Harvard CS 1360, Spring 2025
Instructor: Ariel Procaccia
Scribe: Taig Singh

The Price of Anarchy

—

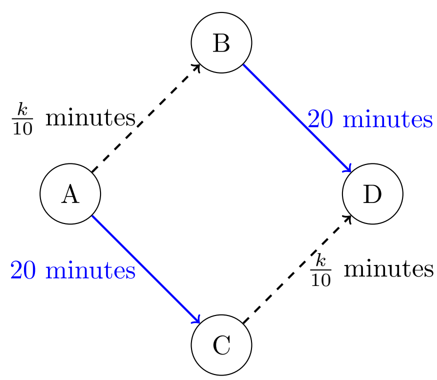
Lecture 4

We'll start off this lecture by looking at an interesting phenomenon: closing roads oftentimes speeds up traffic. For example, a well-known instance of this is in Seoul, South Korea, where the city closed a major highway along the Cheonggyecheon Stream to revitalize the area. Counterintuitively, among other benefits, traffic flow actually improved as drivers redistributed across other routes. This counterintuitive effect, known as Braess's Paradox, challenges our typical assumptions about routing games.

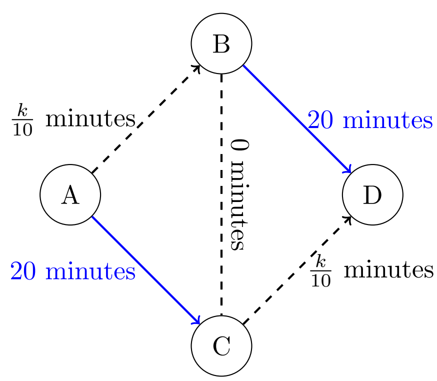
Example 1 (Braess' Paradox)**.** Consider a situation in which 200 travelers want to get from work (A) to home (D). Initially, each traveler can take one of two paths:

- Go from A to B, taking $\frac{k}{10}$ minutes for each traveler where k is the number of travelers who take this path, and then go from B to D, which takes 20 minutes for each traveler.
- Go from A to C, taking $\frac{k}{10}$ minutes for each traveler where k is the number of travelers who take this path, and then go from C to D, which takes 20 minutes for each traveler.

A graphical depiction of these paths is provided below.



Here, in Nash equilibrium, 100 travelers take each route and the travel time for each traveler is 30 minutes, because both A to B and C to D take $\frac{100}{10} = 10$ minutes. Note that this is a Nash equilibrium because a unilateral deviation by any traveler results in their traveling time being $20 + \frac{101}{10} = 30.1$ minutes. Now, consider the addition of a 2-way street between B and C as shown below:



Now, in Nash equilibrium, all travelers take the path from $A \to B \to C \to D$, and so every traveler takes $\frac{200}{10} + \frac{200}{10} = 40$ minutes to get to D. Note that no traveler has a useful unilateral deviation because flipping to any of $A \to B \to D$, $A \to C \to D$, or $A \to C \to B \to D$ all result in the same 40 minutes of travel.

This displays Braess' paradox: the Nash equilibrium outcome by adding a road results in longer travel times than the Nash equilibrium without the road.

We will now provide a general definition for routing games:

Definition 1 (Routing Games)**.** An (atomic) routing game consists of

- A set of players $N = \{1, \ldots, n\}$
- A directed graph $G = (V, E)$
- For each $e \in E$, a nonnegative and nondecreasing cost function $c_e : \mathbb{N} \to \mathbb{R}^+$
- For each $i \in N$, a source and sink $a_i, b_i \in V$

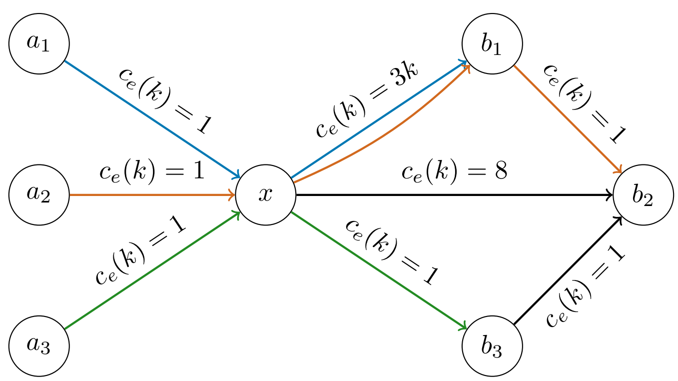
The strategy set S_i of each i is all paths from a_i to b_i. In a strategy profile $\mathbf{s}$, $n_e(s)$ players are using edge e. The cost of player i is

$$\text{cost}_i(\mathbf{s}) = \sum_{e \in s_i} c_e(n_e(\mathbf{s}))$$

Finally, the social cost is

$$\text{cost}(\mathbf{s}) = \sum_{i \in N} \text{cost}_i(\mathbf{s})$$

Example 2 (Routing Game)**.** Consider the following routing game between 3 players. Player i's source is labeled as a_i and their sink is labeled as b_i. Each edge's cost function c_e is written next to the edge as a function of the number of players k who traverse that edge.



If player 1 takes the blue path, player 2 takes the orange path, and player 3 takes the green path, player 1's cost is $1 + 3 \cdot 2 = 7$, player 2's cost is $1 + 3 \cdot 2 + 1 = 8$, and player 3's cost is $1 + 1 = 2$. Thus, we get a social cost of 17. In the optimal solution, player 1 takes $a_1 \to x \to b_1$, player 2 takes $a_2 \to x \to b_3 \to b_2$, and player 3 takes $a_3 \to x \to b_3$. Thus, player 1's cost is $1 + 3 \cdot 1 = 4$, player 2's cost is $1 + 1 + 1 = 3$, and player 3's cost is $1 + 1 = 2$, resulting in a social cost of 9.

We are talking about pure Nash equilibria, but how do we know they even exist? We will characterize the existence of pure Nash equilibria by first defining the notion of an "exact potential game."

Definition 2 (Exact Potential Games)**.** A game is an exact potential game if there exists a function $\Phi : \prod_{i=1}^n S_i \to \mathbb{R}$ such that for all $i \in N$, for all $\mathbf{s} \in \prod_{i=1}^n S_i$, and for all $s_i' \in S_i$,

$$\text{cost}_i(s_i', \mathbf{s}_{-i}) - \text{cost}_i(\mathbf{s}) = \Phi(s_i', \mathbf{s}_{-1}) - \Phi(\mathbf{s})$$

In other words, for any strategy profile, when player i unilaterally deviates from their original strategy, the change in value of Φ equals the change in player i's cost. We call Φ an exact potential function for the game.

The existence of an exact potential function implies the existence of a pure Nash equilibrium, and the set of pure Nash equilibria can be found by locating the local minima of the exact potential function. This is because, at these local minima, no player can improve their cost by switching routes, as the potential function exactly maps the increases of individual costs by deviations of individual players. Thus, as players repeatedly make cost-reducing moves, the potential function decreases and converges to a local minimum, where no player can improve their cost by switching routes.

We will now show that routing games actually do fall under this category of games.

Theorem 1. *Routing games are exact potential games.*

Proof. Define the potential function

$$\Phi(\mathbf{s}) = \sum_{e \in E} \sum_{k=1}^{n_e(\mathbf{s})} c_e(k)$$

Now, suppose player i deviates from path s_i to path s_i'. Then, the difference $\text{cost}_i(s_i', s_{-i}) - \text{cost}_i(\mathbf{s})$ is just

$$\sum_{e \in s_i' \setminus s_i} c_e(n_e(\mathbf{s}) + 1) - \sum_{e \in s_i \setminus s_i'} c_e(n_e(\mathbf{s}))$$

Note that here, the first term is the cost imposed by the edges added to $\mathbf{s}$ and the second term is the cost removed by the edges removed from $\mathbf{s}$. We claim that this is precisely $\Phi(s_i', s_{-i}) - \Phi(\mathbf{s})$. This follows because to move from $\Phi(\mathbf{s})$ to $\Phi(s_i', s_{-i})$, only the terms for the edges that are in s_i' and were not in s_i, and the edges that were in s_i and are now not in s_i', are affected. For the edges e that are in s_i' and were not in s_i, we must add $c_e(n_e(\mathbf{s} + 1))$ because now player i is using that edge and they were not before, so we now have $\sum_{k=1}^{n_e(\mathbf{s})+1} c_e(k)$ within the term for e in the formula for $\Phi(s_i', s_{-i})$. Further, for the edges e that are not in s_i' but were in s_i, we must remove $c_e(n_e(\mathbf{s}))$ because now player i is now not using that edge, so we now have $\sum_{k=1}^{n_e(\mathbf{s})-1} c_e(k)$ within the term for e in the formula for $\Phi(s_i', s_{-i})$.

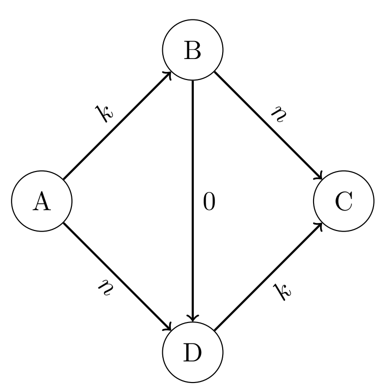
Thus, routing games are exact potential games, implying that all routing games have a pure Nash equilibrium. □

We will now look at the upper bounds of how bad the outcomes of equilibria can be as compared to the optimal solution in different classes of games. This is known as the price of anarchy

Definition 3 (Price of Anarchy (PoA))**.** Fix a class of games, an objective function, and an equilibrium concept. The price of anarchy is the worst-case ratio between the worst objective function value of an equilibrium of the game, and that of the optimal solution.

For the rest of this lecture, our objective function will be the social cost, and our equilibrium concept will be pure Nash equilibria.

Example 3 (PoA of Routing Games)**.** Consider the following routing game, where n players wish to go from A to C. Here, k represents the number of players going through the path.



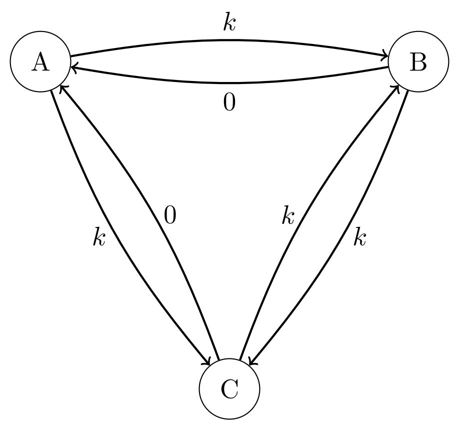
In this game, a pure Nash equilibrium is all n players taking the path $A \to B \to D \to C$, resulting in a travel time of $2n$ for everyone. Note that this is a Nash equilibrium because any single player's deviation to $A \to D \to C$ or $A \to B \to C$ results in the same $2n$ travel time. Here, the social cost is just $2n \cdot n = 2n^2$.

However, in this case, the social-cost-optimal solution is to send $\frac{n}{2}$ players along the path $A \to B \to C$ and $\frac{n}{2}$ players along the path $A \to D \to C$, resulting in a social cost of $\frac{n}{2} \cdot \frac{n}{2} + \frac{n}{2} \cdot n + \frac{n}{2} \cdot n + \frac{n}{2} \cdot \frac{n}{2} = \frac{3n^2}{2}$. Thus, we get that a lower bound for the price of anarchy in this case is just

$$\frac{2n^2}{\frac{3n^2}{2}} = \frac{4}{3}$$

Now, consider another routing game, with 4 players where the set of (a_i, b_i) pairs is

$$\{(A, B), (A, C), (B, C), (C, B)\}$$



In this game, the optimal solution is for all players to take the path with cost k directly to their sink. Thus, player 1 will take $A \to B$, player 2 will take $A \to C$, player 3 will take $B \to C$, and player 4 will take $C \to B$. This results in an overall social cost of

$$1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 = 4$$

Note that this optimal solution is also clearly a Nash equilibrium. To see this, for any player, consider any other path they could take to their sink. In every other path, the player must traverse at least one edge with cost k, and thus this would not be a useful deviation for them. However, this is not the only Nash equilibrium in the game.

Consider player 1 taking $A \to C \to B$, player 2 taking $A \to B \to C$, player 3 taking $B \to A \to C$, and player 4 taking $C \to A \to B$. Here, player 1's cost is $2 + 1 = 3$, player 2's cost is $2 + 1 = 3$, player 3's cost is $0 + 2 = 2$, and player 4's cost is $0 + 2 = 2$, resulting in an overall social cost of 10. To see that this is a Nash equilibrium, it is clear that any player's deviation to a path of length 3 or more is not useful, and deviating to a path of length 1 will result in the same exact individual cost. To see this more clearly, if player 1 deviated to $A \to B$, players 1. 2. and 4 would be taking $A \to B$ so player 1 will have a cost of 3. Similarly, if player 2 deviated to $A \to C$ they would have a cost of 3, if player 3 deviated to $B \to C$ they would have a cost of 2, and if player 4 deviated to $C \to B$ they would have a cost of 2. Thus, this strategy profile is Nash equilibrium. Therefore, we get that a lower bound for the price of anarchy, in this case, is just

$$\frac{10}{4} = 2.5$$

We can actually upper bound the price of anarchy for any routing game with linear cost functions with the following theorem.

Theorem 2. *For any routing game with linear cost functions, the price of anarchy is at most* 2.5

Now, we have that all in all,

- For any routing game with linear cost functions, for any Nash equilibrium $\mathbf{s}$, we know that $\text{cost}(\mathbf{s}) \leq 2.5 \cdot \text{cost}(\text{OPT})$.
- There exists a routing game with linear cost functions, such that it has a Nash equilibrium $\mathbf{s}$ where $\text{cost}(\mathbf{s}) \geq 2.5 \cdot \text{cost}(\text{OPT})$

Example 4 (Routing Game with Non-linear Cost Functions)**.** Consider a routing game in which there are n players and two edges from a common source to a common sink, one with cost 1 and one with cost $(k/n)^p$. We wish to find a lower bound on the PoA as $p \to \infty$.

Note that a Nash equilibrium will always be for all n people to take the path with cost $(k/n)^p$, because in this case, the cost imposed on each individual player is just $(n/n)^p = 1$ so deviating to the other edge for a fixed cost of 1 is not useful. This Nash equilibrium results in a total social cost of

$$n \cdot \left(\frac{n}{n}\right)^p = n \cdot 1 = n$$

On the other hand, consider the strategy profile where 1 player chooses to take the edge with a fixed cost of 1 and $n-1$ players take the path with a cost of $(k/n)^p$. Then, the total social cost is

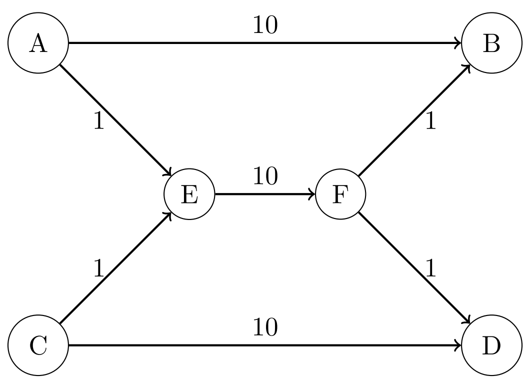
$$n \cdot \left(\frac{n-1}{n}\right)^p + 1$$

However, as $p \to \infty$, the first term goes to 0. Thus, we get a lower bound of the price of anarchy in this game is just $\frac{n}{1} = n$

We will now introduce a different, but related class of games known as "cost sharing games."

Definition 4 (Cost Sharing Games)**.** Cost sharing games are the same as routing games with one exception: each edge has a fixed cost c_e that is split among the players using it.

Example 5 (Basic Cost Sharing Game)**.** Consider the following cost sharing game played between two players. The first player wishes to travel from A to B and the second player wishes to travel from C to D.



Consider the strategy profile where player 1 takes $A \to B$ and player 2 takes $C \to D$. This strategy profile has a social cost of $10 + 10 = 20$. This is a Nash equilibrium as if either player deviates to taking the middle path to get to their sink, their cost will increase from 10 to 12. However, note that the optimal solution is when player 1 takes $A \to E \to F \to B$ and player 2 takes $C \to E \to F \to D$ for a total social cost of $2(1 + 5 + 1) = 14$.

Example 6 (Lower Bounding PoA for Cost Sharing Games)**.** Consider a cost sharing game with n players and two edges between a common source and a common sink with costs 1 and n. Now, consider a strategy profile where all n players take the edge with cost n, achieving a total social cost of n as this cost is split across the n players where each player has an individual cost of 1. This strategy profile is a Nash equilibrium, as no player would benefit from deviating to the path with a cost of 1. However, the optimal solution is for all players to take the path with a cost of 1 achieving a total social cost of 1. Thus, we get a lower bound for the price of anarchy of cost sharing games of $\frac{n}{1} = n$.

It turns out that this bound of n is an upper bound for the PoA of cost sharing games as well!

Theorem 3. *The price of anarchy of cost sharing games is at most n*

Proof. Let $\mathbf{s}$ be a Nash equilibrium and let $\mathbf{s}^\star$ be an optimal solution. For all i, it holds that $\text{cost}_i(\mathbf{s}) \leq \text{cost}_i(s_i^\star, \mathbf{s}_{-i})$ because i can't gain from unilaterally deviating in a NE. But $\text{cost}_i(s_i^\star, \mathbf{s}_{-i}) \leq n \cdot \text{cost}_i(\mathbf{s}^\star)$, because in the worst case i pays for its path alone in the former and splits each edge cost n ways in the latter. Summing over i, we get that $\text{cost}(\mathbf{s}) \leq n \cdot \text{cost}(\mathbf{s}^\star)$, and so

$$\frac{\text{cost}(\mathbf{s})}{\text{cost}(\mathbf{s}^\star)} \leq n$$

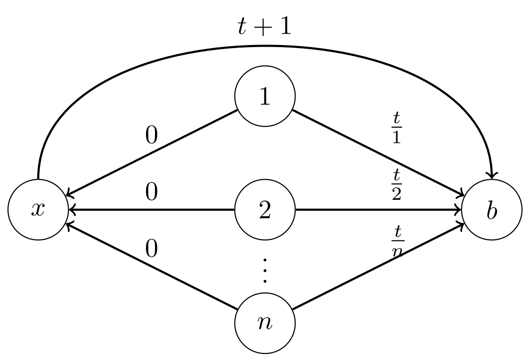
as desired. □

Definition 5 (Price of Stability (PoS))**.** Fix a class of games, an objective function, and an equilibrium concept. The price of stability is the worst-case ratio between the best objective function value of an equilibrium of the game, and that of the optimal solution.

Example 7 (PoS)**.** Consider a cost sharing game with n players and two edges between a common source and a common sink with costs 1 and n. Previously, we identified that the strategy profile where all players take the path with cost 1 is the socially optimal solution. We further see that this is a Nash equilibrium because the unilateral deviation of any player to the path of cost n will increase their individual cost from $\frac{1}{n}$ to n. Thus, we find that a lower bound on price of stability in this case is just

$$\frac{\text{cost(Nash)}}{\text{cost(OPT)}} = \frac{1}{1} = 1$$

Example 8 (A Tighter Lower Bound on PoS for Cost Sharing Games)**.** Consider a the following cost sharing game with n players, where each player i wishes to travel from node i to b.



Here, every player i has the option of choosing their individual path for a cost of t/i or they can jointly build a path that has a total cost of $t+1$. Consider when $t \to \infty$. Every player taking their own path for t/i is a Nash equilibrium because no individual player would benefit from paying $t+1$ by going through node x.

Further, we know that this is the only Nash equilibrium. We will show this by proving no player can go through x in a Nash equilibrium. Assume for the sake of contradiction that player n goes through x in a NE. Then, the least they could possibly pay is $\frac{t+1}{n}$, which is when all n players go through x, but this can't be a NE because player n is better off going straight to b by paying $\frac{t}{n}$. Thus, player n can't go through x in a NE. Similarly, assume for the sake of contradiction that player $n-1$ goes through x in a NE. Then, the least they could pay by going through x is $\frac{t+1}{n-1}$, which is when players 1 through $n-1$ go through x. But then player $n-1$ is better off going straight to b by paying $\frac{t}{n-1}$. Thus, player $n-1$ can't go through x in a NE. This process continues for all other players, showing that the aforementioned strategy profile is the only Nash equilibrium.

The social cost of this Nash equilibrium is

$$\frac{t}{1} + \frac{t}{2} + \cdots + \frac{t}{n} = tH(n)$$

where $H(n)$ is the nth harmonic number.

However, note that the socially optimal solution is when all n players go through x, incurring a total social cost of $t+1$. We then see that as $t \to \infty$, we get a lower bound on the PoS of $H(n) = \Omega(\log n)$.

Like routing games, cost sharing games exact potential games with essentially the same potential function:

$$\Phi(\mathbf{s}) = \sum_{e \in E} \sum_{k=1}^{n_e(\mathbf{s})} \frac{c_e}{k}$$

Using this, we now present a theorem on the price stability of cost sharing games.

Theorem 4. *The price stability of cost sharing games is $O(\log n)$.*

Proof. We know that $\text{cost}(\mathbf{s}) \leq \Phi(\mathbf{s}) \leq H(n) \cdot \text{cost}(\mathbf{s})$. Thus, if we take a strategy profile $\mathbf{s}^\star$ that minimizes Φ, as described earlier, we know that $\mathbf{s}^\star$ is a NE. Thus, we get that

$$\text{cost}(\mathbf{s}^\star) \leq \Phi(\mathbf{s}^\star) \leq \Phi(\text{OPT}) \leq H(n) \cdot \text{cost}(\text{OPT})$$

as desired. □

In summary, we have that for cost sharing games

- Upper bounds: For any cost sharing game,
  - The PoA gives us that for any NE $\mathbf{s}$

    $$\text{cost}(\mathbf{s}) \leq n \cdot \text{cost}(\text{OPT})$$

  - The PoS gives us that there exists a NE $\mathbf{s}$ such that

    $$\text{cost}(\mathbf{s}) \leq H(n) \cdot \text{cost}(\text{OPT})$$

- Lower bounds: There exists a cost sharing game such that
  - The PoA gives us that there exists a NE $\mathbf{s}$ such that

    $$\text{cost}(\mathbf{s}) \geq n \cdot \text{cost}(\text{OPT})$$

  - The PoS gives us that for any NE $\mathbf{s}$

    $$\text{cost}(\mathbf{s}) \geq H(n) \cdot \text{cost}(\text{OPT})$$