Towards Equilibrium Theory in Data Markets

Bhaskar Ray Chaudhury
University of Illinois at Urbana-Champaign

December 16, 2025

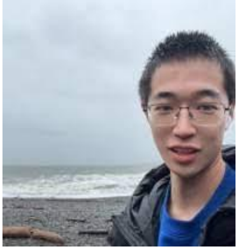
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Ongoing Work

Jugal Garg
UIUC

Aniket Murhekar
UIUC -> Northwestern

Eklavya Sharma
UIUC



Jiaxin Song
UIUC

Data: An Invaluable Asset in the Digital World

ACUMEN RESEARCH AND CONSULTING

Home | Industries | Services | Press Releases | About Us | Blogs

Home > Industry > ICT > Big Data Market

Big Data Market Size - Global Industry, Share, Analysis, Trends and Forecast 2022 - 2030

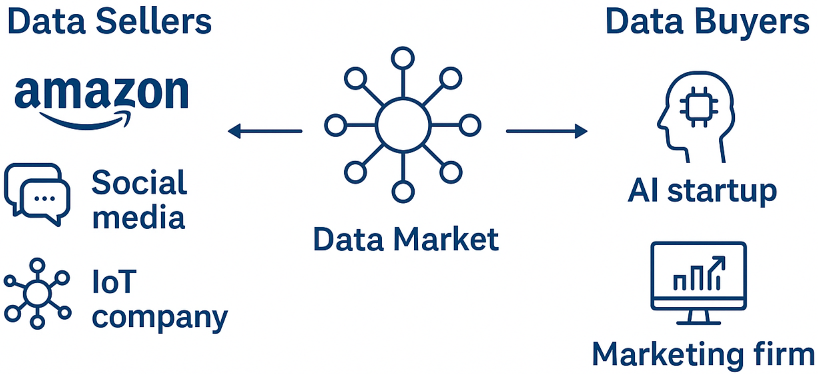
Published : Dec 2022 Report ID: ARC3093 Pages : 250 Format :

Summary | Table of Content | Customization | Download Sample | Infographics

The Global Big Data Market Size accounted for USD 163.5 Billion in 2021 and is projected to occupy a market size of USD 473.6 Billion by 2030 growing at a CAGR of 12.7% from 2022 to 2030.

Big data is primarily intended to analyze, process, and extract information from massive amounts of data and extremely complex structures. Big data analytics are widely associated with many other massively augmented technologies such as artificial intelligence (AI), deep learning, machine learning, and the Internet of Things (IoT) among

Equilibrium Theory in Data Markets



Equilibrium Prices

Market Structures– Fundamentals

Monopoly

One seller dominates the market

Seller is *price-setter*

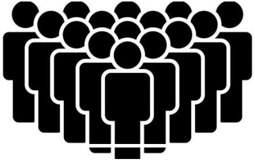
Price is set to *maximize revenue*

Oligopoly

Few sellers compete strategically

Sellers play a *pricing game*

Price is set to a *NE* of the *pricing game*



Perfect Competition

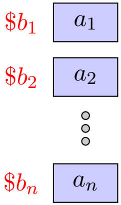
Many sellers

Sellers are *price-takers*

Price set to *match demand and supply*

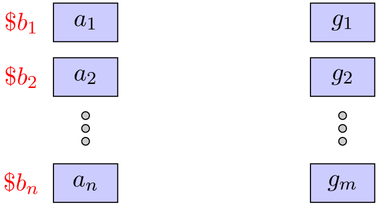
Overview of Solution Concept(s) in Traditional (*Rivalrous*) Markets

(Rivalrous) Fisher Market



n buyers with budgets $\boldsymbol{b}$

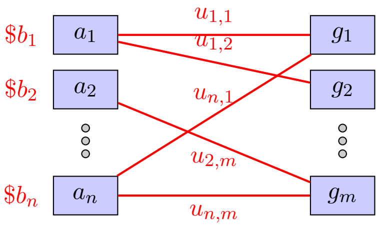
(Rivalrous) Fisher Market



n buyers with budgets $\boldsymbol{b}$

m divisible goods

(Rivalrous) Fisher Market

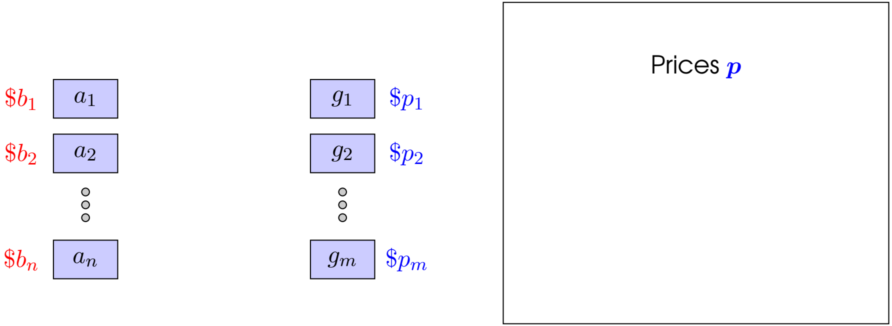


n buyers with budgets $\boldsymbol{b}$

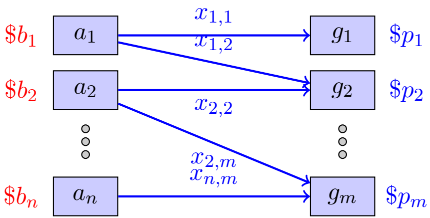
m divisible goods

Utility matrix $\boldsymbol{u}$

Perfect Competition Solution Concept– Competitive Equilibrium (CE)



Perfect Competition Solution Concept– Competitive Equilibrium (CE)



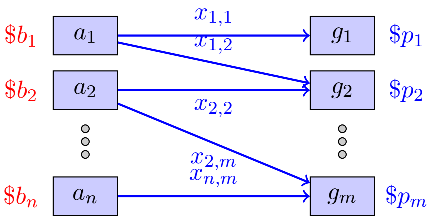
Prices $\boldsymbol{p}$

Optimum demand bundles $\boldsymbol{x}$

$x_i = (x_{i,1}, \ldots x_{i,m})$ is utility

maximizing bundle for i at $\boldsymbol{p}$

Perfect Competition Solution Concept– Competitive Equilibrium (CE)



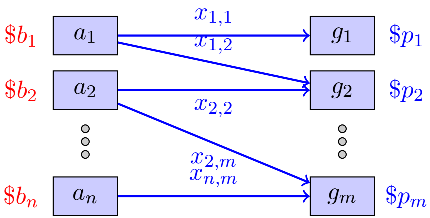
Prices $\boldsymbol{p}$

Optimum demand bundles $\boldsymbol{x}$

$\boldsymbol{x}_i$ maximizes $\sum_j u_{i,j} x_{i,j}$

subject to $\sum_j p_j x_{i,j} \leq b_i$

Perfect Competition Solution Concept– Competitive Equilibrium (CE)

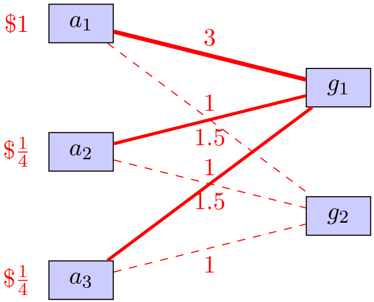


Prices $\boldsymbol{p}$

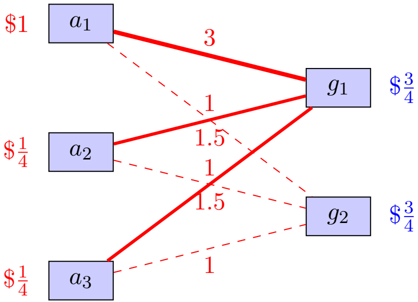
Optimum demand bundles $\boldsymbol{x}$

$(\boldsymbol{p}, \boldsymbol{x})$ is a CE iff $\sum_i x_{i,j} = s_j \;\; \forall j$

Perfect Competition Solution Concept– Competitive Equilibrium (CE)



Perfect Competition Solution Concept– Competitive Equilibrium (CE)

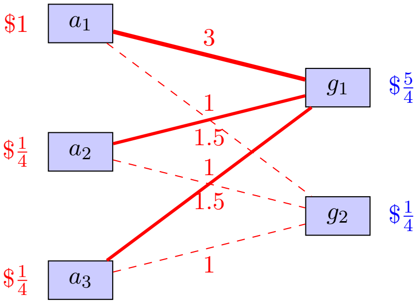


$\boldsymbol{x}_1 = (\frac{4}{3}, 0)$, $\boldsymbol{x}_2 = (\frac{1}{3}, 0)$, $\boldsymbol{x}_3 = (\frac{1}{3}, 0)$

$\sum_i x_{i,1} = 2$ (over-demanded)
$\sum_i x_{i,2} = 0$ (under-demanded)

Not a CE !

Perfect Competition Solution Concept– Competitive Equilibrium (CE)

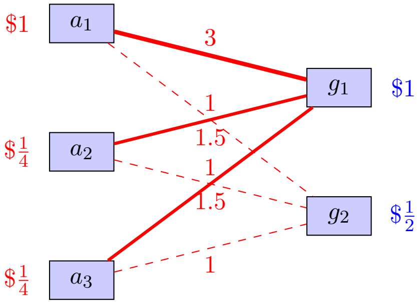


$\boldsymbol{x}_1 = (0,4)$, $\boldsymbol{x}_2 = (0,1)$, $\boldsymbol{x}_3 = (0,1)$

$\sum_i x_{i,1} = 0$ (under-demanded)
$\sum_i x_{i,2} = 6$ (over-demanded)

Not a CE !

Perfect Competition Solution Concept– Competitive Equilibrium (CE)



$\boldsymbol{x}_1 = (1, 0)$, $\boldsymbol{x}_2 = (0, \frac{1}{2})$, $\boldsymbol{x}_3 = (0, \frac{1}{2})$

$\sum_i x_{i,1} = 1$
$\sum_i x_{i,2} = 1$

CE !

Existence and Computation of CE

- A CE always exists [**Arrow and Debreu, *Econometrica*'1954**]
- Convex program exists [**Eisenberg and Gale. *Management Science*'1968**]
- Polynomial time algorithms exist [**Devanur, Papadimitriou, Saberi, Vazirani, *Journal of the ACM*'08**]
- Strongly Polynomial time algorithm exist [**Orlin, *STOC*'10**]
- Intuitive dynamics with fast convergence exist. [**Codenotti, McCune, Varadarajan, *STOC*'05**]

Oligopoly (Fixed Supply) – Pricing Game

- Seller j prices good at p_j,

Oligopoly (Fixed Supply) – Pricing Game

- Seller j prices good at p_j,
- Utility/ Revenue of seller j is $rev_j(p_j, \boldsymbol{p}_{-j}) = p_j \cdot \sum_i x^*_{i,j}(\boldsymbol{p})$,

Oligopoly (Fixed Supply) – Pricing Game

- Seller j prices good at p_j,
- Utility/ Revenue of seller j is $rev_j(p_j, \boldsymbol{p}_{-j}) = p_j \cdot \sum_i x^*_{i,j}(\boldsymbol{p})$,
- $\boldsymbol{p}$ is stable if it is a pure Nash Equilibrium (NE) of the pricing game.

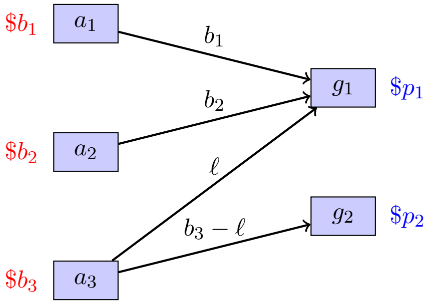
Oligopoly (Fixed Supply) – Pricing Game

- Seller j prices good at p_j,
- Utility/ Revenue of seller j is $rev_j(p_j, \boldsymbol{p}_{-j}) = p_j \cdot \sum_i x^*_{i,j}(\boldsymbol{p})$,
- $\boldsymbol{p}$ is stable if it is a pure Nash Equilibrium (NE) of the pricing game.

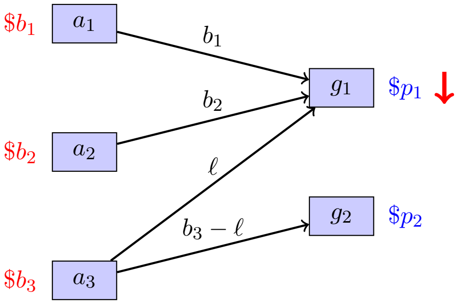
CE $\implies$ NE

If $(\boldsymbol{p}, \boldsymbol{x})$ is a CE, then $\boldsymbol{p}$ is a NE of the pricing game.

Oligopoly (Fixed Supply) – Pricing Game

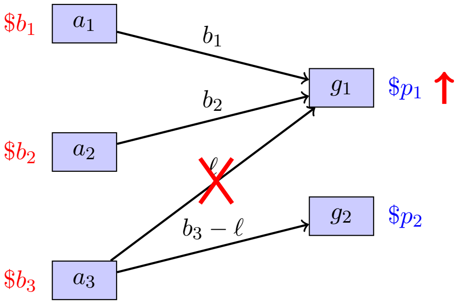


Oligopoly (Fixed Supply) – Pricing Game



$\sum_i x^*_{i,j}$ does not increase

Oligopoly (Fixed Supply) – Pricing Game



$\sum_i p_j \cdot x^*_{i,j}$ does not increase

Monopoly

- Seller prices j at p_j,

Monopoly

- Seller prices j at p_j,
- Buyer i demands $\boldsymbol{x}_i^*(\boldsymbol{p})$,

Monopoly

- Seller prices j at p_j,
- Buyer i demands $\boldsymbol{x}_i^*(\boldsymbol{p})$,
- $\boldsymbol{p}$ is stable if it is a Stackelberg Equilibrium (SE), i.e., it maximizes $\sum_j p_j(\sum_i x_{i,j}^*(\boldsymbol{p}))$.

Monopoly

- Seller prices j at p_j,
- Buyer i demands $\boldsymbol{x}_i^*(\boldsymbol{p})$,
- $\boldsymbol{p}$ is stable if it is a Stackelberg Equilibrium (SE), i.e., it maximizes $\sum_j p_j(\sum_i x_{i,j}^*(\boldsymbol{p}))$.

Observation

Problem is trivial if buyers have no value for money, i.e., $u_i(\boldsymbol{x}_i)$ is independent of $\sum_j p_j x_{i,j}$.

Monopoly with Quasi-Linear Utilities

- Seller prices j at p_j,
- Buyer i demands $\boldsymbol{x}_i^*(\boldsymbol{p})$, where, $\boldsymbol{x}_i^*(\boldsymbol{p})$ maximizes $\sum_j (u_{i,j} - p_j) x_{i,j}$ such that $x_{i,j} \in [0, 1]$ and $\sum_j p_j x_{i,j} \leq b_i$,
- $\boldsymbol{p}$ is stable if it is a Stackelberg Equilibrium (SE), i.e., it maximizes $\sum_j p_j (\sum_i x_{i,j}^*(\boldsymbol{p}))$.

Monopoly with Quasi-Linear Utilities

- Seller prices j at p_j,
- Buyer i demands $\boldsymbol{x}_i^*(\boldsymbol{p})$, where, $\boldsymbol{x}_i^*(\boldsymbol{p})$ maximizes $\sum_j (u_{i,j} - p_j) x_{i,j}$ such that $x_{i,j} \in [0, 1]$ and $\sum_j p_j x_{i,j} \leq b_i$,
- $\boldsymbol{p}$ is stable if it is a Stackelberg Equilibrium (SE), i.e., it maximizes $\sum_j p_j (\sum_i x_{i,j}^*(\boldsymbol{p}))$.

Finster, Goldberg and Lock '24

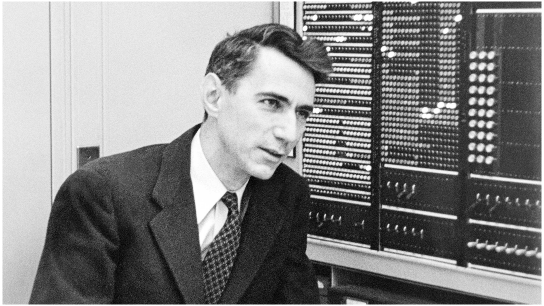
CE $\implies$ SE when agents have quasi-linear utilities.

Data as a Homogeneous Commodity and Value for Data

Value of Data

Shannon's Information Theoretic Perspective

"Data reduces uncertainty"



Value of Data

Data as a Signal

- Agent has prior belief of an unknown state θ,

Value of Data

Data as a Signal

- Agent has prior belief of an unknown state θ,
- Data is a *signal* s about θ,

Value of Data

Data as a Signal

- Agent has prior belief of an unknown state θ,
- Data is a *signal* s about θ,
- Agent updates belief to the posterior of $\theta \mid s$,

Value of Data

Data as a Signal

- Agent has prior belief of an unknown state θ,
- Data is a *signal* s about θ,
- Agent updates belief to the posterior of $\theta \mid s$,
- Value of data $=$ reduction in variance from θ to $\theta \mid s$

Value of Data

- Each data record of a seller is a noisy signal $s = \theta + \eta$, where $\eta \sim \mathcal{D}$,

Value of Data

- Each data record of a seller is a noisy signal $s = \theta + \eta$, where $\eta \sim \mathcal{D}$,
- Value of the data record is $\mathrm{Var}(\theta) - \mathbb{E}[\mathrm{Var}(\theta \mid s)]$

Value of Data

- Each data record of a seller is a noisy signal $s = \theta + \eta$, where $\eta \sim \mathcal{D}$,
- Value of the data record is $\mathrm{Var}(\theta) - \mathbb{E}[\mathrm{Var}(\theta \mid s)] = \mathrm{Var}(\mathbb{E}[\theta \mid s])$ (law of total variance)

Value of Data Bundles

- n *buyers*. Buyer i has budget b_i, and an unknown state θ_i to predict,

Value of Data Bundles

- n *buyers*. Buyer i has budget b_i, and an unknown state θ_i to predict,
- m *sellers*. Seller j owns dataset D_j with s_j data records,

Value of Data Bundles

- n *buyers*. Buyer i has budget b_i, and an unknown state θ_i to predict,
- m *sellers*. Seller j owns dataset D_j with s_j data records,
- Each data record of seller j, is a noisy signal about θ_i to buyer i, of the form $s_{i,j} = \theta_i + \eta_{i,j}$, where $\eta_{i,j} \sim \mathcal{D}_{i,j}$

Value of Data Bundles

- n *buyers*. Buyer i has budget b_i, and an unknown state θ_i to predict,
- m *sellers*. Seller j owns dataset D_j with s_j data records,
- Each data record of seller j, is a noisy signal about θ_i to buyer i, of the form $s_{i,j} = \theta_i + \eta_{i,j}$, where $\eta_{i,j} \sim \mathcal{D}_{i,j}$
- Buyer i's *data bundle* $\boldsymbol{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,m})$, where $x_{i,j}$ is the amount of data records of seller j.

Value of Data Bundles

- n *buyers*. Buyer i has budget b_i, and an unknown state θ_i to predict,
- m *sellers*. Seller j owns dataset D_j with s_j data records,
- Each data record of seller j, is a noisy signal about θ_i to buyer i, of the form $s_{i,j} = \theta_i + \eta_{i,j}$, where $\eta_{i,j} \sim \mathcal{D}_{i,j}$
- $S(\boldsymbol{x}_i)$ = set of all observed signals from data records in $\boldsymbol{x}_i$,

$$u_i(\boldsymbol{x}_i) = \mathrm{Var}(\theta) - \mathbb{E}[\mathrm{Var}(\theta \mid S(\boldsymbol{x}_i)] = \mathrm{Var}(\mathbb{E}[\theta \mid S(\boldsymbol{x}_i)])$$

Value of Data Bundles

Assumptions

- $\theta_i \sim \mathcal{N}(\mu_i, \tau_i^{-1})$

Value of Data Bundles

Assumptions

- $\theta_i \sim \mathcal{N}(\mu_i, \tau_i^{-1})$
- $\eta_{i,j} \sim \mathcal{N}(0, \tau_{i,j}^{-1})$,

Value of Data Bundles

Assumptions

- $\theta_i \sim \mathcal{N}(\mu_i, \tau_i^{-1})$
- $\eta_{i,j} \sim \mathcal{N}(0, \tau_{i,j}^{-1})$, implying $\eta_{i,j} \perp \eta_{i,j'}$ and $\eta_{i,j} \perp \theta_i$

Value of Data Bundles

Assumptions

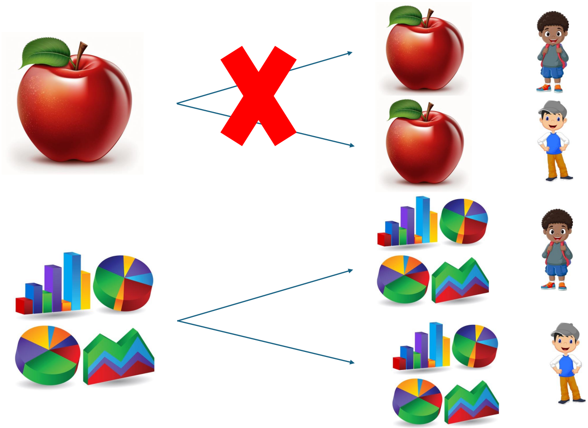
- $\theta_i \sim \mathcal{N}(\mu_i, \tau_i^{-1})$
- $\eta_{i,j} \sim \mathcal{N}(0, \tau_{i,j}^{-1})$, implying $\eta_{i,j} \perp \eta_{i,j'}$ and $\eta_{i,j} \perp \theta_i$

Implication

- $u_i(\boldsymbol{x}_i) = \tau_i{}^{-1} - (\tau_i + \sum_j \tau_{i,j} x_{i,j})^{-1}$
- $\max_{\boldsymbol{x} \in P} u_i(\boldsymbol{x}_i) \iff \max_{\boldsymbol{x} \in P} \sum_j \tau_{i,j} x_{i,j}$.
- Alternatively, define, $u_i(\boldsymbol{x}_i) = \mathbb{E}[\mathrm{Pre}(\theta_i \mid S(\boldsymbol{x}_i))] - \mathrm{Pre}(\theta_i) \iff u_i(\boldsymbol{x}_i) = \sum_j \tau_{i,j} x_{i,j}$.

Data Markets

The Non-Rivalry of Data



Perfect Competition

- $\sum_i x^*_{i,j}(\boldsymbol{p}) = s_j$ does not make sense! (data is non-rival),

Perfect Competition

- $\sum_i x^*_{i,j}(\boldsymbol{p}) = s_j$ does not make sense! (data is non-rival),
- Principle of CE: "*Everything produced is consumed*".

Perfect Competition

- $\sum_i x^*_{i,j}(\boldsymbol{p}) = s_j$ does not make sense! (data is non-rival),
- Principle of CE: "*Everything produced is consumed*".
- New market clearing: $\max_i x^*_{i,j}(\boldsymbol{p}) = s_j \ \ \forall j$

Perfect Competition

- $\sum_i x^*_{i,j}(\boldsymbol{p}) = s_j$ does not make sense! (data is non-rival),
- Principle of CE: "*Everything produced is consumed*".
- New market clearing: $\max_i x^*_{i,j}(\boldsymbol{p}) = s_j \quad \forall j$

CE in Data Markets

$(\boldsymbol{p}, \boldsymbol{x})$ is CE iff

- $\boldsymbol{x}^*_i \in \arg\max_{\boldsymbol{y} | \boldsymbol{p}^T \boldsymbol{y} \,\leq\, b_i} u_i(\boldsymbol{y}) \quad \forall i$, and
- $\max_i x^*_{i,j} = s_j \quad \forall j$

CE: Rivalrous vs.Data Markets

Rivalrous Markets

- CE exists and CE price is unique, rational

Data Markets (Our Results)

CE: Rivalrous vs.Data Markets

Rivalrous Markets

- CE exists and CE price is unique, rational

Data Markets (Our Results)

- CE exists and CE price is unique, rational

CE: Rivalrous vs.Data Markets

Rivalrous Markets

- CE exists and CE price is unique, rational
- Set of CE allocations is convex

Data Markets (Our Results)

- CE exists and CE price is unique, rational

CE: Rivalrous vs.Data Markets

Rivalrous Markets

- CE exists and CE price is unique, rational
- Set of CE allocations is convex

Data Markets (Our Results)

- CE exists and CE price is unique, rational
- Set of CE allocations is **non-convex**

CE: Rivalrous vs.Data Markets

Rivalrous Markets

- CE exists and CE price is unique, rational
- Set of CE allocations is convex
- Convex program known (Eisenberg–Gale)

Data Markets (Our Results)

- CE exists and CE price is unique, rational
- Set of CE allocations is **non-convex**

CE: Rivalrous vs.Data Markets

Rivalrous Markets

- CE exists and CE price is unique, rational
- Set of CE allocations is convex
- Convex program known (Eisenberg–Gale)

Data Markets (Our Results)

- CE exists and CE price is unique, rational
- Set of CE allocations is **non-convex**
- Convex Program not known

CE: Rivalrous vs.Data Markets

Theorem [**C.**, Garg, Murhekar, Song]

An ε-CE can be computed through an auction-style algorithm in poly$(n, m, 1/\varepsilon, \max_{i,j} \log(\tau_{i,j}))$.

CE: Rivalrous vs.Data Markets

Theorem [**C.**, Garg, Murhekar, Song]

An ε-CE can be computed through an auction-style algorithm in poly$(n, m, 1/\varepsilon, \max_{i,j} \log(\tau_{i,j}))$.

Open Problems

- What is the complexity of finding an exact CE in data markets?
- Are there market dynamics that converge to a CE in data markets?

Oligopoly (Pricing Game)

Oligopoly (Pricing Game)

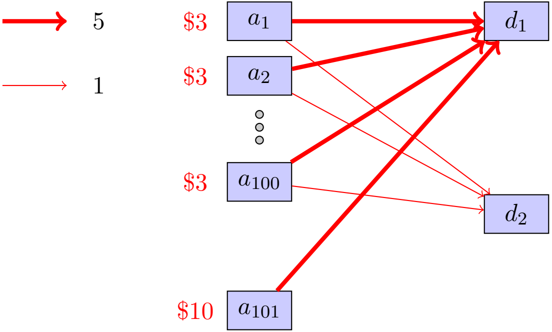
CE $\not\Rightarrow$ NE in Oligopolistic Data Markets.

Oligopoly (Pricing Game) is unstable

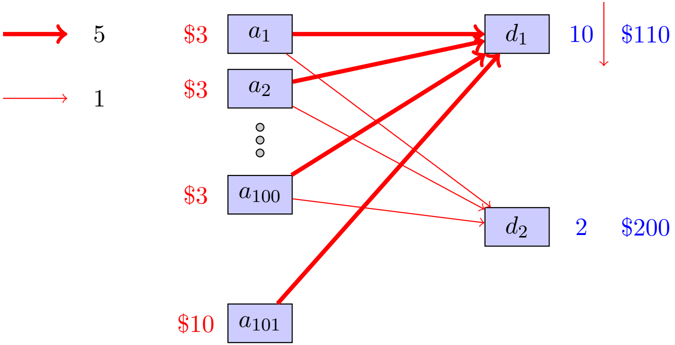
CE $\not\Rightarrow$ NE in Oligopolistic Data Markets.

Recall: CE $\implies$ NE in rivalrous markets. So NE exists and is computable in poly-time.

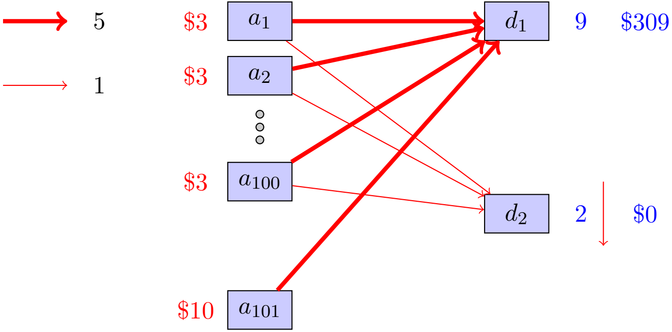
Oligopoly (Pricing Game) is unstable



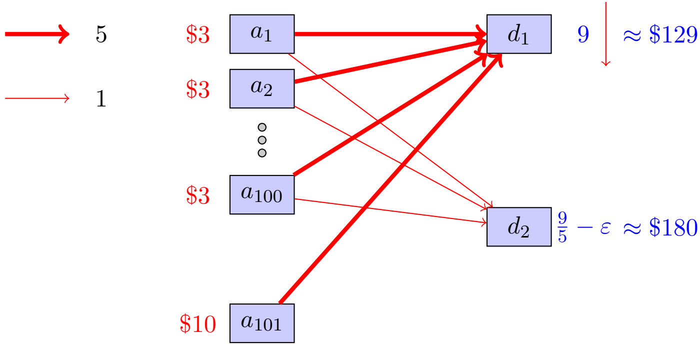
Oligopoly (Pricing Game) is unstable



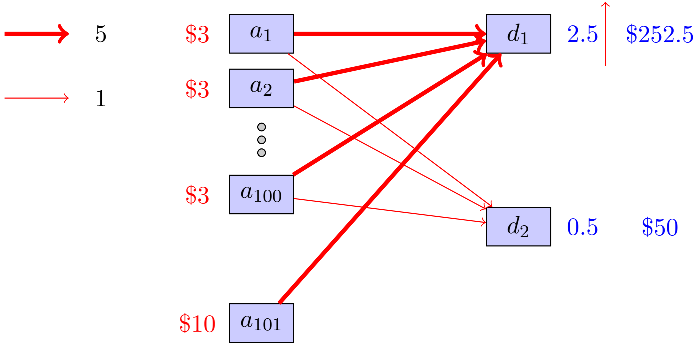
Oligopoly (Pricing Game) is unstable



Oligopoly (Pricing Game) is unstable



Oligopoly (Pricing Game) is unstable



Monopoly (Revenue Maximization)

Recall: Agents have value for money,

$$u_i(\boldsymbol{x}, \boldsymbol{p}) = \alpha_i(\sum_j \tau_{i,j} x_{i,j}) - \sum_j p_j x_{i,j}$$

and the goal is to find $\boldsymbol{p}$, such that

$$\max_{\boldsymbol{p}, \boldsymbol{x} \in OPT_i(\boldsymbol{p})} \sum_{i,j} p_j x_{i,j}$$

where $OPT_i(\boldsymbol{p}) = \arg\max_{\{\boldsymbol{y} | \boldsymbol{p}^T \boldsymbol{y} \,\leq\, b_i\}} u_i(\boldsymbol{x}, \boldsymbol{p})$

Revenue Maximization

Theorem **[C.**, Garg, Sharma, Song]

Revenue maximization in data markets is APX-hard.

Recall: CE $\Longrightarrow$ revenue maximization (SE) in rivalrous markets. Computing SE is therefore in P.

Revenue Maximization

Theorem **[C.**, Garg, Sharma, Song]

Revenue maximization in data markets is APX-hard.

Recall: CE $\implies$ revenue maximization (SE) in rivalrous markets. Computing SE is therefore in P.

Theorem **[C.**, Garg, Sharma, Song]

There exists an *online* 2-approximation algorithm for maximizing revenue in data markets.

Connection to k-submodularity

Monotone k-Submodularity

Given a ground set U, a function f defined on k-tuple disjoint subsets of U is k-submodular iff

- $f(S_1, \ldots, S_k) \leq f(T_1, \ldots, T_k)$ if all $S_i \subseteq T_i$, and
- $f(S_1, \ldots, S_r \cup \{g\}, \ldots S_k) - f(S_1, \ldots, S_k) \geq$ $f(T_1, \ldots, T_r \cup \{g\}, \ldots T_k) - f(T_1, \ldots, T_k)$, where $S_i \subseteq T_i$.

Connection to k-submodularity

Monotone k-Submodularity

Given a ground set U, a function f defined on k-tuple disjoint subsets of U is k-submodular iff

- $f(S_1, \ldots, S_k) \leq f(T_1, \ldots, T_k)$ if all $S_i \subseteq T_i$, and
- $f(S_1, \ldots, S_r \cup \{g\}, \ldots S_k) - f(S_1, \ldots, S_k) \geq f(T_1, \ldots, T_r \cup \{g\}, \ldots T_k) - f(T_1, \ldots, T_k)$, where $S_i \subseteq T_i$.

Ward and Zivny SODA'14

There exists an online 2-approximation greedy algorithm for maximizing k submodular functions.

Connection to k-submodularity

Key Observations

- $u_i(\boldsymbol{x}, \boldsymbol{p}) = \sum_j (\alpha_i \tau_{i,j} - p_j) x_{i,j}$,

Connection to k-submodularity

Key Observations

- $u_i(\boldsymbol{x}, \boldsymbol{p}) = \sum_j (\alpha_i \tau_{i,j} - p_j) x_{i,j}$,
- Money spend by buyer i is $\min(b_i, \sum_{j \mid \alpha_i \tau_{i,j} \geq p_j} p_j)$,

Connection to k-submodularity

Key Observations

- $u_i(\boldsymbol{x}, \boldsymbol{p}) = \sum_j (\alpha_i \tau_{i,j} - p_j) x_{i,j}$,
- Money spend by buyer i is $\min(b_i, \sum_{j \mid \alpha_i \tau_{i,j} \geq p_j} p_j)$,
- At an optimal $\boldsymbol{p}$, $p_j = \alpha_i \tau_{i,j}$ for some i,

Connection to k-submodularity

Key Observations

- $u_i(\boldsymbol{x}, \boldsymbol{p}) = \sum_j (\alpha_i \tau_{i,j} - p_j) x_{i,j}$,
- Money spend by buyer i is $\min(b_i, \sum_{j | \alpha_i \tau_{i,j} \geq p_j} p_j)$,
- At an optimal $\boldsymbol{p}$, $p_j = \alpha_i \tau_{i,j}$ for some i,
- Let $S_i = \{j \mid p_j = \alpha_i \tau_{i,j}\}$,

Connection to k-submodularity

Key Observations

- $u_i(\boldsymbol{x}, \boldsymbol{p}) = \sum_j (\alpha_i \tau_{i,j} - p_j) x_{i,j}$,
- Money spend by buyer i is $\min(b_i, \sum_{j | \alpha_i \tau_{i,j} \geq p_j} p_j)$,
- At an optimal $\boldsymbol{p}$, $p_j = \alpha_i \tau_{i,j}$ for some i,
- Let $S_i = \{j \mid p_j = \alpha_i \tau_{i,j}\}$, and $f(S_1, S_2, \ldots, S_n) =$ total revenue when prices are defined by $S_1, \ldots, S_n$. Prices of datasets $\notin \cup_i S_i$ is set to ∞.

Connection to k-submodularity

Key Observations

- $u_i(\boldsymbol{x}, \boldsymbol{p}) = \sum_j (\alpha_i \tau_{i,j} - p_j) x_{i,j}$,
- Money spend by buyer i is $\min(b_i, \sum_{j \mid \alpha_i \tau_{i,j} \geq p_j} p_j)$,
- At an optimal $\boldsymbol{p}$, $p_j = \alpha_i \tau_{i,j}$ for some i,
- Let $S_i = \{j \mid p_j = \alpha_i \tau_{i,j}\}$, and $f(S_1, S_2, \ldots, S_n) =$ total revenue when prices are defined by $S_1, \ldots, S_n$. Prices of datasets $\notin \cup_i S_i$ is set to ∞.
- **$f(S_1, \ldots, S_n)$ is n-submodular**

Connection to k-submodularity

$f(S_1, \dots, S_n)$ is n-submodular

- $\boldsymbol{p}(S_1, \dots, S_n)$, be the prices corresponding to $S_1, \dots, S_n$.
- Let $S_1 \subseteq T_1, \dots, S_n \subseteq T_n$,
- Buyers have less remaining budget in $\boldsymbol{p}(T_1, \dots, T_n)$ than in $\boldsymbol{p}(S_1, \dots, S_n)$,
- Marginal revenue increase in pricing a new good is more in $p(S_1, \dots, S_n)$ than in $\boldsymbol{p}(T_1, \dots, T_n)$.

Connection to k-submodularity

Theorem [**C.**, Garg, Sharma, Song]

One can formulate revenue maximization in data markets as submodular optimization subject to partition matroid constraint and get a $(1 - 1/e)^{-1}$-approximation.

Beyond Uniform Pricing

Core-Question

- Why restrict ourselves to pricing functions that are linear?
- What happens when agents (monopolist or oligopolist) optimize over all pricing functions?

Beyond Uniform Pricing

Core-Question

- Why restrict ourselves to pricing functions that are linear?
- What happens when agents (monopolist or oligopolist) optimize over all pricing functions?

Theorem [**C.**, Garg, Sharma, Song]

When agents optimize over all pricing strategies,

- A SE can computed in polynomial time.
- An *approximate* NE exists in the pricing game.

Summary

- Framework for studying *equilibrium theory* in data markets.

Summary

- Framework for studying *equilibrium theory* in data markets.
- *Rivalry* of data calls for a new CE notion.

Summary

- Framework for studying *equilibrium theory* in data markets.
- *Rivalry* of data calls for a new CE notion.
- *Uniform pricing*: oligopolies unstable, monopolist's strategy hard.

Summary

- Framework for studying *equilibrium theory* in data markets.
- *Rivalry* of data calls for a new CE notion.
- *Uniform pricing*: oligopolies unstable, monopolist's strategy hard.
- *Beyond-uniform pricing* fixes existence and computation issues.

Summary

- Framework for studying *equilibrium theory* in data markets.
- *Rivalry* of data calls for a new CE notion.
- *Uniform pricing*: oligopolies unstable, monopolist's strategy hard.
- *Beyond-uniform pricing* fixes existence and computation issues.
- Computational directions: *complexity, hardness, and dynamics*.

Summary

- Framework for studying *equilibrium theory* in data markets.
- *Rivalry* of data calls for a new CE notion.
- *Uniform pricing*: oligopolies unstable, monopolist's strategy hard.
- *Beyond-uniform pricing* fixes existence and computation issues.
- Computational directions: *complexity, hardness, and dynamics*.
- Modeling directions: *complimentary signals, Cournot Oligopoly.*

Thank You!