

Partners in Crime: Manipulating the Deferred Acceptance Algorithm through an Accomplice

Theodora Bendlin, Hadi Hosseini

Department of Computer Science
Rochester Institute of Technology, Rochester, NY, USA
trb7281@rit.edu, hhvcs@rit.edu

Abstract

We introduce a new manipulation strategy available to women in the men-proposing stable matching, called manipulation through an accomplice. In this strategy, a woman can team up with a potential male “accomplice” who manipulates on her behalf to obtain a better match for her. We investigate the stability of the matching obtained after this manipulation, provide an algorithm to compute such strategies, and show its benefit compared to single-woman manipulation strategies.

Introduction

The stable matching problem has been at the forefront of many centralized market problems, such as the residency assignment problem or the college admission problem (Gale and Shapley 1962; Roth and Sotomayor 1989). The Deferred Acceptance (DA) algorithm, proposed by Gale and Shapley, is an elegant procedure that guarantees a stable solution and can be computed in polynomial time (Gale and Shapley 1962). In the men-proposing DA algorithm, men iteratively “propose” to women according to their preferences. Each woman accepts a proposal if she prefers it to her current tentative match. The DA algorithm guarantees a stable outcome, is men-optimal, and strategy-proof for men (Dubins and Freedman 1981). However, the algorithm is also susceptible to manipulation, as its pessimal result for women provides incentives for them to misrepresent their preferences.

The men-proposing DA algorithm is manipulable by a woman either by (1) truncation or (2) permutation of her preference list. Gale and Sotomayor proved that women can induce a woman-optimal matching by “truncating” part of their preference list (Gale and Sotomayor 1985). Teo, Sutherland and Tan developed an $O(n^3)$ algorithm for an optimal single-woman manipulation through the permutation of her preference list (Teo, Sethuraman, and Tan 1999). Extensions of this work yielded a polynomial time algorithm that can produce a manipulation that is nearly indistinguishable from the original preferences, i.e., a list that differs only in the promotion of a single man (Vaish and Garg 2017). These single-woman manipulation strategies are somewhat limited: there may be other, more desirable, matches available to a woman, should she receive help from the proposing side.

Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

We consider a manipulation strategy where a woman can potentially team up with a man to manipulate her outcome, called *manipulation through an accomplice*. We imagine a scenario like the school-choice problem, where a student wishes to use her connections to conspire with a more knowledgeable administrator that is willing to manipulate on her behalf without bearing a noticeable loss for the school. In particular, we investigate whether it is possible for a woman to collaborate with a man to obtain a better match with minimal cost to her accomplice. We aim to answer the following questions: 1) *Is a matching stable with respect to true preferences after manipulating with an accomplice?* and 2) *Can manipulation through an accomplice result in an outcome that is better than optimal single-woman manipulation?*

Preliminaries

A **stable matching problem** consists of two finite and disjoint sets of men (M) and women (W), where $|W| = |M|$, and a preference profile $\succ = (\succ_M, \succ_W)$ where \succ_M and \succ_W are the preference lists of all men and women, respectively. Each agent has an associated preference list \succ_i representing their strict total order preferences over the members of the opposite set. Thus, a man m prefers woman w to w' if $w \succ_m w'$, and a woman w prefers man m to m' if $m \succ_w m'$. A matching μ is a bijective mapping between the two sets such that for all $m \in M$ and $w \in W$, $\mu(w) \in M$ and $\mu(m) \in W$. We use $DA(\succ)$ to denote the matching returned by the men-proposing DA algorithm.

Given a matching μ , a pair of agents (m, w) is called a **blocking pair** if $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$, i.e. both m and w prefer each other to their respective partners in μ . A matching μ is **stable** with respect to the preference profile \succ if it contains no blocking pairs.

A **manipulative pair** (m, w) consists of an accomplice $m \in M$ who willingly permutes his true preference \succ_m to \succ'_m , and a woman w who truthfully reports her preference. Given the men-optimality of DA, an accomplice may be matched with a less preferred partner after the manipulation, which we call **regret**. This may result in an unstable match with respect to the true preference profile. Thus, we introduce a stability property with respect to the manipulative pair’s strategy. Given a manipulative pair (m, w) , a matching μ is **m-stable** with respect to \succ if either: (1) no blocking pairs exist in μ or (2) for all blocking pairs (m', w') , we have

Algorithm 1 Manipulation through an Accomplice

Input: \succ_M, \succ_W , and a woman w **Output:** A preference list $P(m, w, i, m')$, or null $\mu \leftarrow DA(\succ_M, \succ_W)$ Initialize $P'(m, w, i, m')$ to hold optimal preference list**for all** $m \in M$ **do****for all** positions i before $\mu(m)$ **do** $\mu' = DA(\succ'_M, \succ_W)$, with $\succ'_m = P(m, w, i)$ **if** $\mu(m) = \mu'(m)$ and $\mu'(w) \succ_w m'$ **then** $P' \leftarrow P(m, w, i, \mu'(w))$ **return** P'

 $m' = m$, i.e. the man m' is the accomplice.

Manipulation through an Accomplice

We formulate Algorithm 1, which returns the best no-regret permutation strategy for some accomplice who manipulates for w , if such an accomplice exists. We assume there exists a manipulating pair (m, w) , and a structure $P'(m, w, i, m')$ which will hold the optimal permutation strategy for m that returns m' as w 's new match after promoting her to position i in \succ'_m , which we denote as $P(m, w, i)$. By performing this promotion, an accomplice aims to “block” another man from proposing to w to induce a new progression of proposals that may result in a better match for w .

Theorem 1. *Given a manipulative pair (m, w) , manipulation through an accomplice is guaranteed to be m -stable with respect to true preferences.*

Proof. Suppose for a contradiction that there exists a manipulative pair (m, w) whose manipulation μ' is not m -stable. Then, there must be a blocking pair (m', w') that blocks μ' with respect to \succ where $m' \neq m$. For this pair to exist, we have $w' \succ_{m'} \mu'(m')$ and $m' \succ_{w'} \mu'(w')$. We first claim that m' does not submit $\succ'_{m'}$, as our model assumes that m is the only manipulator. By the men-optimality of the DA algorithm, we assume that w' appears before $\mu'(m')$ in $\succ_{m'}$. Therefore, m' must have proposed to w' before $\mu'(m')$, or else $\mu'(m') \succ_{m'} w'$, which contradicts the blocking pair condition. Next, consider w' . If $m' \succ_{w'} \mu'(w')$, then w' must have accepted and maintained m' 's proposal, as women in the DA algorithm only reject proposals when they receive a proposal from a more preferred man. Therefore, w' must be matched with a man $\mu'(w') \succ_{w'} m'$, thus contradicting the blocking pair condition. A contradiction is thus unavoidable, and m' must be m . \square

Manipulation through an accomplice can be better than a single-woman manipulation. Thus far, our goal was to find an accomplice that can help a woman without regret, which is not always possible. We now relax our constraint, and prove that our strategy can result in a better match when compared to a single-woman manipulation.

Theorem 2. *There exists an instance where manipulation through an accomplice produces a more preferred allocation than an optimal single-woman manipulation.*

Proof. Consider the following preferences:

| | | | | | | | | | |
|--------|-------|-------|-------|-------|--------|-------|-------|-------|-------|
| $m_1:$ | w_3 | w_1 | w_2 | w_4 | $w_1:$ | m_4 | m_3 | m_2 | m_1 |
| $m_2:$ | w_3 | w_2 | w_1 | w_4 | $w_2:$ | m_2 | m_1 | m_4 | m_3 |
| $m_3:$ | w_2 | w_4 | w_1 | w_3 | $w_3:$ | m_2 | m_3 | m_4 | m_1 |
| $m_4:$ | w_2 | w_1 | w_4 | w_3 | $w_4:$ | m_4 | m_2 | m_3 | m_1 |

The original match for w_2 , the optimal man after single-woman manipulation, and the optimal man when manipulating with m_1 as the accomplice are shaded with dark gray, medium gray, and light gray, respectively. The woman w_2 prefers m_2 to her optimal match, m_1 , via single-woman manipulation. \square

Concluding Remarks

We devised a new manipulation strategy for women that uses potential male “accomplices”. In light of our findings, there are several plausible directions to be explored: does manipulation through an accomplice always guarantee a more preferred match compared to single-woman manipulation? Can we devise a manipulation algorithm that will return the best accomplice while considering both the woman’s potential gain as well as the accomplice’s potential regret? Several intriguing algorithmic and computational questions lie within these directions.

References

- Dubins, L. E., and Freedman, D. A. 1981. Machiavelli and the gale-shapley algorithm. *The American Mathematical Monthly* 88(7):485–494.
- Gale, D., and Shapley, L. S. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69(1):9–15.
- Gale, D., and Sotomayor, M. 1985. Ms. Machiavelli and the stable matching problem. *The American Mathematical Monthly* 92(4):261–268.
- Roth, A. E., and Sotomayor, M. 1989. The college admissions problem revisited. *Econometrica* 57(3):559–570.
- Teo, C.-P.; Sethuraman, J.; and Tan, W.-P. 1999. Gale-shapley stable marriage problem revisited: Strategic issues and applications. In *Proceedings of the 7th International IPCO Conference on Integer Programming and Combinatorial Optimization*, 429–438. London, UK, UK: Springer-Verlag.
- Vaish, R., and Garg, D. 2017. Manipulating gale-shapley algorithm: Preserving stability and remaining inconspicuous. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 437–443.