On the Problem of Assigning PhD Grants

Katarína Cechlárová[1], **Laurent Gourvès**[2] and **Julien Lesca**[2]
[1]P.J. Šafárik University, Košice, Slovakia
[2]Université Paris-Dauphine, PSL, CNRS, LAMSADE, Paris, France
katarina.cechlarova@upjs.sk, {laurent.gourves, julien.lesca}@dauphine.fr

Abstract

In this paper, we study the problem of assigning PhD grants. Master students apply for PhD grants on different topics and the number of available grants is limited. In this problem, students have preferences over topics they applied to and the university has preferences over possible matchings of student/topic that satisfy the limited number of grants. The particularity of this framework is the uncertainty on a student's decision to accept or reject a topic offered to him. Without using probability to model uncertainty, we study the possibility of designing protocols of exchanges between the students and the university in order to construct a matching which is as close as possible to the optimal one i.e., the best achievable matching without uncertainty.

1 Introduction

Matching is a fundamental problem at the intersection of computer science, mathematics and economics [Manlove, 2013]. The restriction of this problem to bipartite graphs has attracted lot of attention in the research community. It has been used to match students with colleges [Goto *et al.*, 2016; Hamada *et al.*, 2017] or schools [Pathak, 2017], doctors with hospitals [Roth, 1984; Deng *et al.*, 2017], etc.

In our countries, the market for PhD grants is decentralized[1], and each university organizes its own procedure to hire PhD candidates. The procedure can even vary between the different disciplines of the same institution. Each university competes to obtain the most successful candidates in the market. A student may apply for different grants in various universities in order to increase his chances to obtain a grant. This phenomenon is well documented in economics where various papers study the strategical aspects of constructing a portfolio of applications maximizing the chances of obtaining a good position [Chade and Smith, 2006; Chade *et al.*, 2014]. One obvious conclusion of these works is that it is always safer for a candidate to apply for various positions. If a candidate receives multiple proposals, he can choose among them his most preferred one and reject the others. As a consequence, when a university makes a grant offer to a candidate in this decentralized market, it faces the uncertainty that this offer may be rejected in favor of an "outside option".

This uncertainty on the answer of a candidate to an offer is a special feature of a decentralized market. If the procedure for assigning students to grants was centralized among universities, and a stable matching algorithm such as deferred acceptance [Gale and Shapley, 1962] was in use, then no offer (or very few) would be rejected by students. Unfortunately imposing a centralized procedure to the various actors of this market is a difficult task. There are various examples of markets which have failed to impose centralized procedures, including gastroenterology fellowships [Niederle *et al.*, 2006], and psychology postdoctoral training [Bodin *et al.*, 2017]. Economists have identified various theoretical reasons to explain this phenomenon. One of them is that it is not always a dominating strategy for a university to participate in a centralized procedure [Ekmekci and Yenmez, 2014] and it may be better off by opting out. Another one is that even if (student optimal) deferred acceptance is strategy-proof for students, it can be manipulated in various ways by universities [Sönmez, 1999; Kesten, 2012]. In fact, there is no stable procedure which is non-manipulable by both students and universities [Roth, 1984]. Apart from these theoretical reasons, and unless it is enforced by law, an institute refrains to charge a central authority to choose their PhD candidates as it is perceived as a loss of autonomy.

In this paper, we study the design of efficient protocols between some candidates and a university in this decentralized market. In these protocols, the university offers grants to students who may accept or reject the proposals, depending on their unknown outside options. We assume that the university has cardinal preferences over possible matchings but we do not resort to probabilistic models to represent the uncertainty faced by the university over students' acceptance. Instead, we opt for a robust approach [Kouvelis and Yu, 1997] and we perform a worst case analysis by considering the greatest "regret" incurred by the protocol. In other words, our objective is to design protocols returning a matching which is as close as possible to the best achievable matching without uncertainty. This framework relies on approximation algorithms

[1]As it is the case for the postdoctoral job market for economists in North America [Roth, 2008], and for college admission in US, Korea and Japan [Che and Koh, 2016].

[Vazirani, 2001] but the barrier is the lack of information instead of the computational complexity.

Plan. The different aspects of the model are introduced in Section 2. Afterwards, two types of protocols are considered: *sequential protocols* where offers are issued one by one (Section 3), and *mixed protocols* where a first set of offers is made in parallel, followed by a sequence of offers (Section 4). Some proofs are omitted due to space limitation.

2 Preliminaries

The model proposed in this paper intends to represent the PhD grant allocation in LAMSADE. A bounded number of PhD grants, namely k, are offered by the university to the laboratory in order to hire PhD students. The members of the department are encouraged to propose topics, and master students apply to the grants under different topics. Then, a jury decides to whom will be given the grants and with which topics. In the current system, a grant is given to a student/topic pair. Let $\mathcal{S} = \{s_1, \ldots, s_n\}$ denote the set of students and $\mathcal{T} = \{t_1, \ldots, t_m\}$ the set of topics. Student s_i applies for a subset of topics $\mathcal{C}_i$ which is his *candidacy set*; $\mathcal{C}_i \subseteq \mathcal{T}$. The preference of student s_i over the topics of $\mathcal{C}_i$ is represented by total preorder $\succsim_i$ such that $t_j \succsim_i t_\ell$ if and only if student s_i weakly prefers t_j to t_ℓ. The fact that both $t_j \succsim_i t_\ell$ and $t_\ell \succsim_i t_j$ (both $t_j \succsim_i t_\ell$ and $t_\ell \not\succsim_i t_j$, respectively) hold is denoted by $t_j \sim_i t_\ell$ ($t_j \succ_i t_\ell$, respectively). We assume that students reveal their preferences over topics truthfully and do not try to manipulate. For any topic $t_j \in \mathcal{C}_i$, we denote by $R_i(t_j)$ the set of topics that are at least as good as t_j according to $\succsim_i$, and we denote by $E_i(t_j)$ the $\succsim_i$-equivalence class of $\mathcal{C}_i$ which contains t_j. More formally, $R_i(t_j) = \{t_\ell \in \mathcal{C}_i : t_\ell \succsim_i t_j\}$ and $E_i(t_j) = \{t_\ell \in \mathcal{C}_i : t_\ell \sim_i t_j\}$ hold.

Actually, each student s_i applies to a set of job opportunities $\mathcal{J}_i$ which is a superset of $\mathcal{C}_i$. The job opportunities of $\mathcal{J}_i \setminus \mathcal{C}_i$ are either PhD grants proposed by other universities or other types of positions. The preference $\succsim_i$ of student s_i may be extended to $\mathcal{J}_i$. Ultimately, student s_i will receive a subset of propositions $\mathcal{P}_i$ from $\mathcal{J}_i \setminus \mathcal{C}_i$. Let j_i denote the most preferred job opportunity of student s_i inside this set of propositions $\mathcal{P}_i$. Job opportunity j_i will define the set of *acceptable topics* $\mathcal{A}_i$ for student s_i, which is the subset of $\mathcal{C}_i$ containing topics as least as preferred as j_i according to $\succsim_i$. More formally, $\mathcal{A}_i = R_i(j_i)$. The jury knows the candidacy set $\mathcal{C}_i$ of each student s_i but it does not know his set of job opportunities $\mathcal{J}_i$ as well as his set of acceptable topics $\mathcal{A}_i$.

Let $\mathcal{C}$ and $\mathcal{A}$ denote sets $\{(s_i, t_j) \in \mathcal{S} \times \mathcal{T} : t_j \in \mathcal{C}_i\}$ and $\{(s_i, t_j) \in \mathcal{C} : t_j \in \mathcal{A}_i\}$, respectively. A matching is a subset of $\mathcal{C}$ containing at most k pairs and such that no student nor topic appears twice. Note that some topics may not be assigned if $k < m$. Matching M is *acceptable*, or is an $\mathcal{A}$-matching, if $M \subseteq \mathcal{A}$. The jury provides its preference over matchings through preorder $\succsim$ such that $M \succsim M'$ iff matching M is at least as good as matching M' according to the jury's preference. The preference of the jury may model multiple aspects of this problem, including the preference over topics that may be selected in the matching. The jury has preferences over topics for different reasons: a topic seems more promising than another, the supervising team has received (or not) a funding recently, the supervisor representing a topic has successfully supervised multiple PhD students, etc. These preferences may also model the synergy over student/topic pairs: a student may not be a good match for one topic but would be a perfect match for another one, according to his capabilities and skills. Matching M is *optimal* if there is no matching M' such that $M' \succ M$. M is $\mathcal{A}$-*optimal* if it is an $\mathcal{A}$-matching and there is no $\mathcal{A}$-matching M' such that $M' \succ M$. Our objective is to find an $\mathcal{A}$-optimal matching. Thereafter, we denote by M^* an $\mathcal{A}$-optimal matching.

In matching theory literature, it is common to assume that the preferences over subsets are responsive [Roth, 1985]. In this paper, we further assume that the jury's preferences are additive. This means that there exists a value function $v : \mathcal{C} \rightarrow \mathbb{R}_{\geq 0}$ such that for any matchings M and M', $M \succsim M'$ if and only if $v(M) \geq v(M')$, where $v(M) := \sum_{(s_i,t_j) \in M} v(s_i, t_j)$. The use of a value function leaves the possibility to quantify the relative importance of a student/topic pair. For example, if $v(s_i, t_j) > v(s_\ell, t_p) + v(s_r, t_q)$ holds then it means that the jury prefers to assign a single grant to student s_i with topic t_j rather than assigning two grants to student s_ℓ and s_r with topics t_p and t_q, respectively. Furthermore, this value function is a good measure of the worth of a matching according to the jury [Biro and Gudmondsson, 2019], and comparisons between matchings can be easily done.

For any subset $E \subseteq \mathcal{C}$, let G_E denote the edge-weighted bipartite graph with vertex set $\mathcal{S} \cup \mathcal{T}$, edge set E and where weights are provided by v. Our objective is therefore to find an optimal matching M^* in $G_{\mathcal{A}}$. This objective would be easy to achieve if $\mathcal{A}$ was known by the jury i.e., the jury knows which topics are acceptable to the students. However only $\mathcal{C}$ is known by the jury, and this can force them to make offers that will be rejected by the students. As we will see later, this can lead the jury to select a matching which is not an $\mathcal{A}$-optimal matching. Sometimes, no protocol can produce an $\mathcal{A}$-optimal matching. In that case, we will search for an $\mathcal{A}$-matching that is as close as possible to the $\mathcal{A}$-optimal matching M^*. An algorithm achieves an *approximation ratio* of $\alpha \leq 1$ if for any instance of the problem, it computes an $\mathcal{A}$-matching M such that $v(M)/v(M^*) \geq \alpha$.

A matching is constructed through an interaction between the candidates and the jury, called protocol, where the jury makes offers to students. In this paper, we consider two types of offers: *fixed* and *flexible offers*. In both cases, some topic, say t_j, is offered to some student, say s_i, i.e. s_i receives a grant for working on t_j. If student s_i accepts a fixed offer then he will be ultimately assigned to topic t_j. On the other hand, if student s_i accepts a flexible offer he may be assigned to another topic t_ℓ under the condition that $t_\ell \succsim_i t_j$ holds. More formally, *offer* $O_i \subseteq \mathcal{C}_i$ is a subset of topics proposed to student s_i. A fixed offer contains a single topic whereas a flexible offer may contain several topics. A flexible offer O_i will be of the form $R_i(t_j)$ for some $t_j \in \mathcal{C}_i$. If student s_i accepts offer O_i then he is guaranteed to receive one of the topics of O_i at the end of the protocol. By accepting offer O_i, student s_i also commits to accept any topic in O_i. We assume that student s_i accepts offer O_i if and only if it is a subset of $\mathcal{A}_i$. Once a student has accepted an offer, it is no

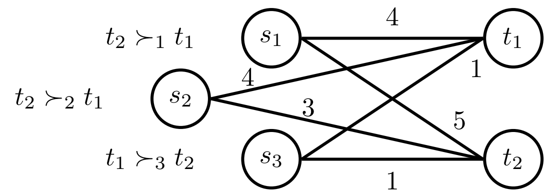

Figure 1: Example with 3 students, 2 topics and 2 grants

longer possible for the jury to make another offer to him. On the other hand, if a student rejects an offer then the jury can make other offers to him.

The use of fixed offers looks constraining for the jury, and one can ask whether only flexible offers can be considered. However, fixed offers somehow take into account the preference of the supervisor of a topic over students. Once an offer for a given topic t_j is accepted by student s_i, the supervisor of this topic may disagree to obtain another student s_ℓ in order to enlarge the global matching, and especially if student s_i seems to be more suited than student s_ℓ for topic t_j. With fixed offers, this situation cannot occur.

Example. 1 *Consider the instance described in Figure 1, and where two grants are available ($k = 2$). The set of applications $\mathcal{C}$ contains (s_i, t_j) for each $s_i \in \mathcal{S}$ and $t_j \in \mathcal{T}$. In other words, each student applies to each topic. However, student s_2 will only accept t_2 and student s_3 will not accept any topic. More formally, $\mathcal{A}_1 = \{t_1, t_2\}$, $\mathcal{A}_2 = \{t_2\}$ and $\mathcal{A}_3 = \emptyset$ hold. If the protocol starts by a fixed offer $O_1 = \{t_2\}$ to student s_1 then he accepts and t_2 must be assigned to him. In that case, only topic t_1 is still available. Topic t_1 is unacceptable for any other student than s_1, therefore any further offer will be rejected. The resulting matching $M = \{(s_1, t_2)\}$ has value 5 whereas the $\mathcal{A}$-optimal matching $M^* = \{(s_1, t_1), (s_2, t_2)\}$ has value 7. On the other hand, if the protocol starts with flexible offer $O_1 = \{t_1, t_2\}$ then student s_1 accepts. Whatever offers are made by the jury thereafter, only offer $O_2 = \{t_2\}$ will be accepted, leading to the $\mathcal{A}$-optimal matching M^* which is consistent with the offers.*

3 Sequential Protocols

In this section, the protocols under consideration consist of a single phase of exchanges between the jury and the students. During this phase, the jury makes offers one by one and waits for the answer of a student before making a new offer, even if the new offer is not addressed to the same student. We call this type of protocols *sequential*.

In Subsection 3.1, we provide a sequential protocol restricted to fixed offers which achieves the best possible approximation ratio of $1/2$. In Subsection 3.2, we propose a sequential protocol which provides an $\mathcal{A}$-optimal matching.

3.1 Sequential Protocols with Fixed Offers

The following proposition shows that an approximation ratio better than $1/2$ cannot be achieved with fixed offers.

Proposition. 1 *No sequential protocol restricted to fixed offers achieves a higher approximation ratio than $1/2$.*

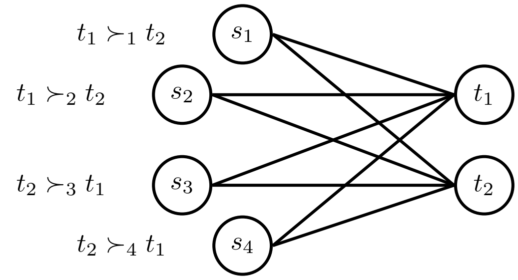

Figure 2: Value function v equals 1 for each pair of $\mathcal{C}$

Proof: Consider the instance described by Figure 2. Assume that two grants are available. Any sequential protocol restricted to fixed offers starts by proposing a fixed offer to one of the students. Assume that it first proposes $O_i = \{t_j\}$ to student s_i and he accepts. Let s_ℓ be another student who prefers t_j to the other topic. If $\mathcal{A}_i = \{t_1, t_2\}$, $\mathcal{A}_\ell = \{t_j\}$ and the set of acceptable topics is empty for the other students then $M^* = \{(s_i, t_{3-j}), (s_\ell, t_j)\}$ has value 2 whereas the protocol outputs $M = \{(s_i, t_j)\}$ with value 1 since no student will accept any further offer. Therefore, the protocol achieves an approximation ratio of at most $1/2$. ■

Consider now the following sequential protocol, called *Max*: by considering each pair (s_i, t_j) of $\mathcal{C}$ by non-increasing value of v, the jury proposes $O_i = \{t_j\}$ to student s_i if s_i did not accept a former offer until either k offers have been accepted or all pairs of $\mathcal{C}$ have been considered. It obviously returns an $\mathcal{A}$-matching. The following proposition shows that this protocol achieves the best possible approximation ratio[2].

Proposition. 2 *Max achieves an approximation ratio of $1/2$.*

3.2 Sequential Protocols with Flexible Offers

In this subsection, we present a sequential protocol which returns an $\mathcal{A}$-optimal matching. The underlying algorithm relies on the notion of *alternating path*. A path is assumed to be a sequence of edges belonging to $G_\mathcal{C}$. For a given matching M, an alternating path P with respect to M is a path in $G_\mathcal{C}$ whose edges belong alternatively to M and $\mathcal{C} \setminus M$. We assume that an alternating path is either a cycle or does not contain a cycle. In the former case, we refer to it as an *alternating cycle*. When it is clear from context, we do not refer to the assignment associated with an alternating path. An alternating path is *augmenting* for M (*decreasing* for M, resp.) if it starts from an unassigned student (assigned student, resp.) in M and ends with an unassigned topic (assigned topic, resp.) in M. Finally, an alternating path is an *even path* if it is neither augmenting, decreasing nor cycle.

For any two sets $A, B \subseteq \mathcal{C}$, let $A \Delta B = (A \setminus B) \cup (B \setminus A)$ denote the symmetric difference of A and B. Operator Δ can be used to construct a new matching from an alternating path as follows. If M is a matching and P an alternating path then $M \Delta P$ is a matching constructed from M by replacing the edges of $M \cap P$ with the edges of $P \setminus M$. Note that $M' =$

[2]Proposition 2 is closely related to the greedy approach for obtaining a $1/2$-approximation for the maximum matching problem [Karp *et al.*, 1990]. Its proof is similar and is therefore omitted.

Algorithm 1 Optimal sequential protocol

Input: Candidacy set $\mathcal{C}$, initial matching M^0.

```
M ← M^0.
Compute a M-optimal augmenting path P in G_C.
while v_M(P) ≥ 0 and |M| < k do
    Let s_i be the unassigned student in M visited by P.
    Propose O_i = C_i to student s_i.
    if s_i accepts O_i then
        M ← MΔP.
    else
        Let E ⊆ C_i be the ≿_i-equivalence class containing
        the least preferred topics according to ≿_i. Remove
        E from C_i and (s_i, t_j) from C for any t_j ∈ E.
    Compute M-optimal augmenting path P in G_C.
return M.
```

$M\Delta P$ implies $M\Delta M' = P$. More generally, if M and M' are two matchings then $M\Delta M'$ contains the set of alternating paths $P_1 \dots, P_s$ such that $M' = M\Delta P_1\Delta \dots \Delta P_s$. The *value of alternating path* P with respect to M is $v_M(P) = \sum_{(s_i,t_j)\in P\setminus M} v(s_i,t_j) - \sum_{(s_i,t_j)\in M\cap P} v(s_i,t_j)$, and it corresponds to the marginal value incurred by switching from M to $M\Delta P$. More formally, $v(M\Delta P) = v(M) + v_M(P)$. It is worth noting that $M' = M\Delta P$ implies that P is an alternating path for both M and M', and $v_M(P) = -v_{M'}(P)$. Augmenting path P for M is *M-optimal* if there is no augmenting path P' for M such that $v_M(P') > v_M(P)$.

The sequential protocol proposed to obtain an $\mathcal{A}$-optimal matching is presented in Algorithm 1. The initial matching M^0 used in this protocol is $\emptyset$. If no augmenting path exists in $G_\mathcal{C}$ then $P = \emptyset$ with value 0. In order to simplify notation, we assume that the candidacy set $\mathcal{C}$ can be updated, and set $\mathcal{C}_i$ for any $s_i \in \mathcal{S}$, derives from $\mathcal{C}$. In essence, Algorithm 1 starts with an empty matching and at every step augments it by proposing its candidacy set to the unmatched student visited by an optimal augmenting path. If he accepts the offer then he is included in the matching, and otherwise his candidacy set is revised by removing his least preferred topics.

Example. 2 *We apply Algorithm 1 to the instance of Example 1. Parameter M^0 is set to $\emptyset$, and initially $M = \emptyset$. The set of augmenting paths for M is therefore $\mathcal{C}$ and their values correspond to the ones returned by v. The M-optimal augmenting path is $P = \{(s_1,t_2)\}$ with value 5. Offer $O_1 = \mathcal{C}_1 = \{t_1,t_2\}$ is made to student s_1 who accepts since $\mathcal{A}_1 = \mathcal{C}_1$. Then, matching M becomes $M\Delta P = \{(s_1,t_2)\}$. The augmenting paths for M of size 1 are the edges $\{s_2,t_1\}$ and $\{s_3,t_1\}$. The augmenting paths for M of size 3 are $\{(s_1,t_1),(s_1,t_2),(s_2,t_2)\}$ and $\{(s_1,t_1),(s_1,t_2),(s_3,t_2)\}$ with values 2 and 0, respectively. The M-optimal augmenting path is $\{(s_2,t_1)\}$ with value 4. Offer $O_2 = \{t_1,t_2\}$ is made to student s_2 who rejects it. Since $t_2 \succ_2 t_1$, t_1 is removed from $\mathcal{C}_2$. During the next iteration, the M-optimal augmenting path is $\{(s_1,t_1),(s_1,t_2),(s_2,t_2)\}$. Offer $O_2 = \{t_1\}$ is made to student s_2 who accepts since $O_2 = \mathcal{A}_2$. Then, M becomes $M\Delta P = \{(s_1,t_1),(s_2,t_2)\}$.*

The next result states that the corresponding sequential protocol returns an $\mathcal{A}$-optimal matching. It is a corollary to Proposition 6 which is given and proved in Subsection 4.2.

Corollary. 1 *Algorithm 1 describes a sequential protocol which returns an $\mathcal{A}$-optimal matching when $M^0 = \emptyset$.*

An M-optimal augmenting path can be computed through the *incremental digraph* which is an oriented version of $G_\mathcal{C}$ where each edge (s_i,t_j) of M is oriented from topic to student with weight $-v(s_i,t_j)$, and each edge (s_i,t_j) of $\mathcal{C}\setminus M$ is oriented from student to topic with weight $v(s_i,t_j)$. An oriented path from an unmatched student to an unmatched topic corresponds to an augmenting path for M. If P is an alternating path then the sum of its weights in the incremental digraph is equal to $v_M(P)$. Hence, finding an optimal augmenting path for M amounts to finding an oriented path in the incremental digraph of maximum weight. This problem is in general NP-hard. However, it is easy to show that the incremental digraph does not contain a cycle with positive weight. Therefore, a generalized version of Bellman-Ford algorithm [Gondran and Minoux, 2008] can be used to compute the oriented path of maximum weight.

4 Mixed Protocols

In practice, a sequential protocol may require a lot of time in order to fill the k grants. This could be problematic if not enough offers are accepted before the market closure (for PhD grants' allocation, it would be before the start of the academic year). This problem, called congestion, is well documented in the matching literature [Roth, 2008]. In order to reduce the duration of the process, it is natural to try to make multiple offers in parallel. However, since the jury commits to give a grant to every accepted offer, parallel offers should be issued carefully. Moreover, we will see that parallelization has an impact on the performance guarantee.

The *mixed protocols* under consideration in this section comprise two different phases of exchanges between the jury and the students. As opposed to sequential protocols, multiple offers are made in parallel, but only during the first phase. Phase one ends when the students who received an offer have answered. A second phase is initiated if at least one student declined. In the second phase, offers are made to the students who did not receive any offer or declined. As for the sequential protocol, the jury makes offers sequentially: one waits for a student's answer before making a new offer. Mixed protocols are similar to the procedures used in the American entry-level market for clinical psychologists [Roth and Xing, 1997].

During the first phase, a set of offers $\mathcal{O} = \{O_{i_1},\dots,O_{i_\ell}\}$ will be made in parallel to students $i_1,\dots,i_\ell$. We will say that a matching M *complies* with a set of offers $\mathcal{O}$ if for every $O_i \in \mathcal{O}$, student s_i is matched in M with a topic of O_i. A matching M compliant with a set of offers $\mathcal{O}$ is *optimal* if no other matching M' compliant with $\mathcal{O}$ has greater value. Thus, the optimal compliant matching is what we can best do if all offers are accepted. A set of offers $\mathcal{O}$ is said to be *feasible* if it satisfies 3 requirements: (i) each student receives at most one offer, (ii) at least one matching is compliant with $\mathcal{O}$, and (iii) for any superset $\mathcal{O}'$ of $\mathcal{O}$, no matching compliant with $\mathcal{O}'$ has a greater value than the optimal matching compliant with $\mathcal{O}$. In other words, requirement (iii) implies that there is no additional offer, made to a student who did not receive

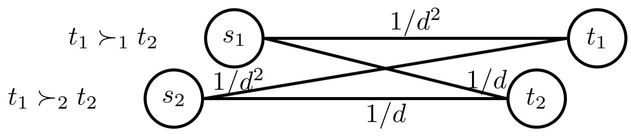

Figure 3: Two available grants, $d < c$ is an arbitrary positive value

any, which can be added to a feasible set in order to improve the value of the optimal matching compliant with it. To illustrate this property, consider an instance with two students, two topics and two grants such that $\mathcal{C}_1 = \{t_1, t_2\}, \mathcal{C}_2 = \{t_2\}$, $t_2 \succ_1 t_1$, $v(s_1, t_1) = v(s_2, t_2) = 1$ and $v(s_1, t_2) = 3$. The set of offers $\mathcal{O} = \{O_1\}$, where $O_1 = \{(s_1, t_2)\}$, is feasible since an additional offer to student s_2 would lead to a compliant matching $\{(s_1, t_1), (s_2, t_2)\}$ which has a strictly lower value than $\{(s_1, t_2)\}$.

In this work, we will only consider feasible sets of offers $\mathcal{O}$ for the first phase. Each time a student accepts an offer, the jury commits to assign him a topic of the offer. If all answers to $\mathcal{O}$ are positive, then the protocol returns an optimal matching compliant with $\mathcal{O}$, and the second phase can be skipped. The goal of the second phase is to complete and improve the optimal matching compliant with the accepted offers.

In Section 4.1, we show that no constant approximation ratio can be achieved by a mixed protocol if either a student applies to less than k topics or the protocol is restricted to fixed offers. In Section 4.2, we provide an upper bound of $1/\sqrt{2}$ for the approximation ratio achievable by a mixed protocol when each student applies to at least k topics, and we provide a mixed protocol with approximation ratio $1/2$.

4.1 Cases of Unbounded Approximation Ratio

The following proposition shows that no mixed protocol achieves a constant approximation ratio.

Proposition. 3 *For any constant $c \leq 1$, no mixed protocol achieves an approximation ratio of c.*

Under the light of this impossibility result, we consider an additional constraint on the instances leading to constant approximation ratios. We assume in the remaining of the section that the number of candidacies for each student is at least k. This assumption is realistic since the jury can invalidate a student's application if it comprises less than k topics. The following proposition shows that even with this restriction, no mixed protocol restricted to fixed offers achieves a constant approximation ratio.

Proposition. 4 *For any constant $c \leq 1$, no mixed protocol restricted to fixed offers achieves an approximation ratio of c, even if $|\mathcal{C}_i| \geq k$ holds for all $s_i \in \mathcal{S}$.*

Proof: Consider the instance described by Figure 3. During phase one, either ($O_1 = \{t_1\}$ and $O_2 = \{t_2\}$) or ($O_1 = \{t_2\}$ and $O_2 = \{t_1\}$) are proposed. In the former case, if $\mathcal{A}_1 = \emptyset$ and $\mathcal{A}_2 = \{t_1, t_2\}$ then s_1 declines and s_2 accepts. In the latter case, if $\mathcal{A}_1 = \{t_1, t_2\}$ and $\mathcal{A}_2 = \emptyset$ then s_1 accepts but s_2 declines. In both cases, topic t_1 remains unassigned whereas it could be assigned to the student who accepted the offer. Therefore, the resulting matching has value $1/d$ whereas M^* has value $1/d^2$. This implies that no mixed protocol restricted to fixed offers achieves a better approximation ratio than d and $d < c$. ■

4.2 Mixed Protocol with Flexible Offers

The following proposition provides an upper bound on the approximation ratio for a mixed protocol with flexible offers.

Proposition. 5 *No mixed protocol achieves an approximation ratio larger than $1/\sqrt{2}$, even if $|\mathcal{C}_i| \geq k$ for any $s_i \in \mathcal{S}$.*

Finally, we present a mixed protocol, called *Max-sum*, which achieves an approximation ratio of $1/2$. During phase one, the protocol starts by computing an optimal matching $M_\mathcal{C}$ in $G_\mathcal{C}$. Let S denote the set of students matched in $M_\mathcal{C}$. The set of offers $\mathcal{O}$ contains offer $O_i := \mathcal{C}_i$ for each student $s_i \in S$. Let $A \subseteq S$ denote the set of students who accept their offer, and let $\mathcal{O}_A$ be the set of offers that they have received. For each student $s_i \in S \setminus A$, candidacy set $\mathcal{C}_i$ is updated by removing the topics contained in his least preferred $\succsim_i$-equivalence class. Let M_A be an optimal matching compliant with $\mathcal{O}_A$. During phase two, Algorithm 1 is used with input $\mathcal{C}$ and M_A.

Let $\mathcal{M}_A$ be the set of all $\mathcal{A}$-matchings such that each student $s_i \in A$ is assigned a topic of $\mathcal{C}_i$. We are going to see that using Algorithm 1 with input M_A leads to an $\mathcal{A}$-matching of maximum value within $\mathcal{M}_A$. This is more general than Proposition 1 (input $M^0 = M_A$ instead of $M^0 = \emptyset$) which is a corollary to the following result.

Proposition. 6 *Algorithm 1 with input $\mathcal{C}$ and M_A returns an $\mathcal{A}$-matching of maximum value within $\mathcal{M}_A$.*

Proof: By definition, $\mathcal{A}$ is a subset of $\mathcal{C}$. During its execution, Algorithm 1 prunes some elements of $\mathcal{C} \setminus \mathcal{A}$. Eventually, the set of candidacies is equal to $\mathcal{C}'$, and $\mathcal{C} \supseteq \mathcal{C}' \supseteq \mathcal{A}$ holds. Let $\mathcal{M}'_A$ be the set of matchings in $G_{\mathcal{C}'}$ such that each student $s_i \in A$ is assigned a topic of $\mathcal{C}_i$. Note that an $\mathcal{A}$-matching which is optimal within $\mathcal{M}'_A$ is also optimal within $\mathcal{M}_A$. Therefore, our proving strategy is to show that Algorithm 1 with input $\mathcal{C}$ and M_A provides a matching which is optimal within $\mathcal{M}'_A$.

Let M^ℓ denote the current solution during the execution of Algorithm 1. More precisely, M^0 is the initial solution. Each time the solution is modified (in line 7), its index is incremented. Thus, M^ℓ contains $|M^0| + \ell$ edges. We show by induction on ℓ that M^ℓ is (ℓ)-optimal i.e., optimal within $\mathcal{M}'_A$ restricted to the matchings of size at most $|M^0| + \ell$.

Base case ($\ell = 0$): By construction, M^0 is (0)-optimal.

Induction step: By induction hypothesis, M^ℓ is (ℓ)-optimal. We show that $M^{\ell+1}$ is $(\ell+1)$-optimal. Let P denote the M^ℓ-optimal augmenting path computed at the end of iteration ℓ and such that $M^{\ell+1} = M^\ell \Delta P$. Note that $v(M^{\ell+1}) = v(M^\ell) + v_{M^\ell}(P) \geq v(M^\ell)$ holds, and therefore $M^{\ell+1}$ is (ℓ)-optimal. By contradiction, assume that there exists a matching $M' \in \mathcal{M}'_A$ of size $|M^0| + \ell + 1$ such that $v(M') > v(M^{\ell+1})$. Without loss of generality, we assume that M' is $(\ell+1)$-optimal and as close as possible to M^ℓ i.e., there is no matching M in $\mathcal{M}'_A$ of size $|M^0| + \ell + 1$ such that $v(M) = v(M')$ and $M^\ell \cap M' \subset M^\ell \cap M$ hold. Let $P_1, \ldots, P_s$ be the alternating paths of $M^\ell \Delta M'$. Note that each of these paths belongs to $G_{\mathcal{C}'}$ since both M^ℓ and M' belong to $G_{\mathcal{C}'}$.

We claim that no alternating path P_i is a cycle or an even path. By contradiction, assume that P_i is either a cycle or an even path. Let M be the matching in $G_{\mathcal{C}'}$ of size $|M^0| + \ell + 1$ such that $M = M' \Delta P_i$. Note that M belongs to $\mathcal{M}'_A$ since both M^ℓ and M' belong to $\mathcal{M}'_A$. If $v_{M^\ell}(P_i) \leq 0$ then $v(M') = v(M) - v_{M'}(P_i) = v(M) + v_{M^\ell}(P_i) \leq v(M)$ implies that M is also $(\ell + 1)$-optimal, a contradiction since M is closer to M^ℓ than M'. On the other hand, if $v_{M^\ell}(P_i) > 0$ then matching $M^\ell \Delta P_i$ of size $|M^\ell|$ has value $v(M^\ell) + v_{M^\ell}(P_i) > v(M^\ell)$, a contradiction with the fact that M^ℓ is (ℓ)-optimal since $M^\ell \Delta P_i$ belongs to $\mathcal{M}'_A$.

Since $|M^\ell| < |M'|$, there must be at least one augmenting path P_j for M^ℓ in $M' \Delta M^\ell$. We claim that $M' \Delta M^\ell$ does not contain any decreasing path for M^ℓ. The existence of a decreasing path P_r for M^ℓ leads to a contradiction. This can be shown by using the same argument as above (replace P_i by $P_r \cup P_j$). Since $M^\ell \Delta M'$ only contains augmenting paths for M^ℓ and $|M'| = |M^\ell| + 1$, $M' \Delta M^\ell$ contains a single path P_1. Therefore, $v(M') = v(M^\ell) + v_{M^\ell}(P_1)$ holds and it implies $v_{M^\ell}(P_1) > v_{M^\ell}(P)$ since both $v(M^{\ell+1}) = v(M^\ell) + v_{M^\ell}(P)$ and $v(M') > v(M^{\ell+1})$ hold, a contradiction with P is M^ℓ-optimal since P_1 belongs to $G_{\mathcal{C}'}$.

Now we know that M^ℓ is (ℓ)-optimal. If the **while** loop of Algorithm 1 stops because $|M^\ell| = k$, then M^ℓ is outputted and must be optimal within $\mathcal{M}'_A$ because k is the maximum cardinality of a matching. If the **while** loop of Algorithm 1 stops because $v_{M^\ell}(P) < 0$ for every augmenting path P for M^ℓ then we shall see that M^ℓ is optimal within $\mathcal{M}_A$. Assume by contradiction that there exists a matching M' within $\mathcal{M}'_A$ of size greater than $|M^\ell|$ such that $v(M') > v(M^\ell)$. With arguments similar to the ones used in the above induction, we can check that $M^\ell \Delta M'$ contains only augmenting paths $P_1, P_2, \ldots, P_{|M'|-|M^\ell|}$ for M^ℓ. Since $v(M') = v(M^\ell) + \sum_{i=1}^{|M'|-|M^\ell|} v_{M^\ell}(P_i)$ and $v(M') > v(M^\ell)$ hold, there must be at least one augmenting path P_i for M^ℓ such that $v_{M^\ell}(P_i) > 0$, a contradiction since P_i belongs to $G_{\mathcal{C}'}$. ■

Proposition 6 shows that the second phase of Max-sum suffers from no loss of value. Bad decisions can be made during the first phase, but their impact is limited.

Proposition. 7 *The approximation ratio of Max-sum is* $1/2$.

5 Related Work

The problem of assigning PhD grants to students is closely related to the student-project allocation problem [Abraham *et al.*, 2007; Abu El-Atta and Moussa, 2009] where projects are proposed by lecturers who have preferences over student/topic pairs. The main novelty of our model lies in the fact that it is decentralized and the preferences of the students (i.e., the set of topics that they may accept) are unknown when the jury builds the matching. The model proposed by Che and Koh [2016] shares also some similarity with our problem. They study a decentralized college admission problem where the students' preferences are unknown and they define the equilibria in the related games.

Matching problems where parameters are unknown or partially observable have already been investigated in the literature. One of the closest model to our problem is the stable matching with partially observable preferences [Drummond and Boutilier, 2013] (see also Rastegari et al. [2013; 2014]). The notion of maximum regret used to quantify the quality of a solution is close to our notion of approximation. Probabilistic models have also been used to represent uncertainty over the preferences of agents for stable matchings [Aziz *et al.*, 2017a] and assignment problems [Aziz *et al.*, 2017b].

Another related matching problem has been proposed by Anshelevich and Sekar [2016] where we know how the edges compare but the exact edges' weights are unknown. Another problem close to ours is the stochastic matching problem [Blum *et al.*, 2015] which aims at finding a maximum matching in a graph where the existence of the edges is only probabilistically known. Similarly to our problem, the matching is constructed through a protocol which checks at each step the existence of a given edge. However, the fact that the protocol checks the existence of a given edge does not imply that one of its extremities should be selected if the edge exists.

Finally, the mixed protocols presented in this paper bear resemblance with the robust matching problem [Hassin and Rubinstein, 2002]. This problem consists in finding a matching whose restriction to its k-best edges provides a good approximation of the optimal matching of size k, for any integer k. The authors show that no algorithm achieves a better approximation than $1/\sqrt{2}$, and they provide an algorithm achieving this ratio. A robust matching is not necessarily a good candidate for the first phase of a mixed protocol because it may contain edges of large value which are unacceptable, and edges of low value which are acceptable. However, the algorithm to produce a robust matching is a good candidate to design an efficient mixed protocol.

6 Conclusion and Future Works

In this paper, we have proposed a decentralized matching model for the problem of assigning grants to PhD candidates where the answers of students are not known in advance by the university. This setting can cover other applications (e.g. allocation of internships in a company). We have provided a theoretical analysis of the best approximation ratio achievable by a protocol of exchanges between the jury and the students.

An open question is whether the $1/2$-approximation can be improved for mixed protocols (in polynomial time or not). It would also be interesting to relax the notion of feasible set of offers imposed during the first phase of a mixed protocol. For example, a feasible set could be instead the largest set of offers that can be made to students without losing the possibility to obtain an $\mathcal{A}$-optimal matching.

Acknowledgements

This work is supported by the bilateral Slovak-French grant of Campus France PHC STEFANIK 2018, 40562NF and Slovak Research and Development Agency APVV SK-FR-2017-0022. Cechlárová is also supported by VEGA grants 1/0311/18 and 1/0056/18. Gourvès and Lesca are supported by the project ANR-14-CE24-0007-01 CoCoRICo-CoDec.

References

[Abraham *et al.*, 2007] D. Abraham, R. Irving, and D. Manlove. Two algorithms for the student-project allocation problem. *J. of Discrete Algo.*, 5:73–90, 2007.

[Abu El-Atta and Moussa, 2009] A. Abu El-Atta and M. Moussa. Student project allocation with preference lists over (student, project) pairs. In *Proc. of ICCEE*, pages 375–379, 2009.

[Anshelevich and Sekar, 2016] E. Anshelevich and S. Sekar. Blind, greedy, and random: Algorithms for matching and clustering using only ordinal information. In *Proc. of AAAI*, pages 390–396, 2016.

[Aziz *et al.*, 2017a] H. Aziz, P. Biró, T. Fleiner, S. Gaspers, R. de Haan, N. Mattei, and B. Rastegari. Stable matching with uncertain pairwise preferences. In *Proc. of AAMAS*, pages 344–352, 2017.

[Aziz *et al.*, 2017b] H. Aziz, R. de Haan, and B. Rastegari. Pareto optimal allocation under uncertain preferences. In *Proc. of IJCAI*, pages 77–83, 2017.

[Biro and Gudmondsson, 2019] P. Biro and J. Gudmondsson. Complexity of finding Pareto-efficient allocations of highest welfare. Working papers, 2019.

[Blum *et al.*, 2015] A. Blum, J. Dickerson, N. Haghtalab, A. Procaccia, T. Sandholm, and A. Sharma. Ignorance is almost bliss: Near-optimal stochastic matching with few queries. In *Proc. of ACM EC*, pages 325–342, 2015.

[Bodin *et al.*, 2017] D. Bodin, J. Schmidt, R. Lemle, B. Roper, R. Goldberg, K. Hill, C. Parrish, S. Williams, A. Kuemmel, and W. Siegel. Recruitment and selection in health service psychology postdoctoral training: A review of the history and current issues. *Training and Education in Professional Psychology*, 12, 11 2017.

[Chade and Smith, 2006] H. Chade and L. Smith. Simultaneous search. *Econometrica*, 74(5):1293–1307, 2006.

[Chade *et al.*, 2014] H. Chade, G. Lewis, and L. Smith. Student Portfolios and the College Admissions Problem. *The Review of Economic Studies*, 81(3):971–1002, 02 2014.

[Che and Koh, 2016] Y. Che and Y. Koh. Decentralized college admissions. *J. of Poly. Eco.*, 124(5):1295–1338, 2016.

[Deng *et al.*, 2017] Y. Deng, D. Panigrahi, and B. Waggoner. The complexity of stable matchings under substitutable preferences. In *Proc. of AAAI*, pages 480–486, 2017.

[Drummond and Boutilier, 2013] J. Drummond and C. Boutilier. Elicitation and approximately stable matching with partial preferences. In *Proc. of IJCAI*, pages 97–105, 2013.

[Ekmekci and Yenmez, 2014] M. Ekmekci and M. Yenmez. Integrating Schools for Centralized Admissions. Working paper, 2014.

[Gale and Shapley, 1962] D. Gale and L. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

[Gondran and Minoux, 2008] M. Gondran and M. Minoux. *Graphs, dioids and semirings: new models and algorithms*. Springer Science & Business Media, 2008.

[Goto *et al.*, 2016] M. Goto, A. Iwasaki, Y. Kawasaki, R. Kurata, Y. Yasuda, and M. Yokoo. Strategyproof matching with regional minimum and maximum quotas. *Artificial Intelligence*, 235:40 – 57, 2016.

[Hamada *et al.*, 2017] N. Hamada, C. Hsu, R. Kurata, T. Suzuki, S. Ueda, and M. Yokoo. Strategy-proof school choice mechanisms with minimum quotas and initial endowments. *Artificial Intelligence*, 249:47 – 71, 2017.

[Hassin and Rubinstein, 2002] R. Hassin and S. Rubinstein. Robust matchings. *SIAM J. on Discrete Mathematics*, 15:530–537, 2002.

[Karp *et al.*, 1990] R. Karp, U. Vazirani, and V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proc. of STOC*, pages 352–358, 1990.

[Kesten, 2012] O. Kesten. On two kinds of manipulation for school choice problems. *Eco. Theory*, 51:677–693, 2012.

[Kouvelis and Yu, 1997] P. Kouvelis and G. Yu. *Robust discrete optimization and its applications*. Kluwer Academic Publishers Group, 1997.

[Manlove, 2013] D. Manlove. *Algorithmics of matching under preferences*. World Scientific, 2013.

[Niederle *et al.*, 2006] M. Niederle, D. Proctor, and A. Roth. What will be needed for the new GI fellowship match to succeed? *Gastroenterology*, 130:218–24, 02 2006.

[Pathak, 2017] P. Pathak. *What Really Matters in Designing School Choice Mechanisms*, page 176–214. Econometric Society Monographs. Cambridge University Press, 2017.

[Rastegari *et al.*, 2013] B. Rastegari, A. Condon, N. Immorlica, and K. Leyton-Brown. Two-sided matching with partial information. In *ACM EC*, pages 733–750, 2013.

[Rastegari *et al.*, 2014] B. Rastegari, A. Condon, N. Immorlica, R. Irving, and K. Leyton-Brown. Reasoning about optimal stable matchings under partial information. In *Proc. of ACM EC*, pages 431–448, 2014.

[Roth and Xing, 1997] A. Roth and X. Xing. Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *J. of Political Economy*, 105(2):284–329, 1997.

[Roth, 1984] A. Roth. The evolution of the labor market for medical interns and residents: A case study in game theory. *J. of Political Economy*, 92(6):991–1016, 1984.

[Roth, 1985] A. Roth. The college admissions problem is not equivalent to the marriage problem. *J. of Economic Theory*, 36(2):277 – 288, 1985.

[Roth, 2008] A. Roth. What have we learned from market design? *The Economic Journal*, 118(527):285–310, 2008.

[Sönmez, 1999] T. Sönmez. Can pre-arranged matches be avoided in two-sided matching markets? *J. of Economic Theory*, 86(1):148 – 156, 1999.

[Vazirani, 2001] Vijay V. Vazirani. *Approximation algorithms*. Springer, 2001.