CSC2556
Lecture 4
Impartial Selection;
PageRank; Facility Location
Announcements

• Hope to add a homework question by next lecture

• Reports tentatively due around Feb end
  ➢ But it will help to decide the topic earlier, and start working.

• I’ll put up a list of possible project ideas (in case you cannot find something related to your research)
  ➢ Will also be available to have more meetings during the next two months to help select projects
Recap

- **Utilities:** Voters have underlying numerical utilities
  - Utility of voter \( i \) for alternative \( a = u_i(a) \)
  - Normalization: \( \sum_a u_i(a) = 1 \) for all voters \( i \)

- **Preferences:** Observed ranked preferences of voters are consistent with their implicit utilities
  - \( a \succ_i b \iff u_i(a) > u_i(b) \)

- **Goal:** maximize the (utilitarian) social welfare
  - Ideally, select \( a^* \in \arg\max_a \sum_i u_i(a) \)
  - Cannot achieve this given only ranked preferences
Recap

• **Modified goal:** Achieve the best worst-case approximation to social welfare
  - Denote the utilities of all voters collectively by \( \tilde{u} \)
  - \( \mathcal{U}(P) = \) set of \( \tilde{u} \) consistent with preference profile \( P \)
  - \( sw(a, \tilde{u}) = \sum_i u_i(a) \)
  - **Distortion** when choosing \( a \) is worst-case approximation:

\[
\max_{\tilde{u} \in \mathcal{U}(P)} \frac{\max_b sw(b, \tilde{u})}{sw(a, \tilde{u})}
\]

• **Uniquely optimal rule**
  - Given profile \( P \), choose \( a \) that minimizes the above term
Recap

• **Theorem** [Boutilier et al. ‘12]: Given ranked preferences, the optimal randomized voting rule has distortion $O(\sqrt{m} \cdot \log^* m)$, $\Omega(\sqrt{m})$.

• **Proof:**
  - **Lower bound:** Construct a profile on which every randomized voting rule $\Omega(\sqrt{m})$ distortion.
  - **Upper bound:** Show *some* randomized voting rule that has $O(\sqrt{m} \cdot \log^* m)$ distortion
    - We’ll do the much simpler $O(\sqrt{m} \cdot \log m)$ distortion
Recap

• Proof (lower bound):
  ➢ Consider a similar profile:
    o $\sqrt{m}$ special alternatives
    o Voting rule must choose one of them (say $a^*$) w.p. at most $1/\sqrt{m}$

  ➢ Bad utility profile $\vec{u}$:
    o All voters ranking $a^*$ first give utility 1 to $a^*$
    o All other voters give utility $1/m$ to each alternative
    o $\frac{n}{\sqrt{m}} \leq sw(a^*, \vec{u}) \leq \frac{2n}{\sqrt{m}}$
    o $sw(a, \vec{u}) \leq n/m$ for every other $a$.
    o Distortion lower bound: $\sqrt{m}/3$ (proof on the board!)
• Proof (upper bound):
  ➢ Given profile $P$, define the harmonic score $sc(a, P)$:
    o Each voter gives $1/k$ points to her $k^{th}$ most preferred alternative
    o Take the sum of points across voters
    o $sw(a, \bar{u}) \leq sc(a, P)$ (WHY?)
    o $\sum_a sc(a, P) = n \cdot \sum_{k=1}^{m} 1/k \leq n \cdot (\ln m + 1)$

➢ Golden rule:
  o W.p. $\frac{1}{2}$: Choose every $a$ w.p. proportional to $sc(a, P)$
  o W.p. $\frac{1}{2}$: Choose every $a$ w.p. $1/m$ (uniformly at random)

➢ Distortion $\leq 2\sqrt{m \cdot (\ln m + 1)}$ (proof on the board!)
  o Two cases by comparing $sc(a, P)$ to $n \sqrt{m \cdot (\ln m + 1)/m}$
Recap: Voting in General

• Incentives: a bit tricky to set right
  ➢ Gibbard-Satterthwaite impossibility
  ➢ Ways to circumvent it, but on average, manipulations are not that hard to find

• Approaches to voting
  ➢ Even if we forget about incentives, how do we select a reasonable voting rule?

• Today: Voting in restricted domains
Impartial Selection
Impartial Selection

• “How can we select $k$ people out of $n$ people?”
  ➢ Applications: electing a student representation committee, selecting $k$ out of $n$ grant applications to fund using peer review, ...

• Model
  ➢ Input: a directed graph $G = (V, E)$
  ➢ Nodes $V = \{v_1, \ldots, v_n\}$ are the $n$ people
  ➢ Edge $e = (v_i, v_j) \in E$: $v_i$ supports/approves of $v_j$
    o We do not allow or ignore self-edges $(v_i, v_i)$
  ➢ Output: a subset $V' \subseteq V$ with $|V'| = k$
  ➢ $k \in \{1, \ldots, n - 1\}$ is given
Impartial Selection

• Impartiality: A \( k \)-selection rule \( f \) is impartial if \( v_i \in f(G) \) does not depend on the outgoing edges of \( v_i \)
  - \( v_i \) cannot manipulate his outgoing edges to get selected
  - Q: But the definition says \( v_i \) can neither go from \( v_i \not\in f(G) \) to \( v_i \in f(G) \), nor from \( v_i \in f(G) \) to \( v_i \not\in f(G) \). Why?

• Societal goal: maximize the sum of in-degrees of selected agents \( \sum_{v \in f(G)} |in(v)| \)
  - \( in(v) \) = set of nodes that have an edge to \( v \)
  - \( out(v) \) = set of nodes that \( v \) has an edge to
  - Note: OPT will pick the \( k \) nodes with the highest indegrees
Optimal ≠ Impartial

• An optimal 1-selection rule must select $v_1$ or $v_2$
• The other node can remove his edge to the winner, and make sure the optimal rule selects him instead
• This violates impartiality
Goal: Approximately Optimal

• $\alpha$-approximation: We want a $k$-selection system that always returns a set with total indegree at least $\alpha$ times the total indegree of the optimal set

• Q: For $k = 1$, what about the following rule?

  Rule: “Select the lowest index vertex in $\text{out}(v_1)$. If $\text{out}(v_1) = \emptyset$, select $v_2$.”

  ➢ A. Impartial + constant approximation
  ➢ B. Impartial + bad approximation
  ➢ C. Not impartial + constant approximation
  ➢ D. Not impartial + bad approximation
No Finite Approximation 😞

• **Theorem** [Alon et al. 2011]
  For every $k \in \{1, \ldots, n - 1\}$, there is no impartial $k$-selection rule with a finite approximation ratio.

• **Proof:**
  - For small $k$, this is trivial. E.g., consider $k = 1$.
    - What if $G$ has two nodes $v_1$ and $v_2$ that point to each other, and there are no other edges?
    - For finite approximation, the rule must choose either $v_1$ or $v_2$
    - Say it chooses $v_1$. If $v_2$ now removes his edge to $v_1$, the rule must choose $v_2$ for any finite approximation.
    - Same argument as before. But applies to any “finite approximation rule”, and not just the optimal rule.
No Finite Approximation 😞

• **Theorem** [Alon et al. 2011]
  For every $k \in \{1, \ldots, n - 1\}$, there is no impartial $k$-selection rule with a finite approximation ratio.

• **Proof:**
  - Proof is more intricate for larger $k$. Let’s do $k = n - 1$.
    - $k = n - 1$: given a graph, “eliminate” a node.
  - Suppose for contradiction that there is such a rule $f$.
  - W.l.o.g., say $v_n$ is eliminated in the empty graph.
  - Consider a family of graphs in which a subset of $\{v_1, \ldots, v_{n-1}\}$ have edges to $v_n$. 
No Finite Approximation 😞

• Proof ($k = n - 1$ continued):
  ➢ Consider star graphs in which a non-empty subset of $\{v_1, \ldots, v_{n-1}\}$ have edge to $v_n$, and there are no other edges
    o Represented by bit strings $\{0,1\}^{n-1}\setminus\{0\}$
  ➢ $v_n$ cannot be eliminated in any star graph
    o Otherwise we have infinite approximation
  ➢ $f$ maps $\{0,1\}^{n-1}\setminus\{0\}$ to $\{1, \ldots, n - 1\}$
    o “Who will be eliminated?”
  ➢ Impartiality: $f(\vec{x}) = i \Leftrightarrow f(\vec{x} + \vec{e}_i) = i$
    o $\vec{e}_i$ has 1 at $i^{th}$ coordinate, 0 elsewhere
    o In words, $i$ cannot prevent elimination by adding or removing his edge to $v_n$
No Finite Approximation 😞

• Proof ($k = n - 1$ continued):
  
  $f: \{0,1\}^{n-1}\setminus\{\vec{0}\} \to \{1, \ldots, n - 1\}$

  $f(\vec{x}) = i \iff f(\vec{x} + \vec{e}_i) = i$
  
  $\circ \vec{e}_i$ has 1 only in $i^{th}$ coordinate

  Pairing implies...
  
  $\circ$ The number of strings on which $f$ outputs $i$ is even, for every $i$.
  $\circ$ Thus, total number of strings in the domain must be even too.
  $\circ$ But total number of strings is $2^{n-1} - 1$ (odd)

  So impartiality must be violated for some pair of $\vec{x}$ and $\vec{x} + \vec{e}_i$
Back to Impartial Selection

• **Question:** So what *can* we do to select impartially?
• **Answer:** Randomization!
  ➢ Impartiality now requires that the probability of an agent being selected be independent of his outgoing edges.

• **Examples:** Randomized Impartial Mechanisms
  ➢ Choose $k$ nodes uniformly at random
    o Sadly, this still has arbitrarily bad approximation.
    o Imagine having $k$ special nodes with indegree $n - 1$, and all other nodes having indegree 0.
    o Mechanism achieves $(k/n) \times OPT \Rightarrow$ approximation $= n/k$
    o Good when $k$ is comparable to $n$, but bad when $k$ is small.
Random Partition

• **Idea:**
  - What if we partition $V$ into $V_1$ and $V_2$, and select $k$ nodes from $V_1$ based only on edges coming to them from $V_2$?

• **Mechanism:**
  - Assign each node to $V_1$ or $V_2$ i.i.d. with probability $\frac{1}{2}$
  - Choose $V_i \in \{V_1, V_2\}$ at random
  - Choose $k$ nodes from $V_i$ that have most incoming edges from nodes in $V_{3-i}$
Random Partition

• Analysis:
  ➢ We want to approximate $I = \# \text{ edges incoming to nodes in } OPT$.
    o Let $OPT_1 = OPT \cap V_1$, and $OPT_2 = OPT \cap V_2$.
    o Let $I_1 = \# \text{ edges incoming to } OPT_1 \text{ from } V_2$.
    o Let $I_2 = \# \text{ edges incoming to } OPT_2 \text{ from } V_1$.

  ➢ Note that $E[I_1 + I_2] = I/2$.  (WHY?)

  ➢ With probability $\frac{1}{2}$, mechanism picks $k$ nodes from $V_1$ that have most incoming edges from $V_2$ (thus at least $I_1$ incoming edges).
    o Because they’re at least as good as $OPT_1$.

  ➢ With probability $\frac{1}{2}$, mechanism picks $k$ nodes from $V_2$ that have most incoming edges from $V_1$ (thus at least $I_2$ incoming edges).

  ➢ The expected total incoming edges is at least
    o $E\left[\left(\frac{1}{2}\right) \cdot I_1 + \left(\frac{1}{2}\right) \cdot I_2\right] = \left(\frac{1}{2}\right) \cdot E[I_1 + I_2] = \left(\frac{1}{2}\right) \cdot \frac{I}{2} = \frac{I}{4}$
Random Partition

- **Generalization**
  - Divide into $\ell$ parts, and pick $k/\ell$ nodes from each part based on incoming edges from all other parts.

- **Theorem [Alon et al. 2011]**:
  - $\ell = 2$ gives a $4$-approximation.
  - For $k \geq 2$, $\ell \sim k^{1/3}$ gives $1 + O\left(\frac{1}{k^{1/3}}\right)$ approximation.
Better Approximations

• Alon et al. [2011] conjectured that for randomized impartial 1-selection...
  ➢ (For which their mechanism is a 4-approximation)
  ➢ It should be possible to achieve a 2-approximation.
  ➢ Recently proved by Fischer & Klimm [2014]
  ➢ Permutation mechanism:
    o Select a random permutation \((\pi_1, \pi_2, ..., \pi_n)\) of the vertices.
    o Start by selecting \(y = \pi_1\) as the “current answer”.
    o At any iteration \(t\), let \(y \in \{\pi_1, ..., \pi_t\}\) be the current answer.
    o From \(\{\pi_1, ..., \pi_t\}\setminus\{y\}\), if there are more edges to \(\pi_{t+1}\) than to \(y\), change the current answer to \(y = \pi_{t+1}\).
Better Approximations

• 2-approximation is tight.
  ➢ In an $n$-node graph, fix $u$ and $v$, and suppose no other nodes have any incoming/outgoing edges.
  ➢ Three cases: only $u \rightarrow v$ edge, only $v \rightarrow u$, or both.
    o The best impartial mechanism selects $u$ and $v$ with probability $1/2$ in every case, and achieves 2-approximation.

• But this is because $n - 2$ nodes are not voting!
  ➢ What if every node must have an outgoing edge?
  ➢ Fischer & Klimm [2014]:
    o Permutation mechanism gives $12/7 = 1.714$ approximation.
    o No mechanism gives better than $2/3$ approximation.
    o Open question to achieve better than $12/7$. 
PageRank
Axiomatization
PageRank

• An extension of the impartial selection problem
  ➢ Instead of selecting \( k \) nodes, we want to rank all nodes

• The PageRank Problem: Given a directed graph, rank all nodes by their “importance”.
  ➢ Think of the web graph, where nodes are webpages, and a directed \( (u, v) \) edge means \( u \) has a link to \( v \).

• Questions:
  ➢ What properties do we want from such a rule?
  ➢ What rule satisfies these properties?
PageRank

• Here is the PageRank Algorithm:
  ➢ Start from any node in the graph.
  ➢ At each iteration, choose an outgoing edge of the current node, uniformly at random among all its outgoing edges.
  ➢ Move to the neighbor node on that edge.
  ➢ In the limit of $T \to \infty$ iterations, measure the fraction of time the “random walk” visits each node.
  ➢ Rank the nodes by these “stationary probabilities”.

• Google uses (a version of) this algorithm
  ➢ It’s seems a reasonable algorithm.
  ➢ What nice axioms might it satisfy?
PageRank

• In a formal sense...
  ➢ Let $p_i$ = stationary probability of visiting $i$.
  ➢ Let $N(i)$ = set of nodes that have an edge to $i$.
  ➢ Then, $p_i = \sum_j p_j / \text{outdeg}(j) \Rightarrow n$ equations, $n$ variables!

• Another way to do this:
  ➢ Let $A$ be a matrix with $A_{i,j} = 1 / \text{outdeg}(i)$ for every $(i,j) \in E$.
  ➢ Then, we are searching for a solution $v$ such that $Av = v$.
  ➢ One method: start from any $v_0$, and compute $\lim_{k \to \infty} A^k v_0$
    o Note: $A^k$ can be computed using $\log k$ matrix multiplications!
Axioms

• Axiom 1 (Isomorphism)
  ➢ Permuting node names permutes the final ranking.

• Axiom 2 (Vote by Committee)
  ➢ Voting through intermediate fake nodes cannot change the ranking.

• Axiom 3 (Self Edge)
  ➢ $v$ adding a self edge cannot change the ordering of the other nodes.

• Axiom 4 (Collapsing)
  ➢ Merging identically voting nodes cannot change the ordering of the other nodes.

• Axiom 5 (Proxy)
  ➢ If $k$ nodes with equal score vote for $k$ other nodes through a proxy, it should be no different than a direct 1-1 voting.
PageRank

• Theorem [Altman and Tennenholtz, 2005]: An algorithm satisfies these five axioms if and only if it is PageRank.
Facility Location
Apprx Mechanism Design

1. Define the problem: agents, outcomes, values

2. Fix an objective function (e.g., maximizing sum of values)

3. Check if the objective function is maximized through a strategyproof mechanism

4. If not, find the strategyproof mechanism that provides the best worst-case approximation ratio of the objective function
Facility Location

- Set of agents \( N \)
- Each agent \( i \) has a true location \( x_i \in \mathbb{R} \)
- Mechanism \( f \)
  - Takes as input reports \( \tilde{x} = (\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n) \)
  - Returns a location \( y \in \mathbb{R} \) for the new facility
- Cost to agent \( i \) : \( c_i(y) = |y - x_i| \)
- Social cost \( C(y) = \sum_i c_i(y) = \sum_i |y - x_i| \)
• Social cost \( C(y) = \sum_i c_i(y) = \sum_i |y - x_i| \)

• **Q:** Ignoring incentives, what choice of \( y \) would minimize the social cost?

• **A:** The median location \( \text{med}(x_1, \ldots, x_n) \)
  - \( n \) is odd → the unique \( \text{“}(n+1)/2\text{”}^{\text{th}} \) smallest value
  - \( n \) is even → \( \text{“}n/2\text{”}^{\text{th}} \) or \( \text{“}(n/2)+1\text{”}^{\text{st}} \) smallest value
  - Why?
Facility Location

- Social cost $C(y) = \sum_i c_i(y) = \sum_i |y - x_i|$
- Median is optimal (i.e., 1-approximation)

What about incentives?

- Median is also strategyproof (SP)!
- Irrespective of the reports of other agents, agent $i$ is best off reporting $x_i$
Median is SP

No manipulation can help
Max Cost

• A different objective function $C(y) = \max_{i} |y - x_i|$

• Q: Again ignoring incentives, what value of $y$ minimizes the maximum cost?

• A: The midpoint of the leftmost ($\min_{i} x_i$) and the rightmost ($\max_{i} x_i$) locations

• Q: Is this optimal rule strategyproof?

• A: No!
Max Cost

- \( C(y) = \max_i |y - x_i| \)

- We want to use a strategyproof mechanism.

- **Question:** What is the approximation ratio of median for maximum cost?
  1. \( \in [1,2) \)
  2. \( \in [2,3) \)
  3. \( \in [3,4) \)
  4. \( \in [4,\infty) \)
Max Cost

• **Answer:** 2-approximation

• Other SP mechanisms that are 2-approximation
  - Leftmost: Choose the leftmost reported location
  - Rightmost: Choose the rightmost reported location
  - Dictatorship: Choose the location reported by agent 1
  - ...

Max Cost

• Theorem [Procaccia & Tennenholtz, ‘09]
  No deterministic SP mechanism has approximation ratio < 2 for maximum cost.

• Proof:
Max Cost + Randomized

• The Left-Right-Middle (LRM) Mechanism
  ➢ Choose $\min x_i$ with probability $\frac{1}{4}$
  ➢ Choose $\max x_i$ with probability $\frac{1}{4}$
  ➢ Choose $(\min x_i + \max x_i)/2$ with probability $\frac{1}{2}$

• Question: What is the approximation ratio of LRM for maximum cost?

• At most \( \frac{(1/4)*2C+(1/4)*2C+(1/2)*C}{C} = \frac{3}{2} \)
Max Cost + Randomized

• Theorem [Procaccia & Tennenholtz, ‘09]: The LRM mechanism is strategyproof.

• Proof:

\[1/4\] \quad 2\delta \quad \delta \quad 1/4 \]

\[1/4\] \quad 1/2 \quad 1/4\]
Max Cost + Randomized

- Exercise for you!
  Try showing that no randomized SP mechanism can achieve approximation ratio $< 3/2$. 