Matching Mechanisms: Top Trading Cycles

1 Overview

So far, we have focused most of our mechanism design discussion on the case where a central agency or auctioneer starts with a collection of goods and wants to allocate them to n players. Today we will consider the case where the players start with items and we want a mechanism for trade. We will be particularly looking at cases where there is no money involved, and as you will see, the setup will look a bit like the setup for social choice except that different voters will end up with different candidates in a sense. Much of this discussion is taken from Aaron Roth's class notes at UPenn.

2 The House Allocation Problem

Consider the following problem, classically called the "house allocation problem". We have n players. Each player i begins with a single item h_i (their house) but they might like someone else's h_j more than their own. In particular, we'll assume that each player has a strict preference ordering over all the items $h_1, ..., h_n$. Our goal will be to find a mechanism for trading that (a) is Individually Rational (IR): no player is ever harmed by participating if they participate truthfully, (b) is Incentive Compatible (IC): no player can ever benefit by misreporting their true preferences, and (c) finds a Pareto-Optimal solution.

Definition 1 (Pareto Optimality) *An allocation A is Pareto Optimal if any allocation A' that makes some player better off than A also makes at least one player worse off than A.*

For example, if player i prefers h_j to h_i and player j prefers h_i to h_j, then the initial allocation is *not* Pareto Optimal since the two could trade houses and both be better off without hurting anyone else.

Applications: Instead of houses, it's natural to think of office space. Let's say that each student i has some seat in a shared office or perhaps they have a cubicle, and we want some way that students can trade if we're not in a Pareto-Optimal allocation already. Or, a more serious example is kidney exchange. When someone has kidney failure and needs a new kidney, often they will have a partner (or friend or relative) willing to donate to them. If they are blood-type and tissue-type compatible, then the donation will happen. But if they are not, they can participate in a kidney exchamge, where each "player" i is a patient and their "house" h_i is their donor. The rank ordering can depend on compatibility and other factors. the goal then is to find swaps that will improve overall social welfare. We'll stick with the "house" or "office space" terminology.

A simple special case: Suppose everyone has the *same* preference ordering over the h's. That is, there is a common understanding of the best office space followed by the second-best office space, etc. Then what can we say about IR mechanisms? They have to return the original allocation, which indeed is Pareto Optimal.

3 The Top Trading Cycles Algorithm

We now give an algorithm for satisying (a), (b), and (c): individual rationality, incentive compatibility, and that finds a Pareto Optimal allocation. This is called the *Top Trading Cycles* algorithm.

First, given a subset S of the n players, define a "favorites graph" G_S as follows: there is one vertex for each player $i \in S$, and a directed edge (i, j) if h_j is i's favorite house from $\{h_k : k \in S\}$. So, this is a directed graph with out-degree 1 and it *can* have self-loops. Notice that one property of such a graph is that it *must have a cycle* (even if only a self-loop).

Given this definition, the algorithm is quite simple.

Top Trading Cycles:

1. Initialize $S_1 = \{1, ..., n\}$.
2. For $t = 1, 2, ...$ do:
   (a) Create the graph G_{S_t} and find all cycles $C_t^1, C_t^2, ...$ in it. Notice that these cycles must be disjoint since the out-degree of the graph is 1, and as noted above there must be at least one cycle.
   (b) "Clear" each cycle C_t^k by making all trades in it. That is, if $(i, j) \in C_t^k$ then give h_j to player i. Notice that all players on the cycle are getting their top choice from among the houses still available.
   (c) Let $S_{t+1} = S_t \setminus (C_t^1 \cup C_t^2 \cup ...)$, i.e., remove all players matched in this round. If S_{t+1} is empty, then halt.

The first thing to notice is that the algorithm will indeed halt after at most n rounds, since each round will find at least one cycle. So, we need to show the three properties. We'll start with Individual Rationality, then Pareto Optimality, and finally Incentive Compatibility.

Theorem 1 *Top Trading Cycles satisfies Individual Rationality.*

Proof: Can anyone see why? If player i provides their true ordering to the mechanism, then their out-edge in the favorites graph will always be to a house they like at least as much as h_i. So, they will never be matched to a house that is worse for them than h_i. ∎

Theorem 2 *Top Trading Cycles finds a Pareto Optimal allocation.*

Proof: Let TTC denote the allocation produced by the Top Trading Cycles algorithm, and let A denote some other allocation in which no player does worse than they did under TTC. We want to show that A must in fact be TTC itself. To show this, consider the players matched in round $t = 1$. TTC gives each of them their top choice, so A must do so as well. Now, consider the players matched in round $t = 2$. TTC gives them each their top choice from among houses not taken in round 1, so A must also do so as well. In particular, A cannot give them the houses matched in round 1 because we already showed that A had to give those houses to the players matched in round 1. More generally, let's inductively assume that A agrees with TTC for all players (and houses) matching in rounds $1, ..., t$. Then A must also agree with TTC for all players (and houses) matched in round $t+1$ for the same reason: TTC gives these players their top choice from among the houses remaining, and A can't give them any of the houses allocated by TTC in previous rounds, so A must agree with TTC in this round as well. ■

Finally, we need to argue incentive-compatibility. To make this easier, we've defined the algorithm as clearing *all* cycles in the graph at each round. Sometimes the algorithm is described as just picking some arbitrary cycle, which means that a misreporting player could potentially change which arbitrary cycle was picked even if the cycle has nothing to do with them, which makes the argument more complicated. The algorithm as stated is cleaner to analyze because any misreporting that doesn't create a cycle or break a cycle on a given round will have no effect on the algorithm's behavior on that round.

Theorem 3 *Top Trading Cycles is incentive-compatible.*

Proof: Fix some player i. We want to argue that no matter what any of the other players do, player i is best off reporting their true preference ordering. To do this, we will show that it can never benefit player i to misreport their true favorite in any favorites graph, even if we allow them to change their mind from graph to graph.

First, as noted above, misreporting that doesn't create a cycle or break a cycle on a given round will have no effect on the algorithm's behavior on that round. Moreover, misreporting that *breaks* a cycle is never a good idea, because if player i was in a cycle, then they would have received their favorite house in the current graph. So, the only kind of misreporting we need to conisder is misreporting that creates a cycle.

Suppose player i were to misreport and create a cycle by pointing to their kth-favorite house in the current graph for some $k > 1$. The algorithm would then give them that house. We need to argue they will do at least as well by reporting truthfully.

Here's why. Consider the path P in the current graph going from player i's kth-favorite house to player i. Notice that until player i is matched, this path never goes away. That's because the only time that a player changes their out-edge is when the endpoint of their out-edge has disappeared. This means that player i can defer pointing to that house until there is no better house to point to (which is what the algorithm does on their behalf when they report truthfully). ■