Anarchy, Stability, and Utopia: Creating Better Matchings

Elliot Anshelevich, Sanmay Das, and Yonatan Naamad

Dept. of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180
{eanshel,sanmay,naamay2}@cs.rpi.edu

Abstract. We consider the loss in social welfare caused by individual rationality in matching scenarios. We give both theoretical and experimental results comparing stable matchings with socially optimal ones, as well as studying the convergence of various natural algorithms to stable matchings. Our main goal is to design mechanisms that incentivize agents to participate in matchings that are socially desirable. We show that theoretically, the loss in social welfare caused by strategic behavior can be substantial. However, under some natural distributions of utilities, we show experimentally that stable matchings attain close to the optimal social welfare. Furthermore, for certain graph structures, simple greedy algorithms for partner-switching (some without convergence guarantees) converge to stability remarkably quickly in expectation. Even when stable matchings are significantly socially sub-optimal, slight changes in incentives can provide good solutions. We derive conditions for the existence of approximately stable matchings that are also close to socially optimal, which demonstrates that adding small switching costs can make socially optimal matchings stable.

1 Introduction

This paper investigates the social quality of stable matchings. The theory of stable matching has received attention because of its many applications, including matching graduating medical students to residency programs [1], and matching kidney donors with recipients [2]. Most of the work on stable matching has assumed that the agents being matched have some preference ordering on who they would like to be matched with, without assigning a concrete utility for agent $i$ being matched with agent $j$ [3,4,5, inter alia]. This is natural, because stability as a concept does not need the stronger requirement of ascribing utilities to outcomes: it only needs the ranking of matchings from the perspective of every agent.

Matching problems, however, often bring with them outcomes that need to be evaluated in terms of cardinal utility. This occurs, for example, in pair programming, a central practice of the software engineering methodology known as Extreme Programming [6]. The utility of a matching is a function of the productivity of a pair of programmers working together. In kidney exchange, as well as many other stable matching scenarios, the goal is not only to form stable matchings, but also to form a matching with high overall utility.
The properties of matching mechanisms determine the utilities received by agents in these situations. A good mechanism for kidney exchange could make donors happier with their decision to donate while arranging the best possible matches for recipients. A good mechanism for pairing programmers would lead to the best possible programming productivity for their employer. Inevitably, there is a tradeoff between stable matchings, which are pairwise (or groupwise) rational, and socially optimal matchings (for our purposes, for the rest of this paper we assume simple additive social utilities, so that the socially optimal matching is the one that maximizes the sum of utilities received by each individual). The central question of mechanism design for matching markets is how to get people into “good” matchings, however “good” is defined. Almost all the work on matching mechanism design has focused on engineering stable matchings. This work has met with significant large-scale success in applications like matching graduating medical students to residency programs, and matching students to public high schools [7, 11]. Some of this work, especially recent work on designing high school student matches, also explicitly seeks to realize the best matchings for one side of the market (in the high school case, the best matchings for students), but the notion of welfare is weak pareto-optimality among the set of stable matches for one side of the market [8].

Our focus is on extending our understanding of matching problems in situations where we are concerned with social welfare in terms of utility, instead of just stability and choice among stable outcomes. Several alternatives may be available in these situations, ranging from purely centralized allocation based on information available to a matchmaker, to purely individual decision-making based on personal preferences. The first set of questions that arises can be divided into three categories: (1) How bad are stable matchings when compared with socially optimal ones? (2) Can agents find stable matchings on their own? What are the outcomes of algorithms they may actually use in practice? (3) How can we incentivize agents to participate in matchings that are socially desirable?

**Our Results.** We initiate an investigation of the questions described above in the context of two-sided matchings, and give both theoretical and experimental results. Specifically, we study the effects of different network structures and utility distributions on the price of anarchy: the ratio of social utilities achieved by stable and optimal matchings respectively. We find that in most cases the stable matching attains close to the optimal social welfare (generally above 90%). We characterize some situations where the price of anarchy can be more substantial, and then study a potential means of incentivizing good stable matchings in Section 5. We consider *approximate stability*, which corresponds to the addition of a switching cost to the mechanism, so that an agent would have to pay in order to deviate from the current matching. We show both theoretically and experimentally that the addition of a small switching cost can greatly improve the quality of stable solutions. Finally, in Section 6 we consider several greedy algorithms for partner-switching, and show experimentally that they converge quickly to stability for some simple yet natural distributions of utilities, as well as prove convergence guarantees.