DS 574 Algorithmic Mechanism Design **Lecture #11**
Fall 2025 Prof. Kira Goldner

Kidney Exchange

- There are more than 92,000 on the waitlist for a kidney transplant in the US; this makes up 87% of the organ transplant list [1].
- Healthy people have two kidneys and can survive fine with only one.
- A donor and recipient must be "compatible" (blood and tissue types).
- Two incompatible patient-donor pairs can agree to a kidney exchange. This is legal. (Compensation for kidneys is not, except in Iran.)

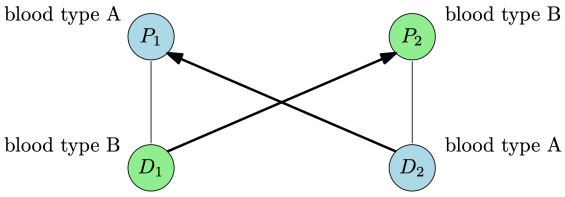

Figure 1: A kidney exchange.

Question: How would one design a centralized mechanism for kidney exchange, where incompatible patient-donor pairs can register and be matched with others?

Idea #1: Use the Top Trading Cycle Algorithm

Vanilla Top Trading Cycles

Consider the housing allocation problem defined by Shapley and Scarf [6]: There are n agents, and each initially owns one house. Each agent has a total ordering over the n houses, and need not prefer their own over the others. How can we reallocate the houses to make the agents better off?

The Top Trading Cycle Algorithm [Gale [6]].

While agents remain:

- Each remaining agent points to its favorite remaining house. This induces a directed graph G on the remaining agents in which every vertex has out-degree 1 (Figure 1).
- The graph G has at least one directed cycle. Self-loops count as directed cycles.
- Reallocate as suggested by the directed cycles, with each agent on a directed cycle C giving its house to the agent that points to it, that is, to its predecessor on C.
- Delete the agents and the houses that were reallocated in the previous step.

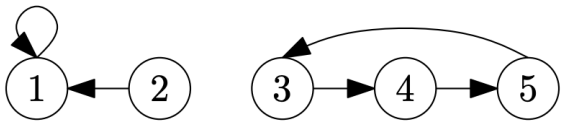

Figure 2: An iteration of the Top Trading Cycle Algorithm (TTCA) with two directed cycles.

Observations:

- This terminates with each agent possessing exactly one house.
- Every agent is only made better off by the algorithm.
- There is no incentive for agents to misreport their preferences. (Requires proof!)

Theorem 1. *The TTCA induces a DSIC mechanism.*

[*Hint:* Divide agents into those allocated to in the jth iteration.]

They key claim is as follows: Let N_j denote the agents allocated in the jth iteration of the TTCA when all agents report truthfully. Agents in N_j are never pointed to by agents of $N_1 \cup \cdots \cup N_{j-1}$ before the jth iteration, and no misreport by N_j can cause this.

Proof. Let N_j denote the agents allocated in the jth iteration of the TTCA when all agents report truthfully. Each agent of N_1 gets its first choice and hence has no incentive to misreport. An agent i of N_2 is not pointed to by any agent of N_1 in the first iteration—otherwise, i would belong to N_1 rather than N_2. Thus, no misreport by i nets a house originally owned by an agent in N_1. Since i gets its first choice outside of the houses owned by N_1, it has no incentive to misreport. In general, an agent i of N_j is never pointed to in the first $j-1$ iterations of the TTCA by any agents in $N_1 \cup \cdots \cup N_{j-1}$. Thus, whatever it reports, i will not receive a house owned by an agent in $N_1 \cup \cdots \cup N_{j-1}$. Since the TTCA gives i its favorite house outside this set, it has no incentive to misreport. □

Now we notice a nice property of the TTCA even stronger than our typical "best response" dynamics.

Definition 1. A *core allocation* is an allocation such that no coalition of agents can make all of its members better off via internal reallocations.

Theorem 2. *For every house allocation problem, the allocation computed by the TTCA is the unique core allocation.*

Proof. To prove the computed allocation is a core allocation, consider an arbitrary subset S of agents. Define N_j as in the proof of Theorem 3.1. Let ℓ be the first iteration in which $N_\ell \cap S \neq \emptyset$, with agent $i \in S$ receiving its house in the ℓth iteration of TTCA. TTCA gives agent i its favorite house outside of those owned by $N_1, \ldots, N_{\ell-1}$. Since no agents of S belong to $N_1, \ldots, N_{\ell-1}$, no reallocation of houses among agents of S can make i strictly better off.

We now prove uniqueness. In the TTCA allocation, all agents of N1 receive their first choice. This must equally be true in any core allocation—in an allocation without this property, the agents of N_1 that didn't get their first choice form a coalition for which internal reallocation can make everyone strictly better off. Similarly, in the TTCA allocation, all agents of N_2 receive their first choice outside of N_1. Given that every core allocation agrees with the TTCA allocation for the agents of N_1, such allocations must also agree for the agents of N_2—otherwise, the agents of N_2 that fail to get their first choice outside N_1 can all improve via an internal reallocation. Continuing inductively, we find that the TTCA allocation is the unique core allocation. □

Modifications for Kidney Exchange

The first attempt was via the TTCA by Roth, Sönmez, and Ünver [3] before the authors talked extensively to doctors. People's "preferences" over kidneys would just be via decreasing probability of success of the transplant.

But kidney exchange is more complicated:

(1) There are patients without living donors, and deceased donors.

(2) The cycles along which reallocations are made can be arbitrarily long.

(3) Modeling preferences as a total ordering over the set of living donors is overkill: empirically, patients don't really care which kidney they get as long as it is compatible with them.

Instead: Binary preferences.

Idea #2: Use a Matching Algorithm

(2) Short reallocation cycles and (3) binary preferences motivate looking for *matchings*, as done in [4].

What's the relevant graph for kidney exchange? Describe the vertices, edges, and what a matching would look like.

The vertex set V corresponds to incompatible patient-donor pairs (one vertex per pair), and we have an undirected edge between compatible exchanges: vertices (P1, D1) and (P2, D2) such that P1 and D2 are compatible and P2 and D1 are compatible. We define the optimal solutions to be the matchings of this graph that have maximum cardinality— that is, we want to arrange as many compatible kidney transplants as possible. By restricting the feasible set to matchings of this graph, we are restricting to pairwise kidney exchanges, and hence "only" 4 simultaneous surgeries.[1]

How do incentives work here? What should the mechanism look like?

Our model for agents is that each vertex i has a true set E_i of incident edges, and can report any subset $F_i \subseteq E_i$ to a mechanism. In practice, proposed kidney exchanges can be refused by a patient for any reason, so one way to implement a misreport is to refuse exchanges in $E_i \setminus F_i$. All that a patient cares about is being matched to a compatible donor. Our mechanism design goal is to compute an optimal solution (i.e., a maximum-cardinality matching) and to be DSIC, meaning that for every agent, reporting its full edge set is a dominant strategy.

Our mechanism takes the following form.

(1) Collect a reported set of incident edges F_i from each agent i.

(2) Form the edge set $E = \{(i,j) : (i,j) \in F_i \cap F_j\}$. That is, include edge (i,j) if and only if both endpoints agree to the exchange.

(3) Return a maximum-cardinality matching of the graph $G = (V, E)$, where V is the (known) set of patient-donor pairs.

But how do we tie-break between maximum-cardinality matchings?

- Different edges for the same vertices: doesn't matter.

[1] These days, 3-way exchanges, corresponding to a directed cycle of 3 patient-donor pairs (with D2 compatible with P1, D3 with P2, and D1 with P3), are increasingly common. The reason is that 3-way exchanges are still logistically feasible and allowing them significantly increases the number of patients that can be saved. Empirically, exchanges involving 4 or more pairs don't really help match more patients, so they are not typically done.

- Different vertices?

Solution: Priority list over patients. In practice, patients are ordered according to some priority, so we can assume that the vertices $1, 2, \ldots, n$ are ordered from highest to lowest priority.[2] Then, we implement step (3) as follows:

(3a) Let M_0 denote the set of maximum matchings of G.

(3b) For $i = 1, 2, \ldots, n$:

(3b.i) Let Z_i denote the matchings in M_{i-1} that match vertex i.

(3b.ii) If $Z_i \neq \emptyset$, set $M_i = Z_i$.

(3b.iii) Otherwise, set $M_i = M_{i-1}$.

(3c) Return an arbitrary matching of M_n.

That is, in each iteration i, we ask if there is a maximum matching that respects previous commitments and also matches vertex i. If so, then we additionally commit to matching i in the final matching. If previous commitments preclude matching i in a maximum-cardinality matching, then we skip i and move on to the next vertex. By induction on i, M_i is a nonempty subset of the maximum matchings of G. Every matching of M_n matches the same set of vertices—the vertices i for which Z_i was non-empty—so the choice of matching in step (3c) is irrelevant.

Theorem 3. *For every collection $\{E_i\}_{i=1}^n$ of edge sets and every ordering of the vertices, the priority matching mechanism above is DSIC: no agent can go from unmatched to matched by reporting a strict subset F_i of E_i rather than E_i.*

Hospital Incentives

Current research is focused on incentive problems at the *hospital* level, rather than at the level of individual patient-donor pairs. Hospitals are the ones who actually report the pairs to the national kidney exchange, but the objectives of a hospital (to match as many of its patients as possible) and of society (to match as many patients overall as possible) are not perfectly aligned.

The Need for Full Reporting. Only reporting pairs who the hospital can't match internally can result in fewer exchanges.

[2]The priority of a patient on a waiting list is determined by numerous factors, such as the length of time it has been waiting on the list, the difficulty of finding a compatible kidney, etc.

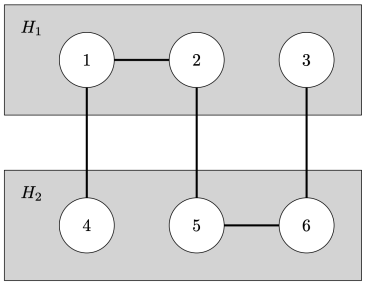

Figure 3: Full reporting by hospitals leads to more matches than with only internal matches.

Hiding patients. If H_1 hides patients 2 and 3 from the exchange (while H_2 reports truthfully), then H_1 guarantees that all of its patients are matched. The unique maximum matching in the report graph matches patient 6 with 7 (and 4 with 5), and H_1 can match 2 and 3 internally. On the other hand, if H_2 hides patients 5 and 6 while H_1 reports truthfully, then all of H_2's patients are matched. In this case, the unique maximum matching in the graph of report matches patient 1 with 2 and 4 with 3, while H_2 can match patients 5 and 6 internally.

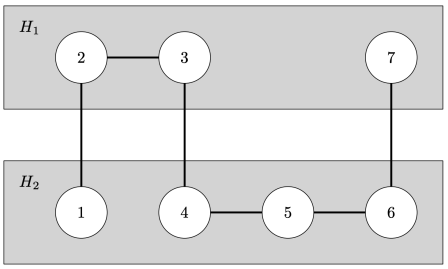

Figure 4: Hospitals can have an incentive to hide patient-donor pairs.

It turns out there cannot be a DSIC mechanism that always computes a maximum-cardinality matching in the full graph.

In light of this example, the revised goal should be to compute an approximately maximum-cardinality matching so that, for each participating hospital, the number of its patients that get matched is approximately as large as in any matching, maximum-cardinality or otherwise. Understanding the extent to which this is possible, in both theory and practice, is an active research topic [2, 7].

Acknowledgements

This lecture was developed in part using materials by Tim Roughgarden, and in particular, his book "Twenty Lectures on Algorithmic Game Theory" [5].

References

[1] American Kidney Fund, Jun 2022.

[2] Itai Ashlagi, Felix Fischer, Ian A Kash, and Ariel D Procaccia. Mix and match: A strategyproof mechanism for multi-hospital kidney exchange. *Games and Economic Behavior*, 91:284–296, 2015.

[3] Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. Kidney exchange. *The Quarterly journal of economics*, 119(2):457–488, 2004.

[4] Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. Pairwise kidney exchange. *Journal of Economic theory*, 125(2):151–188, 2005.

[5] Tim Roughgarden. *Twenty Lectures on Algorithmic Game Theory*. Cambridge University Press, 2016.

[6] Lloyd Shapley and Herbert Scarf. On cores and indivisibility. *Journal of mathematical economics*, 1(1):23–37, 1974.

[7] Panagiotis Toulis and David C Parkes. A random graph model of kidney exchanges: efficiency, individual-rationality and incentives. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 323–332, 2011.