The High Cost of Stability in Two-Sided Matching: How Much Social Welfare Should be Sacrificed in the Pursuit of Stability?

Robert L. Axtell and Steven O. Kimbrough
George Mason University and the University of Pennsylvaina
rax222@gmu.edu, sok@wharton.upenn.edu

Abstract. New results on a class of distributed two-sided matching algorithms are described. In particular, these algorithms can dominate the matches produced by the Gale-Shapley algorithm in terms of both average social welfare and fairness. However, stability is not guaranteed and therefore a natural trade-off arises between stability and social welfare. At one extreme, stability is 'purchased' in Gale-Shapley by significant welfare losses. At the other extreme, matches yielding high social welfare are vulnerable to unraveling when information costs—basically, the cost of acquiring others' ranks—are small. In the face of either large information costs or a 'veil of ignorance' for match participants, it may be easy to justify matches that are not stable. Computational evidence is provided that these extreme Gale-Shapley matchings are brittle in the sense of being not robust to ostensibly minor perturbations. We argue that the nearly universal focus on stable matchings in this literature is misguided at best. As a positive description of any human social process, such as marriage matching or college admissions, Gale-Shapley and related centralized, synchronized algorithms are hopelessly flawed. As normative devices for achieving social good they explore only a tiny part of the Pareto frontier that trades-off stability for welfare, and so do not serve as robust design tools.

Keywords: two-sided matching, Gale-Shapley algorithm, deferred acceptance algorithm, distributed matching.

1 Matching, in Practice and Theory

There are many examples of social phenomena in which distinct sets of heterogeneous actors (people, institutions) get matched in order to subsequently perform some function. Consider firms who are hiring certain kinds of workers in a specific geographical region. There is some set of processes by which job openings are filled by available workers. For purposes of the present paper we will consider such processes to *match* workers and firms. Or consider high school seniors applying for admission to colleges and receiving offers back. The ways in which students are *matched* to colleges is another example of the two-sided matching problem. Perhaps the most well known example of such matching problems is the so-called *marriage* problem, that is, how people with heterogeneous preferences for one another end up in matrimonial pairs. The goal of this paper is to study the properties of such matching processes.

1.1 Positive Matching Models to Describe the Real World

Given the variety of matching processes in human social systems, it is surprising that there has been little formal modeling of extant processes. That is, while there are lots of verbal descriptions of how students pick colleges and how colleges pick students, we are unaware of any general model of the process.

Similarly with marriage. While there are many, many sociological and ethnographic studies of the institution and its operation, there are no credible models of how people actually form couples.

Surely one reason—probably the most important reason—for this dearth of positive work is the peculiar fixation of most previous work on a particular normative notion with peculiar properties, which we describe in some detail next.

1.2 Normative Matching and the Gale-Shapley Algorithm

Early on in the history of the two-sided matching problem a particular matching algorithm was advanced with several ostensibly desirable properties (*1*). This is the 'deferred acceptance' algorithm and we describe it here, after defining the problem more formally.

Consider two sets of agents, not necessarily individuals, denoted by A and B. The cardinality of these sets may or may not be equal. Let each $a \in A$ have a strict preference order over agents in B, and vice versa for all $b \in B$. A *match*, $\mu(A, B)$ is simply a list of pairs, (a, b), for all $a \in A \cup \varnothing$ and $b \in B \cup \varnothing$, where each agent is matched at most q times, where q is its *quota*. In the case of marriage, $q = 1$, while for college admissions, q equals the number of freshman slots available at the college in question. If an agent is matched with the empty set then it is unmatched.

With this specification of the problem now in-hand, we describe the 'deferred acceptance' algorithm as follows:

1. Each $a \in A$ not currently matched proposes a match to that $b \in B$ that it most prefers and has not already turned it down.
2. Each $b \in B$ who has received proposals turns down all but its highest ranked proposer.
3. This process repeats until no new proposals occur.

Gale and Shapley proved that this simple process always terminates. Additionally, the mechanism has other interesting properties.

2 Properties of Matches in General and Gale-Shapley Specifically

A match is called *acceptable* if it does not contain any pairs, (a, b), such that either a prefers being unmatched to being matched with b, or b prefers being unmatched to being matched with a. A match is *individually rational* if it is acceptable and does not exceed any quotas. A match is *Pareto optimal* if there is no alternative match, μ', such that at least one agent has a pairing in μ' that it strictly prefers to μ and no other agents strictly prefer μ to μ'. A match is called *stable* if there do not exist pairs (a, b) and (c, d) such that a prefers d to b and d prefers a to c. Stability is a stronger notion than Pareto optimality, as we can see by considering the marriage problem.

Imagine that there is a marriage match that is Pareto optimal. This means that it is not possible to rearrange any two pairs to make all agents better off. However, it is possible for a match to be Pareto optimal but not stable. Consider the two matched pairs (a, b) and (c, d), but that a prefers d to b and d prefers a to c, but c does not prefer b to d and b does not prefer c to a. This leaves us in the situation that a and d could decide to form the pair (a, d), leaving the unhappy couple (c, b). This rearrangement is not Pareto improving on the original one, since both b and c are less well off. In this sense there are clearly Pareto optimal matches that are not stable while all stable ones are Pareto optimal.

2.3 Stability and Equilibrium

It is a somewhat surprising result that there exist stable matchings for any two-sided matching problem. It seems reasonable to consider a stable matching as an equilibrium of two-sided matching insofar as once such a match is established no agent has any reason to unilaterally deviate. Furthermore, Gale and Shapley proved that the deferred acceptance procedure always generates such a stable matching (their theorem 1). What is more, they also showed (theorem 2) that the matching produced by deferred acceptance is *optimal* with respect to propers' welfare, insofar as there is no other stable matching that yields more welfare to proposers.

2.4 Equity and Fairness

The Gale-Shapley mechanism has an implicit welfare function (SWF) specification. As Knuth (2) first showed, this is a curious SWF indeed. Here we investigate this for populations of agents having random preferences. First, consider the distribution of partner ranks that agents receive in the course of a single Gale-Shapley match, where agents desire their highest choices (low rank, i.e., first, second and so on). These are quite different for men and women in Gale-Shapley. For a population of two thousand men and an equal number of women, we first show the distribution of partner ranks received by men as a probability density function (PDF).

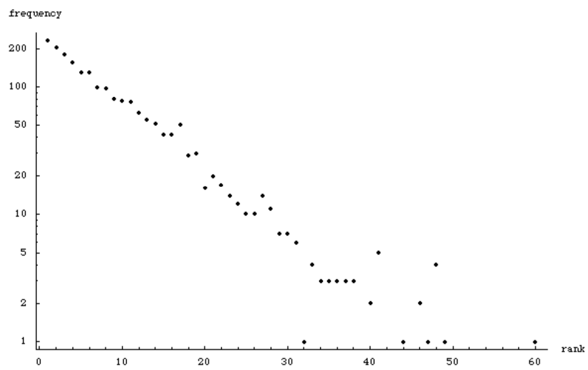

Fig. 1. Distribution of ranks of male matches

Note the logarithmic coordinates on the vertical axis. Since the data are close to a straight line in these coordinates, ranks are exponentially-distributed. The modal value of these data is 1, that is, men getting their highest choice. The median value is between 7 and 8, while the mean is just short of 9.

For women the situation is qualitatively similar but quantitatively very different. In this case the PDF is very erratic, so it is more revealing to look at a counter-cumulative distribution function (CCDF), which is much smoother. This is shown in figure 2, where the mode is once again rank 1, but now the median is about 155 while the mean is nearly 224!

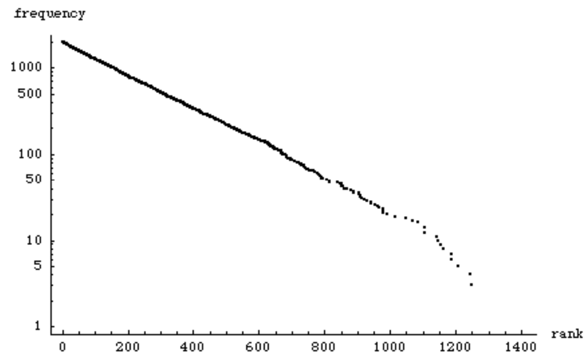

Fig. 2. Counter-CDF of ranks for female matche

As with the men, the women's rank distribution is quite close to being exponential, as evidenced by the nearly linear plot in semilog coordinates.

Next we turn our attention to the average welfare obtained by the agents under Gale-Shapley matching, since this is a well-defined quantity given that the moments are meaningful for an exponential distribution of ranks. We do this as a function of the population size as shown in figure 3.

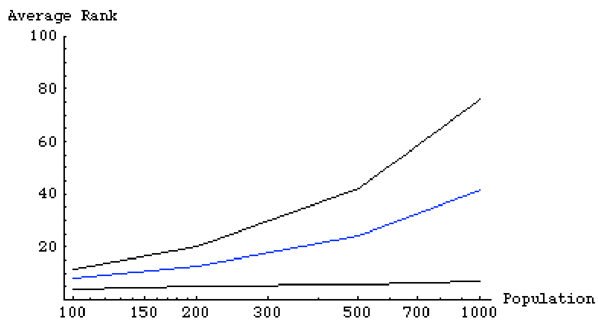

Fig. 3. Average partner rank as a function of the size of the population

The lower line refers to men, while the upper line is for women. Clearly the men do much better for any population size. The blue line is the average across men and women. Note that it is also the case that as the population grows, women do progressively worse than men, i.e., average rank obtained by women/average rank received by men increases.

3 Distributed Matching: Description[1]

In comparison to Gale-Shapley, consider a decentralized matching process in which all agent types can make offers and enter into engagements, subsequently either breaking an engagement off in favor of a different mate or turning the engagement into an unbreakable match. At each step of the model one or more agents are activated to search out a better mate than he or she already has, exhaustively inspecting its preferred mates for someone willing to accept him or her. A single period is defined as a number of agents being active that is equal to the entire population. Clearly this is a more distributed, decentralized process than Gale-Shapley. There are two basic parameters of this 'algorithm': (1) how many periods the agents are active, and (2) the probability that an engagement turns into a match.

4 Distributed Matching: Properties

First we investigate the average welfare properties associated with this matching process. There is no substantive difference between men and women with this distributed matching rule, so we expect only statistical fluctuations to differentiate the two sub-populations. Figure 4 compares the average partner rank received by men and women via this decentralized matching rule (black lines) with that produced by Gale-Shapley (blue line).

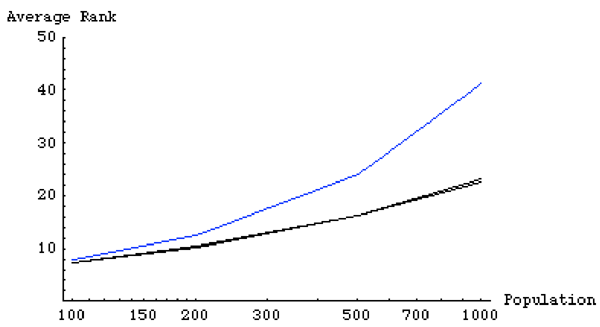

Fig. 4. Average partner rank as a function of the size of the population, for both Gale-Shapley and the decentralized matching process

---

[1] Distributed matching schemes different from those presented here are described by (3).

Clearly, the decentralized matching rules outperform Gale-Shapley on average, with the difference in performance growing with population size. Furthermore, the coincidence of the two black lines means that there is no inequality produced by the decentralized match across genders, in contrast with Gale-Shapley.

Figure 4 was obtained for 100 periods of interaction with the chance that any single engagement turns into an unbreakable match of 1%. We now investigate the extent to which these parameters alter the average partner rank (aka agent welfare) generated by this process. In figure 5 the duration of interaction is varied, from a single period to 1000 periods, and the average partner rank plotted.

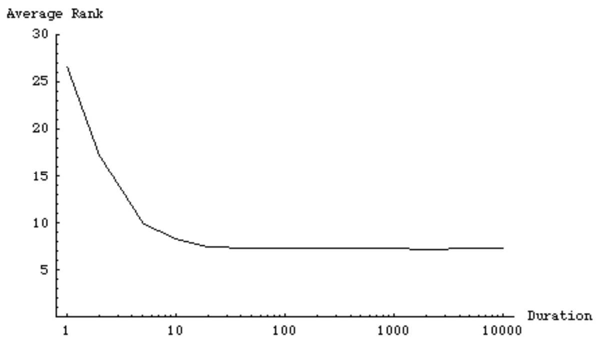

Fig. 5. Effect of the number of periods of interaction on average partner rank for decentralized matching

This shows that few rounds of interaction leave much room for improve-ment, while beyond 20 or rounds little additional welfare is obtained.

The effect of the probability an engagement is turned into a match is shown in figure 6.

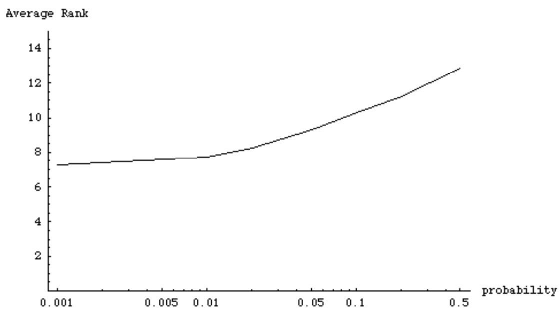

Fig. 6. Dependence of average welfare on the probability an engagement becomes permanent

Note the logarithmic coordinates for the probability. From this figure it is clear that as long as the engagement process does not become frozen into unbreakable matches too quickly the decentralized matching process can yield good matches overall.

Figures 1 and 2 above depict exponential distributions of ranks for Gale-Shapley matching of two thousand men and an equal number of women. Here we investigate the distribution of partner ranks produced by the distributed matching process. The PDF is displayed in figure 7.

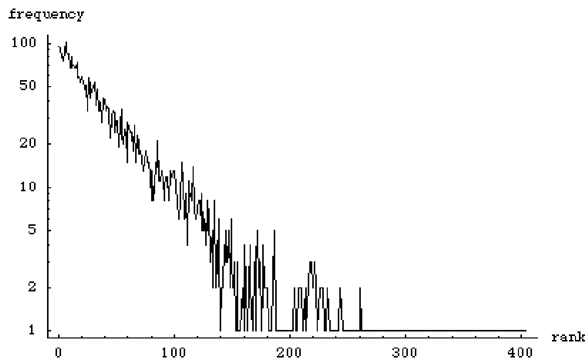

Fig. 7. Exponential distribution of partner ranks under decentralized matching

In these semilog coordinates the presence of a nearly straight line means that once again we have an exponential distribution. The mean rank is much better with this matching process than Gale-Shapley, of course.

The matches produced by the decentralized process are not necessarily stable, and next we investigate how the number of unstable matches grows with the size of the population. Figure 8 shows the average number of unstable matches as a function of the number of agents, for 100 periods of interaction with the chance that any single engagement turns into an unbreakable match of 1%.

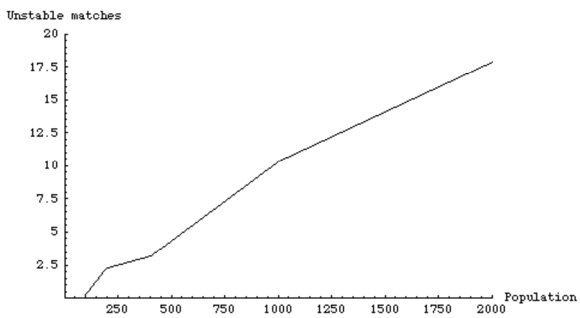

Fig. 8. Slow increase of the number of unstable matches with population size

Note that this is approximately linear. The slope is near 0.01, meaning that about 1% of the matches are unstable with this decentralized matching process. Intuitively, because the agents who make a pair of agent pairs unstable are a small fraction of a large population, it will be difficult for such agents to ever meet up and realize they can do better when part of a large population. For example, the male agent in pairing #253 among the 1000 pairs in a population of 2000 agents may have no idea that he can do 2 ranks better by pairing up with the female agent from pair #791, who would consent because this would provide her an improvement from rank 37 to 33 in her partner. (In any realistic model of (local) information flow, the agents do not know the preferences of arbitrary other agents, only those with whom they interact.) Therefore, the existence of unstable pairings may be of little practical interest, and possibly eliminated only at significant cost to average welfare. It remains to be seen whether modifications to the simple decentralized matching process can yield fewer unstable matches.

5 Summary and Conclusions

The two-sided matching problem has been described and the well-known 'deferred acceptance' algorithm critiqued. We have proposed an alternative, distributed matching process which has some properties that are superior to 'deferred acceptance' and some that are inferior. This naturally introduces a trade-off and the 'best' matching procedures for particular problems will depend on the mixture of properties that one seeks.

Overall, the multi-agent systems perspective on matching adopted here provides us new insights into a well-established problem.

References

1. Gale, D. & Shapley, L. S. (1962) *The American Mathematical Monthly* **69,** 9-15.
2. Knuth, D. E. (1976) *Marriages Stables* (Les Presses de l'Universite de Montreal, Montreal, Canada).
3. Fuku, T., Namatama, A. & Kaizouji, T. (2006) in *Artificial Economics: Agent-Based Methods in Finance, Game Theory and Their Applications*, eds. Mathieu, P., Beaufils, B. & Brandouy, O. (Springer-Verlag, Berlin), Vol. 564.