A Study of Incentive Compatibility and Stability Issues in Fractional Matchings

Extended Abstract

Shivika Narang
Indian Institute of Science
Bengaluru, Karnataka, India
shivika@iisc.ac.in

Yadati Narahari
Indian Institute of Science
Bengaluru, Karnataka, India
narahari@iisc.ac.in

ABSTRACT

The study of stable fractional matchings is fairly recent and moreover, is scarce. This paper reports the first investigation into the important but unexplored topic of incentive compatibility of matching mechanisms to find stable fractional matchings. We focus our attention on matching instances under strict preferences. First, we make the significant observation that there are matching instances for which no mechanism that produces a stable fractional matching is incentive compatible. We then characterize restricted settings of matching instances admitting unique stable fractional matchings. For this class of instances, we prove that every mechanism that produces the unique stable fractional matching is (a) incentive compatible and (b) resistant to coalitional manipulations. We provide a polynomial-time algorithm to compute the stable fractional matching as well. The algorithm uses envy-graphs, hitherto unused in the study of stable matchings.

KEYWORDS

Integral Matchings; Fractional Matchings, Stability; Incentive Compatibility

ACM Reference Format:
Shivika Narang and Yadati Narahari. 2020. A Study of Incentive Compatibility and Stability Issues in Fractional Matchings. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

Matchings have been studied for several decades now, beginning with Gale and Shapley's pioneering work [6]. They introduced the notion of stability and provided algorithms for finding stable matchings. Since then, a considerable amount of work has been carried out on both the theory and applications of stable matchings. Matching mechanisms already in use have also been for their stability and incentive compatibility aspects [3, 5, 7]. The focus of these studies has often been school choice mechanisms or residency matching mechanisms already in practice [1, 2]. In these familiar settings, nodes are wholly or "integrally" matched. We shall call such matchings as integral matchings. A fractional matching is essentially a convex combinations of integral matchings. While they do have a lot of practical relevance, fractional matchings are not applicable to settings such as school choice. Consequently, they have only been studied in literature as a means to produce integral matchings. Even papers that explicitly study fractional allocations only use them towards a deeper understanding of integral matchings [11–13].

Relaxing the integrality constraint in matchings may make the problem harder. For instance, the problem of finding a stable, social welfare maximizing integral matching can be posed as a linear program. However, when we allow for matchings to be fractional, the problem becomes **NP**-Hard. This was shown by Caragiannis et al. [4]. They also show that by allowing the stable matchings to be fractional, we can make large gains in terms of social welfare. Thus, it is of interest to study fractional matchings and design algorithms to find fractional matchings with desirable properties. Another crucial requirement for matching mechanisms is incentive compatibility, i.e. the matching mechanism should induce all participating agents to reveal their true preferences. Previous work [8, 10, 14, 15] explores the incentive compatiblity of stable integral matchings. This paper explores the problem of finding incentive compatible mechanisms that produce stable fractional matchings.

2 SETUP AND RESULTS

2.1 Definitions

We represent a stable matching instance as $I = \langle M, W, U, V \rangle$. Here, $M = \{m_1, \cdots, m_n\}$ is the set of men and $W = \{w_1, \cdots, w_n\}$, is the set of women. The valuations of men and women are captured by matrices $U = [a_{i,j}]_{i,j \in [n]}$ and $V = [b_{i,j}]_{i,j \in [n]}$ respectively. In particular, $a_{i,j}$ is m_i's valuation for being matched integrally to w_j. Analogously, $b_{i,j}$ is w_j's valuation for being matched integrally to m_i. We assume that all entries of U and V are non-negative and that a linear order can be derived from the valuations of one agent. Matching problems are traditionally studied as graph problems. Let us denote the induced bipartite graph for a stable matching instance I as $G = (V, E)$ where $V = M \cup W$ and $(m_i, w_j) \notin E \Leftrightarrow U(i, j) = V(i, j) = 0$. Given $v \in V, e \in E$, we shall use $e \perp v$ to denote that e is incident on v. Fractional matchings can be defined as follows.

Definition 2.1 (Fractional Matching). μ is said to be a fractional matching on $G = (V, E)$ if $\mu : E \to [0, 1]$ such that $\forall v \in V, \sum_{e \perp v} \mu(e) \leq 1$.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

Before defining the stability of fractional matchings, we must define a blocking pair in the context of fractional matchings. We say that (m, w) form a *blocking pair* under matching μ if both get strictly less utility from μ than they get by being matched integrally with each other. The utility of a woman w under fractional matching μ is $\sum_{m \in M} \mu(m, w)V(m, w)$. Thus, it is essentially the weighted sum of the utility from each of the integral matchings in the support of μ. The utility of a man can be analogously defined. Hence, we can now define stability for fractional matchings.

Definition 2.2 (Stable Fractional Matchings). A fractional matching μ is said to be stable is there does not exist a pair of agents $(m, w) \in M \times W$ such that $U(m, w) > U(m, \mu(m))$ and $V(m, w) > V(\mu(w), w)$.

We explore the existence of incentive compatible mechanisms to find stable fractional matchings. This paper aims for what is generally known as Bayesian Incentive Compatibility. That is, we shall say that a mechanism is incentive compatible if **truthful revelation** of preferences by all agents is a Nash Equilibrium for *all input instances*. We show that there does not exist a mechanism to find a stable fractional matching which is incentive compatible for all agents. This clearly negates any possibility of a mechanism where truthful revelation is a dominant strategy even for general settings. We call a man woman pair to be MFP (mutual first preference) if they are each others' *first preferences*. That is, m and w are said to be MFP if $w = \arg\max_{a \in W} U^{(t)}(m, a)$, $m = \arg\max_{a \in M} V^{(t)}(a, w)$. Note that for any stable matching instance I with strict preferences, if there exist a pair of nodes that are MFP, they are matched under every stable matching.

2.2 Main Results

The contribution of this paper is in the important but unexplored topic of incentive compatibility of matching mechanisms to find stable fractional matchings. We focus on matching instances under strict preferences. The complete proofs can be found in the full version of the paper [9].

THEOREM 2.3. *There is no incentive compatible mechanism to find stable fractional matchings which gives incentives for truthful revelation of preferences to all agents on all inputs.*

We establish this by giving a cardinal version of the example [10] and show that for each stable fractional matching, there is an agent who can misreport their values and ensure that any stable matching after the misreporting will give them higher utility. As a result, irrespective of the mechanism used, there is always an agent for whom being truthful is not a best response. In light of this result, we characterize the class of unique stable fractional matchings under strict preferences and show that any stable matching mechanism will be incentive compatible when input instances are restricted to those having a unique stable fractional matching. We use Algorithm 1 to define the class **Conditioned Mutual First Preference (CMFP)**.

Definition 2.4. We say that a stable matching instance I is in class CMFP if and only if Algorithm 1 returns a perfect matching when I is given as input.

Algorithm 1: CMFP_matching

Input: $I = \langle M, W, U, V \rangle$
Output: $\mu, I' = \langle M', W', U', V' \rangle$
$t \leftarrow 0$ and $\mu \leftarrow \varnothing$;
$M^{(0)} \leftarrow M,\ W^{(0)} \leftarrow W,\ U^{(0)} \leftarrow U,\ V^{(0)} \leftarrow V$;
$I^{(0)} = \langle M^{(0)}, W^{(0)}, U^{(0)}, V^{(0)} \rangle$;
while $\exists (m, w) \in M^{(t)} \times W^{(t)}$ *s.t.* (m, w) *is MFP in* $I^{(t)}$ **do**
    $\mu \leftarrow \mu \cup (m, w)$. Set $I^{(t+1)}$ as $I^{(t)}$ without m and w;
    $t \leftarrow t + 1$;
$I' = I^{(t)}$;

LEMMA 2.5. *Given any stable matching instance I, Algorithm 1 returns a matching that is a subset of any stable integral matching on I.*

THEOREM 2.6. *Given any matching instance belonging to CMFP, under any mechanism to find a stable fractional matching, truthful revelation of preferences forms a Nash Equilibrium.*

PROOF. Given a stable matching instance $I \in CMFP$, every mechanism resulting in a stable fractional matching will return the same matching μ^*. We use Algorithm 1 to give us labellings $i_1, \cdots, i_n$ and $j_1, \cdots, j_n$ such that $\mu^* = \{(m_{i_1}, w_{j_1}), \cdots, (m_{i_n}, w_{j_n})\}$ where (m_{i_t}, w_{j_t}) are matched in the t^{th} round in Algorithm 1. Let all other agents be truthful. Clearly, (m_{i_1}, w_{j_1}) have no incentive to misreport their preferences as they are already matched to their first preference. As long as m_{i_1} and w_{j_1} stay truthful, they will continue to be matched to each other, irrespective of how other agents are behaving. Now for $t > 1$ for (m_{i_t}, w_{j_t}) neither can gain by increasing or decreasing their value for any agent matched earlier. This is because the agents who are matched before round t are truthful and will not become MFP pairs with m_{i_t} or w_{j_t}. Thus, those pairings will not change. Of the remaining agents, m_{i_t} and w_{j_t} have highest value for each other and cannot benefit from misreporting their preferences. Consequently, when all other agents are truthful, no agent has an incentive to misreport their preferences. □

THEOREM 2.7. *A stable matching instance I has a unique stable fractional matching if and only if it is in CMFP.*

One side of this claim is clear from Theorem 2.6. In order to establish the other side we give an algorithm to find a stable fractional matching which is not integral, whenever the instance is not in *CMFP*. The complete proof and algorithm can be found in the full version of the paper [9].

3 CONCLUSIONS AND FUTURE WORK

We explore the design of incentive compatible mechanisms for finding stable fractional matchings under strict preferences. While this is not possible under general settings, when the instance has a unique stable fractional matching, no agent has an incentive to misreport. There is much scope for future work. Manipulability of mechanisms for stable fractional matchings must be studied. Further, it may be possible to achieve incentive compatibility by relaxing the stability constraint.

REFERENCES

[1] Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. 2009. Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match. *American Economic Review* 99, 5 (2009), 1954–78.

[2] Atila Abdulkadiroğlu, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez. 2005. The Boston public school match. *American Economic Review* 95, 2 (2005), 368–371.

[3] Surender Baswana, Partha Pratim Chakrabarti, Sharat Chandran, Yashodhan Kanoria, and Utkarsh Patange. 2019. Centralized Admissions for Engineering Colleges in India. *Proceedings of the 2019 ACM Conference on Economics and Computation* (2019), 323–324.

[4] Ioannis Caragiannis, Aris Filos-Ratsikas, Panagiotis Kanellopoulos, and Rohit Vaish. 2019. Stable fractional matchings. *Proceedings of the 2019 ACM Conference on Economics and Computation* (2019), 21–39.

[5] Jose Correa, Rafael Epstein, Juan Escobar, Ignacio Rios, Bastian Bahamondes, Carlos Bonet, Natalie Epstein, Nicolas Aramayo, Martin Castillo, Andres Cristi, and Boris Epstein. 2019. School Choice in Chile. *Proceedings of the 2019 ACM Conference on Economics and Computation* (2019), 325–343.

[6] David Gale and Lloyd S Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69, 1 (1962), 9–15.

[7] Yannai A Gonczarowski, Lior Kovalio, Noam Nisan, and Assaf Romm. 2019. Matching for the Israeli "Mechinot" Gap-Year Programs: Handling Rich Diversity Requirements. *Proceedings of the 2019 ACM Conference on Economics and Computation* (2019).

[8] Fuhito Kojima and Parag A Pathak. 2009. Incentives and stability in large two-sided matching markets. *American Economic Review* 99, 3 (2009), 608–27.

[9] Shivika Narang and Y Narahari. 2020. A Study of Incentive Compatibility and Stability Issues in Fractional Matchings. *arXiv preprint arXiv:2001.05652* (2020).

[10] Alvin E Roth. 1982. The economics of matching: Stability and incentives. *Mathematics of operations research* 7, 4 (1982), 617–628.

[11] Alvin E Roth, Uriel G Rothblum, and John H Vande Vate. 1993. Stable matchings, optimal assignments, and linear programming. *Mathematics of Operations Research* 18, 4 (1993), 803–828.

[12] Jay Sethuraman, Chung-Piaw Teo, and Liwen Qian. 2006. Many-to-one stable matching: geometry and fairness. *Mathematics of Operations Research* 31, 3 (2006), 581–596.

[13] Chung-Piaw Teo and Jay Sethuraman. 1998. The geometry of fractional stable matchings and its applications. *Mathematics of Operations Research* 23, 4 (1998), 874–891.

[14] Chung-Piaw Teo, Jay Sethuraman, and Wee-Peng Tan. 2001. Gale-shapley stable marriage problem revisited: Strategic issues and applications. *Management Science* 47, 9 (2001), 1252–1267.

[15] Rohit Vaish and Dinesh Garg. 2017. Manipulating Gale-Shapley Algorithm: Preserving Stability and Remaining Inconspicuous. *IJCAI* (2017), 437–443.