics*. Jan. 1973, pp. 66–70 (cit. on pp. 8, 133).

[KK15] Yuichiro Kamada and Fuhito Kojima. “Efficient Matching under Distributional Constraints: Theory and Applications”. In: *American Economic Review* 105.1 (2015), pp. 67–99 (cit. on p. 20).

[KK23] Yuichiro Kamada and Fuhito Kojima. “Fair Matching under Constraints: Theory and Applications”. In: *Review of Economic Studies* (2023) (cit. on p. 20).

[KMR16] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. *Inherent Trade-Offs in the Fair Determination of Risk Scores*. 2016 (cit. on p. 17).

[Knu76] D.E. Knuth. *Mariages stables et leurs relations avec d’autres problèmes combinatoires: introduction à l’analyse mathématique des algorithmes*. Chaire Aisenstadt. Presses de l’Université de Montréal, 1976 (cit. on p. 13).

[Kőn31] Dénes Kőnig. “Gráfok és mátrixok”. In: *Matematikai és Fizikai Lapok* 38 (1931), pp. 116–119 (cit. on p. 2).

[KR18] Jon Kleinberg and Manish Raghavan. “Selection Problems in the Presence of Implicit Bias”. In: *Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS ’18)*. 2018 (cit. on p. 18).

[KR21] Jon Kleinberg and Manish Raghavan. “Algorithmic monoculture and social welfare”. In: *Proceedings of the National Academy of Sciences* 118.22 (2021) (cit. on p. 67).

[Kri+22] Prem Krishnaa, Girija Limaye, Meghana Nasre, and Prajakta Nimbhorkar. “Envy-freeness and relaxed stability: hardness and approximation algorithms”. In: *Journal of Combinatorial Optimization* 45 (Dec. 2022) (cit. on p. 20).

[KRY22] Gili Karni, Guy N. Rothblum, and Gal Yona. “On Fairness and Stability in Two-Sided Matchings”. In: *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Leibniz International Proceedings in Informatics (LIPIcs). 2022, 92:1–92:17 (cit. on p. 20).

[Kuh55] Harold W Kuhn. “The Hungarian method for the assignment problem”. In: *Naval research logistics quarterly* 2.1-2 (1955), pp. 83–97 (cit. on p. 39).

[Lip+04] Richard J Lipton, Evangelos Markakis, Elchanan Mossel, and Amin Saberi. “On approximately fair allocations of indivisible goods”. In: *ACM Conference on Electronic Commerce*. 2004 (cit. on p. 25).

[LL20] Jacob D Leshno and Irene Lo. “The Cutoff Structure of Top Trading Cycles in School Choice”. In: *The Review of Economic Studies* 88.4 (Nov. 2020), pp. 1582–1623 (cit. on pp. 66, 114).

[Meh13] Aranyak Mehta. “Online Matching and Ad Allocation”. In: *Found. Trends Theor. Comput. Sci.* 8 (2013), pp. 265–368 (cit. on p. 1).

[Mel22] Ursula Mello. “Centralized Admissions, Affirmative Action, and Access of Low-Income Students to Higher Education”. In: *American Economic Journal: Economic Policy* 14.3 (2022), pp. 166–97 (cit. on p. 65).

[Meu23] Guillaume Meurice. *Petit éloge de la médiocrité*. Les Pérégrines, 2023 (cit. on p. 116).

[ML22] Mathieu Molina and Patrick Loiseau. "Bounding and Approximating Intersectional Fairness through Marginal Fairness". In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh. Vol. 35. Curran Associates, Inc., 2022, pp. 16796–16807 (cit. on p. 114).

[MS15] Bruce Maggs and Ramesh Sitaraman. "Algorithmic Nuggets in Content Delivery". In: *ACM SIGCOMM Computer Communication Review* 45 (July 2015), pp. 52–66 (cit. on p. 3).

[Mul23] Thomas Mulligan. "Meritocracy". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta and Uri Nodelman. Fall 2023. Metaphysics Research Lab, Stanford University, 2023 (cit. on p. 116).

[MV80] Silvio Micali and Vijay V. Vazirani. "An O(v|v| c |E|) algorithm for finding maximum matching in general graphs". In: *21st Annual Symposium on Foundations of Computer Science (sfcs 1980)* (1980), pp. 17–27 (cit. on p. 2).

[MW71] D. G. McVitie and L. B. Wilson. "The Stable Marriage Problem". In: *Commun. ACM* 14.7 (July 1971), pp. 486–490 (cit. on p. 13).

[MX20] Will Ma and Pan Xu. "Group-level Fairness Maximization in Online Bipartite Matching". In: *Adaptive Agents and Multi-Agent Systems*. 2020 (cit. on p. 19).

[MXX23] Will Ma, Pan Xu, and Yifan Xu. "Fairness Maximization among Offline Agents in Online-Matching Markets". In: *ACM Transactions on Economics and Computation* 10 (Feb. 2023) (cit. on p. 19).

[Nis+07] Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, 2007 (cit. on p. 1).

[NR19] Alexandra Niessen-Ruenzi and Stefan Ruenzi. "Sex Matters: Gender Bias in the Mutual Fund Industry". In: *Management Science* 65.7 (2019), pp. 3001–3025 (cit. on p. 1).

[Oxl22] James Oxley. "Matroid theory". In: *Handbook of the Tutte Polynomial and Related Topics*. Chapman and Hall/CRC, 2022, pp. 44–85 (cit. on p. 25).

[PG23] Kenny Peng and Nikhil Garg. *Monoculture in Matching Markets*. 2023. arXiv: `2312.09841 [cs.GT]` (cit. on p. 87).

[Phe72] Edmund Phelps. "The Statistical Theory of Racism and Sexism". In: *American Economic Review* 62.4 (1972), pp. 659–661 (cit. on p. 18).

[Pin96] Fred L Pincus. "Discrimination comes in many forms: Individual, institutional, and structural". In: *American Behavioral Scientist* 40.2 (1996), pp. 186–194 (cit. on p. 64).

[Roe98] John Roemer. *Equality of Opportunity*. Harvard University Press, 1998 (cit. on p. 15).

[Rot82] Alvin E. Roth. "The Economics of Matching: Stability and Incentives". In: *Mathematics of Operations Research* 7.4 (1982), pp. 617–628 (cit. on p. 12).

[Rot84] Alvin E. Roth. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory". In: *Journal of Political Economy* 92.6 (1984), pp. 991–1016 (cit. on p. 13).

[Rot85] Alvin E Roth. "The college admissions problem is not equivalent to the marriage problem". In: *Journal of Economic Theory* 36.2 (1985), pp. 277–288 (cit. on p. 75).

[Rot86] Alvin E. Roth. "On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets". In: *Econometrica* 54.2 (1986), pp. 425–427 (cit. on pp. 3, 13).

[RS92] Alvin E. Roth and Marilda Sotomayor. *Two-sided matching: A study in game-theoretic modeling and analysis*. 1st ed. Handbook of game theory with economic applications. Cambridge University Press, 1992 (cit. on pp. 2, 3).

[RSU05] Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver. "Pairwise kidney exchange". In: *Journal of Economic Theory* 125.2 (2005), pp. 151–188 (cit. on p. 25).

[RV90] A. E. Roth and H. Vande Vate. "Random Paths to Stability in Two-Sided Matching". In: *Econometrica* 58 (1990), pp. 1475–1480 (cit. on p. 101).

[San+21] Govind S. Sankar, Anand Louis, Meghana Nasre, and Prajakta Nimbhorkar. "Matchings with Group Fairness Constraints: Online and Offline Algorithms". In: *IJCAI*. 2021 (cit. on p. 19).

[Sca84] Marco Scarsini. "On measures of concordance". eng. In: *Stochastica* 8.3 (1984), pp. 201–218 (cit. on pp. 70, 100).

[Sen17] Amartya Sen. *Collective Choice and Social Welfare*. Harvard University Press, 2017 (cit. on p. 33).

[Skl59] M. Sklar. "Fonctions de répartition à N dimensions et leurs marges". In: *Annales de l'ISUP* VIII.3 (1959), pp. 229–231 (cit. on p. 68).

[SS74] Lloyd Shapley and Herbert Scarf. "On cores and indivisibility". In: *Journal of Mathematical Economics* 1.1 (1974), pp. 23–37 (cit. on p. 66).

[Ste49] Hugo Steinhaus. "Sur la division pragmatique". In: *Econometrica: Journal of the Econometric Society* (1949), pp. 315–319 (cit. on p. 25).

[SY22] Tayfun Sönmez and M. Yenmez. "Affirmative Action in India via Vertical, Horizontal, and Overlapping Reservations". In: *Econometrica* 90 (Jan. 2022), pp. 1143–1176 (cit. on p. 25).

[Var74] Hal R Varian. "Equity, envy, and efficiency". In: *Journal of Economic Theory* 9.1 (1974), pp. 63–91 (cit. on p. 25).

[VZ23] Vignesh Viswanathan and Yair Zick. "A general framework for fair allocation under matroid rank valuations". In: *ACM Conference on Economics and Computation*. 2023 (cit. on p. 26).

[Wel85] Dietrich Weller. "Fair division of a measurable space". In: *Journal of Mathematical Economics* 14 (1985), pp. 5–17 (cit. on p. 25).

[Wes18] D.B. West. *Introduction to Graph Theory*. Pearson, 2018 (cit. on p. 2).

[Yen13] M. Bumin Yenmez. "Incentive-Compatible Matching Mechanisms: Consistency with Various Stability Notions". In: *American Economic Journal: Microeconomics* 5.4 (2013), pp. 120–41 (cit. on p. 66).

[You94] H. Peyton Young. *Equity: In Theory and Practice*. Princeton University Press, 1994 (cit. on p. 15).

Websites and newspaper articles

[Com23] European Comission. *Commission proposes new measures on skills and talent to help address critical labour shortages*. 2023. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_5740 (cit. on p. 2).

[NRM] NRMP. *National Resident Matching Program website*. URL: http://www.nrmp.org (cit. on p. 3).

[Ref23] United Nations High Commissioner for Refugees. *Resettlement Fact Sheet*. 2023. URL: https://www.unhcr.org/media/resettlement-fact-sheet-2023 (cit. on pp. 2, 133).

Et là normalement il faut une citation latine mais pff... J'en ai marre!

François Rollin

Parce que si on commence à pendre les gens qui gagnent leur vie en racontant des conneries, il est possible que je ne termine pas la semaine.

Guillaume Meurice

List of Figures

List of Tables

Résumé long de la thèse en français

Introduction

De nombreuses situations d'allocation de ressources peuvent être représentées comme des problèmes d'appariement, un modèle théorique basé sur la théorie des graphes et utilisé dans les domaines de la recherche opérationnelle, de l'informatique et de l'économie. L'objectif des problèmes d'appariement est de trouver un "bon" appariement, par exemple trouver le plus grand appariement possible. Les inégalités entre groupes démographiques sont présentes dans tous les domaines de la vie publique, par exemple à l'embauche. Les problèmes d'appariement sont particulièrement concernés, d'autant plus qu'ils sont liés à des sujets sensibles comme les admissions à l'université ou l'accueil et la répartition de réfugiés [Ref23]. Dans cette thèse, j'étudie la question de l'équité dans les problèmes d'appariement et le potentiel compromis entre efficacité et équité. Ce premier chapitre introduit introduit les fondements de la théorie de l'appariement, et discute des différentes notions d'équité existantes dans la littérature.

Étant donné un graphe, un appariement est un ensemble d'arêtes qui ne partagent aucun sommet. En particulier, dans un graphe biparti, il s'agit de faire des paires entre éléments situés de part et d'autre du graphe. UN graphe de taille maximum peut être trouvé en temps polynomial grâce à de nombreux algorithmes [FF56; HK73; Kar73]. Quand les sommets du graphe sont de agents avec des préférences sur les partenaires potentiels, on recherche souvent des appariements stables. Un appariement est stable si il ne peut pas être agrandi et si aucune paire d'agents non appariés ne se préfèrent l'un l'autre à leurs partenaires actuels. Ce modèle peut également s'étendre pour modéliser un problème d'admission à l'université où les universités peuvent admettre plusieurs étudiants, dans la limite d'une capacité fixée. Un appariement stable peut là encore être trouvé en temps polynomial grâce à l'algorithme de Gale et Shapley [GS62].

L'équité peut être définie de nombreuse manière. Dans la littérature, on peut distinguer deux grandes catégories de notions d'équité: l'équité individuelle et l'équité de groupe. L'équité individuelle est le principe selon lequel des individus similaires devraient obtenir des résultats similaires (dans un processus de notation ou d'appariement. Cette définition est assez vague et nécessite de définir la similarité entre les individus et entre les résultats possibles, et dans la majorité des applications de prévient pas les discriminations envers certains groupes de la population. C'est là qu'intervient la deuxième catégorie, l'équité de groupe. Il s'agit cette fois de s'assurer que les groupes de la population (qui doivent être définis au préalable) soient traités de la même manière, par exemple, en imposant à une entreprise d'embaucher autant d'hommes que de femmes. L'équité de groupe ne prend pas en compte les individus, seulement des statistiques de groupe.

Appariement sans préférences: Matroïdes et prix de l'équité

On considère un graphe biparti $G = (U, V, E)$, où les sommets de V sont divisés en K groupes: $V = \bigcup_{i \in [K]} G_i$, $G_i \cap G_j = \emptyset$. On note $\mathcal{M}(\mathcal{G})$ l'ensemble des appariements possibles sur le graphe $\mathcal{G}$. On introduit une représentation géométrique de cet ensemble des appariements au travers de la fonction $X : \mathcal{M} \to \mathbb{R}^K$, $\mu \mapsto= (X_1(\mu), X_2(\mu), ..., X_K(\mu))$, où $X_i(\mu) := \sum_{u \in U} \sum_{v \in G_i} \mu(u, v)$, qui compte le nombre d'agents appariés dans chaque groupe. On montre que $X(\mathcal{M}(\mathcal{G}))$ est un polymatroïde discret, ce qui implique que l'ensemble des appariements maximaux correspond au front de Pareto du polymatroïde, donnant une caractérisation précise des appariements maximaux.

Nous introduisons une classe de contraintes d'équité de groupe qui inclut de nombreuses contraintes classiques de la littérature, que nous appelons équité pondérée. Étant donné un vecteur de poids $(w_1, \ldots, w_K)$, un appariement μ est w-équitable si les $X(\mu)_i/w_i$ sont égaux pour tout $i \in [K]$. Pour un w fixé, la question est alors de savoir s'il existe un appariement qui soit à la fois maximum et w-équitable, et si ce n'est pas le cas, quel est la taille du plus grand appariement x-équitable comparé au plus grand appariement sans contrainte. Le ratio entre ces deux grandeurs est ce que nous appelons le prix de l'équité (PoF pour Price of Fairness). Nous nous concentrons sur une contrainte particulière appelée équité d'opportunité, et cherchons à majorer le prix de l'équité. Nous obtenons des bornes qui sont linéaires en K, en particulier, le PoF est toujours inférieur à $K - 1$, et donc avec deux groupes il est toujours égal à 1, ce qui implique qu'il existe toujours un appariement maximal qui est équitable. Pour aller plus loin, nous étudions également des graphes aléatoires et montrons que le PoF est proche de 1, ce qui montre que les situations ou il est élevé correspondent à des cas très particuliers et dans des applications concrètes où les graphes sont assimilables à des graphes aléatoires le prix de l'équité est faible.

Appariement avec préférences: le rôle de la corrélation des priorités

Dans ce chapitre, on étudie le modèle d'admission à l'université, c'est à dire un problème d'appariement entre des étudiants et des universités où chaque étudiant à un ordre de préférence pour les universités et chaque université a un classement des étudiants. On considère un modèle simplifié avec deux universités A et B et un continuum d'étudiants S. L'ensemble des étudiants est divisé en K groupes $G_1, \ldots, G_K$ comme dans le chapitre précédent. On s'intéresse à la corrélation entre les deux classements des étudiants produits par les universités A et B: chaque étudiant a une note attribuée par chaque université, W^A et W^B, et au sein de chaque groupe, les vecteurs (W^A, W^B) ont une distribution potentiellement différente. On extrait de chacune de ces distributions sa copule (la distribution obtenue en transformant les marginales en distributions uniformes sur $[0, 1]$), et on suppose qu'il existe un famille de copule $(H_\theta)_{\theta\in\Theta}$ telle que pour chaque groupe G_i il existe θ_i tel que la distribution des notes de G_i est H_{θ_i}. Si cette famille vérifie une hypothèse qu'on appelle cohérence, alors les θ_i représentent la corrélation de la distribution pour chaque groupe, et on peut étudier l'effet d'avoir différents niveaux de corrélation sur la qualité de l'appariement stable et sur les potentielles inégalités entre les groupes.

De nombreux résultats sont obtenus dans cette direction. Tout d'abord, on remarque que la proportion d'étudiants obtenant leur premier voeu est la même dans tous les groupes. En revanche, la proporiton d'étudiants qui ne sont admis dans aucune université est plus importante dans les groupes qui ont une corrélation élevée. De plus, augmeter la corrélation pour un seul groupe a pour effet d'augmenter la proportion d'étudiants ayant leur premier voeu dans tous les groupes. Dans le même temps, si on considère deux groupes dont les proportions d'étudiants admis nulle part sont différentes, cette inégalité s'accroit si on augmente la corrélation du groupe désaventagé ou si on réduit celle du groupe avantagé; et réciproqument, l'inégalité se réduit en diminuant la corrélation du groupe désavantagé ou en augmentant celle du groupe avantagé.

Ces résultats peuvent également être appliqué au problème du bris d'égalité, c'est à dire quand les critères de sélection conduisent à de nombreuses égalités dans le classement qui doivent être tranchées aléatoirement. Il existe deux manières classiques de faire: soit chaque université classe aléatoirement les étudiants ex-aequo jusu'à obtenir un calssement sans égalités, ou alors un classment aléatoire des étudiants est crée par une autorité centrale et toutes les universités utilisent ce même classement pour trancher les égalités. Il a été montré par le passé qu'utiliser un classement commun pour trancher les égalités conduit à un nombre d'étudiants obtenant leur premier voeu plus élevé que les classements multiples. Notre modèle permet de retrouver ce résultat car les classements multiples correspondent au cas corrélation 0 et le classement commun à corrélation 1. Il ouvre également la voie à l'étude de problèmes de bris d'égalité avec une corrélation intermédiaire entre 0 et 1, ou même avec une corélation négative.

Conclusion

Je montre dans cette thèse que les inégalités entre groupes arrivent systématiquement si on se contente de chercher un appariement maximum ou stable (suivant le problème étudié) sans prêter attention à l'équité. En revanche, je montre également qu'il existe souvent des appariements qui sont à la fois efficaces et équitables, ou dans le pire des cas qu'il existe un appariement équitable proche de l'optimum, et qu'il n'y a donc souvent pas de compromis à faire entre équité et efficacité.

Les directions à explorer pour approfondir ce travail sont nombreuses. Les plus importantes semblent être:

- intégrer des notions d'équité intersectionnelle, c'est à dire la possibilité que certaines personnes appartiennent à plusieurs groupes, ce qui rend l'analyse des questions d'équité plus complexe;

- étendre les résultats du modèle avec préfférences à un autre appariement que l'appariement stable, nottament celui obtenu par l'algorithme Top Trading Cycles, qui est similaire par certains aspects;

- étudier les variantes dynamiques des deux modèles étudiés, c'est à dire le ,cas où les arrâtes du graphe apparaissent séquentiellement et des décisions doivent être prises à chaque pas de temps sans connaitre le graphe entièrement.

Enfin, puisque les inégalités dans les problèmes d'appariement viennent d'inégalités présentes à tous les niveaux de la société, leur prise en compte leur mitigation ne peut pas venir uniquement de moodèles mathématiques ou d'algorithmes, mais doit émaner de décisions collectives, éclairées d'une part par des analyses théoriques comme celle présentée dans cette thèse mais aussi par d'autres champs de recherche comme la sociologie ou l'économie empirique.

Colophon

This thesis was typeset with $\LaTeX\ 2_{\varepsilon}$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at `http://cleanthesis.der-ric.de/`.