# Too Much Data: Prices and Inefficiencies in Data Markets

Azarakhsh Malekian
University of Toronto

with Daron Acemoglu, Ali Makhdoumi, and Asuman Ozdaglar

Focus Period Symposium on Network Dynamics and Control
September, 2023

# Motivation

- Our lives have been vastly improved via technology advancements.
- We regularly use these products usually for “free” to address our needs.
- These applications and products are founded and funded by data:
  - They are being **improved by the data** gathered from the users.
  - Their revenue is generated from data either via the **advertisers** or via **explicitly selling** data.

# Motivation

- Our lives have been vastly improved via technology advancements.
- We regularly use these products usually for “free” to address our needs.
- These applications and products are founded and funded by data:
  - They are being **improved by the data** gathered from the users.
  - Their revenue is generated from data either via the **advertisers** or via **explicitly selling** data.

# Motivation

- Our lives have been vastly improved via technology advancements.
- We regularly use these products usually for “free” to address our needs.
- These applications and products are founded and funded by data:
  - They are being **improved by the data** gathered from the users.
  - Their revenue is generated from data either via the **advertisers** or via **explicitly selling** data.

# Motivation

- Our lives have been vastly improved via technology advancements.
- We regularly use these products usually for “free” to address our needs.
- These applications and products are founded and funded by data:
  - They are being **improved by the data** gathered from the users.
  - Their revenue is generated from data either via the **advertisers** or via **explicitly selling** data.

# Motivation

- The heavy usage of data and recent data scandals raised public awareness and consequently privacy concerns.
- Recent scandals such as Cambridge Analytica data scandal raised a large number of policy and regularity questions regarding the protection and sharing of information.
- The main question is "what are the appropriate business models for data market"?

# Motivation

- The heavy usage of data and recent data scandals raised public awareness and consequently privacy concerns.
- Recent scandals such as Cambridge Analytica data scandal raised a large number of policy and regularity questions regarding the protection and sharing of information.
- The main question is "what are the appropriate business models for data market"?

# Motivation

- The heavy usage of data and recent data scandals raised public awareness and consequently privacy concerns.
- Recent scandals such as Cambridge Analytica data scandal raised a large number of policy and regularity questions regarding the protection and sharing of information.
- The main question is “what are the appropriate business models for data market”?

# Privacy Market Properties

## Main Features:

- There is no unified definition for privacy.
- People are not clearly aware of how and by whom their data will be used upon taking different actions.
- Because of the correlations among individuals' personal data, when an individual shares her/his information it partially reveals others' information.

## Main Questions

- How does these correlations affect the equilibrium price of personal data?
- Are there equilibria that benefit both users (data holders) and platforms (data buyers)?
- What are the implications of a market for personal data for individuals and society as a whole?
- How can we improve the surplus:
  - treating privacy as another economic good, or
  - based on regulation, treating privacy as a fundamental right?

# Privacy Market Properties

## Main Features:

- There is no unified definition for privacy.
- People are not clearly aware of how and by whom their data will be used upon taking different actions.
- Because of the correlations among individuals' personal data, when an individual shares her/his information it partially reveals others' information.

## Main Questions

- How does these correlations affect the equilibrium price of personal data?
- Are there equilibria that benefit both users (data holders) and platforms (data buyers)?
- What are the implications of a market for personal data for individuals and society as a whole?
- How can we improve the surplus:
    - treating privacy as another economic good, or
    - based on regulation, treating privacy as a fundamental right?

# In This Talk

- We develop a model in which data sharing by one user reveals relevant data about others, and a platform wants to purchase users' data to infer their types.
- We first establish basic properties of our information measure, and then study the equilibrium existence of this game.
- We characterize data market equilibria and their efficiency properties, and provide conditions under which equilibria are inefficient and shutting down data markets improves welfare.
    - We consider also the generalizations: competition, unknown valuations, and correlations.
- Finally, we study schemes to regulate the market and improve its efficiency.
    - With the complete knowledge of the correlations and users' valuation, Pigovian taxation resolves the inefficiency.
    - We then show a solution based on decorrelation reduces inefficiencies.

# Related Literature

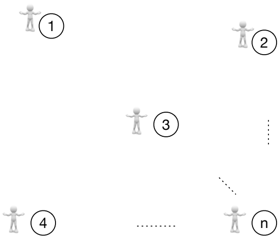
- We are related to two literatures:
  1. Privacy: [Warren and Brandeis 1890], [Westin 1968], [Posner 1981], [Varian 2009], [Goldfarb and Tucker 2012] , [Acquisti and Taylor 2016].
  2. Information markets: [Admati and Pfleiderer 1986], [Taylor 2004] , [Bergemann and Bonatti 2015], [Horner and Skrzypacz 2016], [Bergemann et al. 2018].
- Most closely related are:
  - Early papers on externalities and data sharing: [MacCarthy 2010] and [Fairfield and Engel 2015].
  - More recent work on data externalities and information markets by [Choi et al. 2019], [Bergemann et al. 2021], and [Ichihashi 2020 & 2021]

# Model: Information

- $n$ users (data holders) residing on a platform:
  - Each user has a personal type: $X_i \sim \mathcal{N}(0, \sigma_i^2)$.
  - Users' types are correlated, captured by matrix $\Sigma$, i.e.,

$$X = (X_1, \ldots, X_n) \sim \mathcal{N}(0, \Sigma)$$

- Platform wants to learn $X_1, \ldots, X_n$:
  - Offers to user $i$, price $p_i$, in exchange for her personal data, $S_i = X_i + Z_i$ where $Z_i \sim \mathcal{N}(0, 1)$.

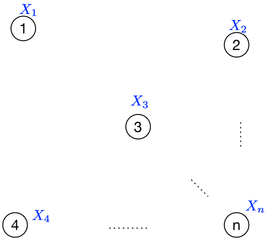


# Model: Information

- $n$ users (data holders) residing on a platform:
  - Each user has a personal type: $X_i \sim \mathcal{N}(0, \sigma_i^2)$.
  - Users' types are correlated, captured by matrix $\Sigma$, i.e.,

$$X = (X_1, \ldots, X_n) \sim \mathcal{N}(0, \Sigma)$$

- Platform wants to learn $X_1, \ldots, X_n$:
  - Offers to user $i$, price $p_i$, in exchange for her personal data, $S_i = X_i + Z_i$ where $Z_i \sim \mathcal{N}(0, 1)$.

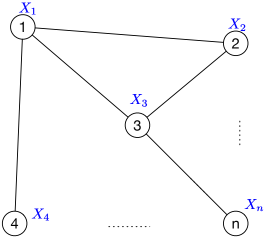


# Model: Information

- $n$ users (data holders) residing on a platform:
  - Each user has a personal type: $X_i \sim \mathcal{N}(0, \sigma_i^2)$.
  - Users' types are correlated, captured by matrix $\Sigma$, i.e.,

$$\mathsf{X} = (X_1, \ldots, X_n) \sim \mathcal{N}(0, \Sigma)$$

- Platform wants to learn $X_1, \ldots, X_n$:
  - Offers to user $i$, price $p_i$, in exchange for her personal data, $S_i = X_i + Z_i$ where $Z_i \sim \mathcal{N}(0, 1)$.

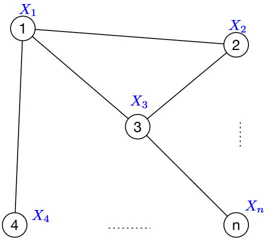


# Model: Information

- $n$ users (data holders) residing on a platform:
  - Each user has a personal type: $X_i \sim \mathcal{N}(0, \sigma_i^2)$.
  - Users' types are correlated, captured by matrix $\Sigma$, i.e.,

$$\mathrm{X} = (X_1, \dots, X_n) \sim \mathcal{N}(0, \Sigma)$$

- Platform wants to learn $X_1, \dots, X_n$:
  - Offers to user $i$, price $p_i$, in exchange for her personal data, $S_i = X_i + Z_i$ where $Z_i \sim \mathcal{N}(0, 1)$.

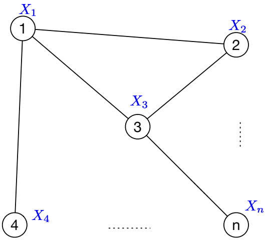


# Model: Information

- $n$ users (data holders) residing on a platform:
  - Each user has a personal type: $X_i \sim \mathcal{N}(0, \sigma_i^2)$.
  - Users' types are correlated, captured by matrix $\Sigma$, i.e.,

$$\mathrm{X} = (X_1, \ldots, X_n) \sim \mathcal{N}(0, \Sigma)$$

- Platform wants to learn $X_1, \ldots, X_n$:
  - Offers to user $i$, price $p_i$, in exchange for her personal data, $S_i = X_i + Z_i$ where $Z_i \sim \mathcal{N}(0, 1)$.



# Model: Leaked Information

- Given price vector $\mathrm{p} = (p_1, \ldots, p_n)$, a set of users decide to share.
- $a_i \in \{0, 1\}$: user $i$'s decision to whether to share her personal data.
- $\mathrm{a} = (a_1, \ldots, a_n)$: users' decision
- $\mathrm{S}_\mathrm{a} := (S_i \ : \ i \in \mathcal{V} \text{ s.t. } a_i = 1)$ is platform's data.

Definition (Leaked information)

*Leaked information* of (or about) user $i \in \mathcal{V}$ is the reduction in the MSE of the best estimator of the type of user $i$:

$$\mathcal{I}_i(\mathrm{a}) = \sigma_i^2 - \min_{\hat{x}_i} \mathbb{E}\left[ \left( X_i - \hat{x}_i\left( \mathrm{S}_\mathrm{a} \right) \right)^2 \right].$$

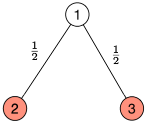
# Model: Leaked Information

- Given price vector $\mathrm{p} = (p_1, \ldots, p_n)$, a set of users decide to share.
- $a_i \in \{0, 1\}$: user $i$'s decision to whether to share her personal data.
- $\mathrm{a} = (a_1, \ldots, a_n)$: users' decision
- $\mathrm{S_a}: = (S_i \ : \ i \in \mathcal{V} \text{ s.t. } a_i = 1)$ is platform's data.

## Definition (Leaked information)

*Leaked information* of (or about) user $i \in \mathcal{V}$ is the reduction in the MSE of the best estimator of the type of user $i$:

$$\mathcal{I}_i(\mathrm{a}) = \sigma_i^2 - \min_{\hat{x}_i} \mathbb{E}\left[\left(X_i - \hat{x}_i\left(\mathrm{S_a}\right)\right)^2\right].$$



$\mathcal{I}_1(a_1 = 0, a_2 = 1, a_3 = 1) = \frac{3}{7}$

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix}$$

# Model: Payoff

- Payoff of Platform:

$$U(\mathrm{a},\mathrm{p}) = \sum_{i\in\mathcal{V}} \mathcal{I}_i(\mathrm{a}) - \sum_{i\in\mathcal{V}:\ a_i=1} p_i.$$

    - $p_i$: denotes payments to user $i$ from the platform (direct payment or service).

- Payoff of user $i$:

$$u_i(a_i,\mathrm{a}_{-i},\mathrm{p}) = \begin{cases} p_i - v_i\mathcal{I}_i\left(a_i=1,\mathrm{a}_{-i}\right), & a_i = 1 \\ -v_i\mathcal{I}_i\left(a_i=0,\mathrm{a}_{-i}\right), & a_i = 0, \end{cases}$$

    - $v_i \geq 0$: user $i$'s value of privacy.

- Utilitarian welfare = social surplus is

$$\text{Social surplus}(\mathrm{a}) = \sum_{i\in\mathcal{V}} (1-v_i)\mathcal{I}_i(\mathrm{a}).$$

# Equilibrium

### Definition (User equilibrium)

For any price vector $\mathrm{p}$, action profile $\mathrm{a} \in \{0,1\}^n$ is a *user equilibrium* if

$$a_i \in \operatorname*{argmax}_{a\in\{0,1\}} u_i(a, \mathrm{a}_{-i}).$$

We let $\mathcal{A}(\mathrm{p})$ be the set of all user equilibria.

- Properties of Information leakage:
  - **Submodularity**: as more users share their information, the marginal increase of information leakage decreases.
  - **Monotonicity**: the information leakage increases as more people share their information.

# Equilibrium

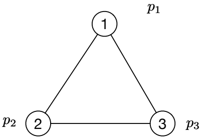
**Definition (User equilibrium)**

For any price vector p, action profile $\mathrm{a} \in \{0,1\}^n$ is a *user equilibrium* if

$$a_i \in \operatorname*{argmax}_{a \in \{0,1\}} u_i(a, \mathrm{a}_{-i}).$$

We let $\mathcal{A}(\mathrm{p})$ be the set of all user equilibria.

- Properties of Information leakage:
  - **Submodularity**: as more users share their information, the marginal increase of information leakage decreases.
  - **Monotonicity**: the information leakage increases as more people share their information.



# Equilibrium

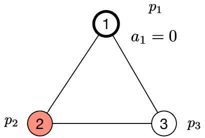
**Definition (User equilibrium)**

For any price vector p, action profile a $\in \{0,1\}^n$ is a *user equilibrium* if

$$a_i \in \operatorname*{argmax}_{a \in \{0,1\}} u_i(a, \mathrm{a}_{-i}).$$

We let $\mathcal{A}(\mathrm{p})$ be the set of all user equilibria.

- Properties of Information leakage:
  - **Submodularity**: as more users share their information, the marginal increase of information leakage decreases.
  - **Monotonicity**: the information leakage increases as more people share their information.



# Equilibrium

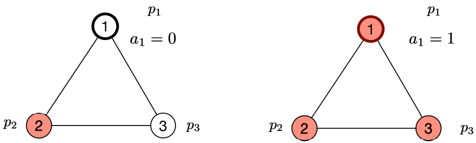
### Definition (User equilibrium)

For any price vector p, action profile a $\in \{0,1\}^n$ is a *user equilibrium* if

$$a_i \in \operatorname*{argmax}_{a \in \{0,1\}} u_i(a, \mathrm{a}_{-i}).$$

We let $\mathcal{A}(\mathrm{p})$ be the set of all user equilibria.

- Properties of Information leakage:
  - **Submodularity**: as more users share their information, the marginal increase of information leakage decreases.
  - **Monotonicity**: the information leakage increases as more people share their information.



# Equilibrium

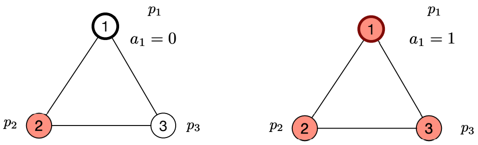
### Definition (User equilibrium)

For any price vector $\mathrm{p}$, action profile $\mathrm{a} \in \{0,1\}^n$ is a *user equilibrium* if

$$a_i \in \operatorname*{argmax}_{a\in\{0,1\}} u_i(a, \mathrm{a}_{-i}).$$

We let $\mathcal{A}(\mathrm{p})$ be the set of all user equilibria.

- Properties of Information leakage:
  - **Submodularity**: as more users share their information, the marginal increase of information leakage decreases.
  - **Monotonicity**: the information leakage increases as more people share their information.



# Existence of Equilibrium

**Proposition (Existence)**

*For any* p*, the set* $\mathcal{A}(\mathrm{p})$ *is a complete lattice. Therefore,* $\mathcal{A}(\mathrm{p})$ *is non-empty and it has largest and least elements.*

- *Proof Idea*: leaked information functions $\mathcal{I}_i(\cdot)$ are *submodular* ⇒ game is supermodular ⇒ Tarski's fixed point theorem shows $\mathcal{A}(\mathrm{p})$ is a complete lattice.

**Definition (Stackelberg equilibrium)**

Action profile $\mathrm{a}^{\mathrm{E}}$ and price vector $\mathrm{p}^{\mathrm{E}}$ is a Stackelberg equilibrium if $\mathrm{a}^{\mathrm{E}} \in \mathcal{A}(\mathrm{p}^{\mathrm{E}})$, and

$$U(\mathrm{a}^{E}, \mathrm{p}^{E}) \geq U(\mathrm{a}, \mathrm{p}), \qquad \forall \mathrm{p}, \forall \mathrm{a} \in \mathcal{A}(\mathrm{p}).$$

**Proposition (Existence)**

*For any* $\Sigma$ *and* v *Stackelberg equilibrium exists.*

# Existence of Equilibrium

**Proposition (Existence)**

*For any* p, *the set* $\mathcal{A}(\mathrm{p})$ *is a complete lattice. Therefore,* $\mathcal{A}(\mathrm{p})$ *is non-empty and it has largest and least elements.*

- *Proof Idea*: leaked information functions $\mathcal{I}_i(\cdot)$ are *submodular* ⇒ game is supermodular ⇒ Tarski's fixed point theorem shows $\mathcal{A}(\mathrm{p})$ is a complete lattice.

**Definition (Stackelberg equilibrium)**

Action profile $\mathrm{a}^{\mathrm{E}}$ and price vector $\mathrm{p}^{\mathrm{E}}$ is a Stackelberg equilibrium if $\mathrm{a}^{\mathrm{E}} \in \mathcal{A}(\mathrm{p}^{\mathrm{E}})$, and

$$U(\mathrm{a}^E, \mathrm{p}^E) \geq U(\mathrm{a}, \mathrm{p}), \qquad \forall \mathrm{p}, \forall \mathrm{a} \in \mathcal{A}(\mathrm{p}).$$

**Proposition (Existence)**

*For any* $\Sigma$ *and* v *Stackelberg equilibrium exists.*

# Existence of Equilibrium

### Proposition (Existence)

*For any* $\mathrm{p}$, *the set* $\mathcal{A}(\mathrm{p})$ *is a complete lattice. Therefore,* $\mathcal{A}(\mathrm{p})$ *is non-empty and it has largest and least elements.*

- *Proof Idea*: leaked information functions $\mathcal{I}_i(\cdot)$ are *submodular* ⇒ game is supermodular ⇒ Tarski's fixed point theorem shows $\mathcal{A}(\mathrm{p})$ is a complete lattice.

### Definition (Stackelberg equilibrium)

Action profile $\mathrm{a}^{\mathrm{E}}$ and price vector $\mathrm{p}^{\mathrm{E}}$ is a Stackelberg equilibrium if $\mathrm{a}^{\mathrm{E}} \in \mathcal{A}(\mathrm{p}^{\mathrm{E}})$, and

$$U(\mathrm{a}^{E}, \mathrm{p}^{E}) \geq U(\mathrm{a}, \mathrm{p}), \qquad \forall \mathrm{p}, \forall \mathrm{a} \in \mathcal{A}(\mathrm{p}).$$

### Proposition (Existence)

*For any* $\Sigma$ *and* $\mathrm{v}$ *Stackelberg equilibrium exists.*

# Existence of Equilibrium

### Proposition (Existence)

*For any* $\mathrm{p}$*, the set* $\mathcal{A}(\mathrm{p})$ *is a complete lattice. Therefore,* $\mathcal{A}(\mathrm{p})$ *is non-empty and it has largest and least elements.*

- *Proof Idea*: leaked information functions $\mathcal{I}_i(\cdot)$ are *submodular* ⇒ game is supermodular ⇒ Tarski's fixed point theorem shows $\mathcal{A}(\mathrm{p})$ is a complete lattice.

### Definition (Stackelberg equilibrium)

Action profile $\mathrm{a}^{\mathrm{E}}$ and price vector $\mathrm{p}^{\mathrm{E}}$ is a Stackelberg equilibrium if $\mathrm{a}^{\mathrm{E}} \in \mathcal{A}(\mathrm{p}^{\mathrm{E}})$, and

$$U(\mathrm{a}^E, \mathrm{p}^E) \geq U(\mathrm{a}, \mathrm{p}), \qquad \forall \mathrm{p}, \forall \mathrm{a} \in \mathcal{A}(\mathrm{p}).$$

### Proposition (Existence)

*For any* $\Sigma$ *and* $\mathrm{v}$ *Stackelberg equilibrium exists.*

# Example: User Equilibria

**Is the total payment monotone in the set of users who share information?**

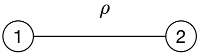
- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$
- Users' equilibria as a function of price vector:

- $\Rightarrow$ for $\rho^2 \geq \frac{7-\sqrt{17}}{4} \approx .71$, the buyer can extract more information ($a_1 = a_2 = 1$ instead of $a_1 = 1$, $a_2 = 0$) with lower overall payment.

# Example: User Equilibria

**Is the total payment monotone in the set of users who share information?**

- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$



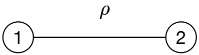
- Users' equilibria as a function of price vector:

- $\Rightarrow$ for $\rho^2 \geq \frac{7-\sqrt{17}}{4} \approx .71$, the buyer can extract more information ($a_1 = a_2 = 1$ instead of $a_1 = 1,\ a_2 = 0$) with lower overall payment.

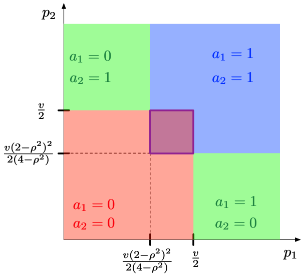
# Example: User Equilibria

**Is the total payment monotone in the set of users who share information?**

- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$



- Users' equilibria as a function of price vector:

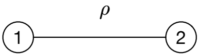


- $\Rightarrow$ for $\rho^2 \geq \frac{7-\sqrt{17}}{4} \approx .71$, the buyer can extract more information ($a_1 = a_2 = 1$ instead of $a_1 = 1$, $a_2 = 0$) with lower overall payment.

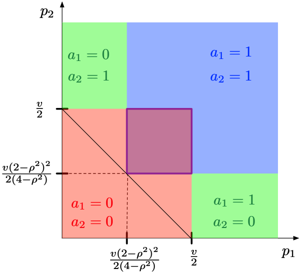
# Example: User Equilibria

**Is the total payment monotone in the set of users who share information?**

- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix}$



- Users' equilibria as a function of price vector:

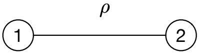


- $\Rightarrow$ for $\rho^2 \geq \frac{7-\sqrt{17}}{4} \approx .71$, the buyer can extract more information ($a_1 = a_2 = 1$ instead of $a_1 = 1$, $a_2 = 0$) with lower overall payment.

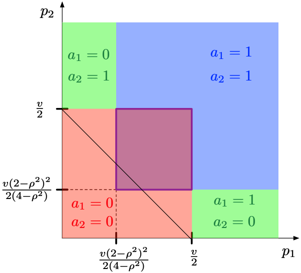
# Example: User Equilibria

**Is the total payment monotone in the set of users who share information?**

- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$



- Users' equilibria as a function of price vector:



- $\Rightarrow$ for $\rho^2 \geq \frac{7-\sqrt{17}}{4} \approx .71$, the buyer can extract more information ($a_1 = a_2 = 1$ instead of $a_1 = 1$, $a_2 = 0$) with lower overall payment.

# Example: Equilibrium

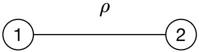
**Is surplus monotone in valuations?**

- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$
- (Stackelberg) Equilibrium is
- For intermediate values of $v$, i.e., $v \in [1, \frac{4}{(2-\rho^2)^2}]$, total surplus is negative.

# Example: Equilibrium

**Is surplus monotone in valuations?**

- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$



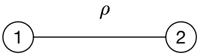
- (Stackelberg) Equilibrium is

- For intermediate values of $v$, i.e., $v \in [1, \frac{4}{(2-\rho^2)^2}]$, total surplus is negative.

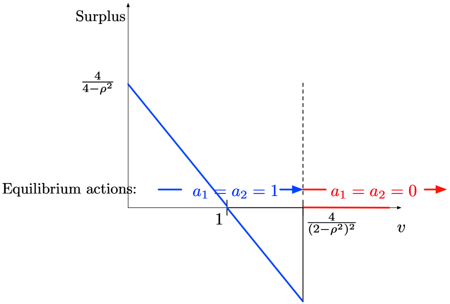
# Example: Equilibrium

**Is surplus monotone in valuations?**

- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$



- (Stackelberg) Equilibrium is

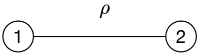


- For intermediate values of $v$, i.e., $v \in [1, \frac{4}{(2-\rho^2)^2}]$, total surplus is negative.

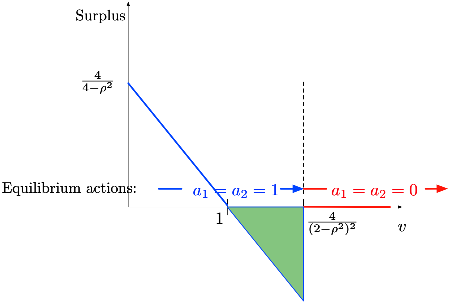
# Example: Equilibrium

**Is surplus monotone in valuations?**

- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$



- (Stackelberg) Equilibrium is

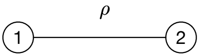


- For intermediate values of $v$, i.e., $v \in [1, \frac{4}{(2-\rho^2)^2}]$, total surplus is negative.

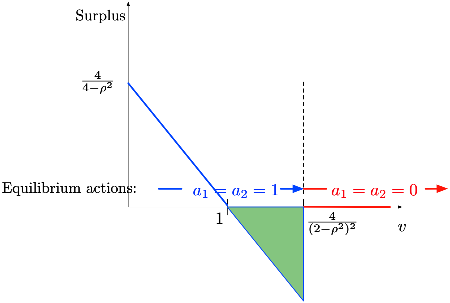
# Example: Equilibrium

**Is surplus monotone in valuations?**

- Two users $n = 2$ with valuations $v_1 = v_2 = v$ and correlation $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$



- (Stackelberg) Equilibrium is



- For intermediate values of $v$, i.e., $v \in [1, \frac{4}{(2-\rho^2)^2}]$, total surplus is negative.

# Price Characterization

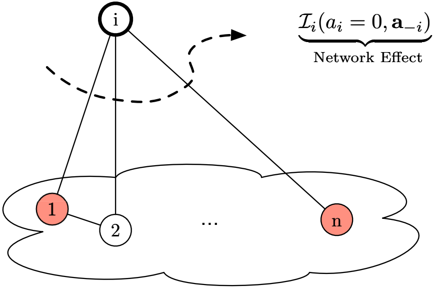
- For any v and $\Sigma$, the prices to sustain actions $a_i = 1$ and $\mathrm{a}_{-i} \in \{0,1\}^{n-1}$ satisfies

$$p_i - v_i \mathcal{I}_i(a_i = 1, \mathrm{a}_{-i}) \geq -v_i \mathcal{I}_i(a_i = 0, \mathrm{a}_{-i}) \Rightarrow p_i^* = v_i \left( \mathcal{I}_i(a_i = 1, \mathrm{a}_{-i}) - \mathcal{I}_i(a_i = 0, \mathrm{a}_{-i}) \right)$$

# Price Characterization

- For any v and $\Sigma$, the prices to sustain actions $a_i = 1$ and $\mathrm{a}_{-i} \in \{0,1\}^{n-1}$ satisfies

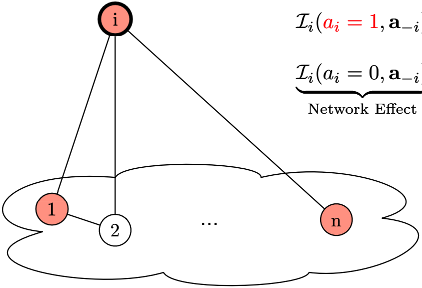
$$p_i - v_i \mathcal{I}_i(a_i = 1, \mathrm{a}_{-i}) \geq -v_i \mathcal{I}_i(a_i = 0, \mathrm{a}_{-i}) \Rightarrow p_i^* = v_i \left(\mathcal{I}_i(a_i = 1, \mathrm{a}_{-i}) - \mathcal{I}_i(a_i = 0, \mathrm{a}_{-i})\right)$$


$$\underbrace{\mathcal{I}_i(a_i = 0, \mathbf{a}_{-i})}_{\text{Network Effect}}$$

# Price Characterization

- For any v and $\Sigma$, the prices to sustain actions $a_i = 1$ and $\mathrm{a}_{-i} \in \{0,1\}^{n-1}$ satisfies

$$p_i - v_i\mathcal{I}_i(a_i = 1, \mathrm{a}_{-i}) \geq -v_i\mathcal{I}_i(a_i = 0, \mathrm{a}_{-i}) \Rightarrow p_i^* = v_i\left(\mathcal{I}_i(a_i = 1, \mathrm{a}_{-i}) - \mathcal{I}_i(a_i = 0, \mathrm{a}_{-i})\right)$$



$$\mathcal{I}_i(a_i = 1, \mathbf{a}_{-i}) =$$

$$\underbrace{\mathcal{I}_i(a_i = 0, \mathbf{a}_{-i})}_{\text{Network Effect}} + \underbrace{\frac{\left(\sigma_i^2 - \mathcal{I}_i(a_i = 0, \mathbf{a}_{-i})\right)^2}{1 + (\sigma_i^2 - \mathcal{I}_i(a_i = 0, \mathbf{a}_{-i}))}}_{\text{User } i\text{'s Effect}}$$

# Price Characterization

### Theorem

*For any* $\mathrm{v}$ *and* $\Sigma$, *the optimal prices to sustain action profile* $\mathrm{a} \in \{0,1\}^n$ *are*

$$p_i = \begin{cases} v_i \frac{\left(\sigma_i^2 - \mathcal{I}_i(a_i=0,\mathrm{a}_{-i})\right)^2}{(\sigma_i^2+1)-\mathcal{I}_i(a_i=0,\mathrm{a}_{-i})}, & \forall a_i = 1, \\ 0, & \forall a_i = 0, \end{cases}$$

*where* $\mathcal{I}_i(a_i = 0, \mathrm{a}_{-i}) = \mathrm{b}_i^T \left(I + B\right)^{-1} \mathrm{b}_i$, $B$ *is obtained by removing row and column* $i$ *from matrix* $\Sigma$ *and all rows and columns* $j$ *for which* $a_j = 0$, *and* $\mathrm{b}_i$ *is its* $i$*-th row.*

- Prices are decreasing in the set of users who share their information.
- $p_i$ is increasing in $\sigma_i^2$.

# Price Characterization

### Theorem

*For any* $\mathrm{v}$ *and* $\Sigma$*, the optimal prices to sustain action profile* $\mathrm{a} \in \{0,1\}^n$ *are*

$$p_i = \begin{cases} v_i \frac{\left(\sigma_i^2 - \mathcal{I}_i(a_i=0,\mathrm{a}_{-i})\right)^2}{(\sigma_i^2+1)-\mathcal{I}_i(a_i=0,\mathrm{a}_{-i})}, & \forall a_i = 1, \\ 0, & \forall a_i = 0, \end{cases}$$

*where* $\mathcal{I}_i(a_i = 0, \mathrm{a}_{-i}) = \mathrm{b}_i^T \left(I + B\right)^{-1} \mathrm{b}_i$*,* $B$ *is obtained by removing row and column* $i$ *from matrix* $\Sigma$ *and all rows and columns* $j$ *for which* $a_j = 0$*, and* $\mathrm{b}_i$ *is its* $i$*-th row.*

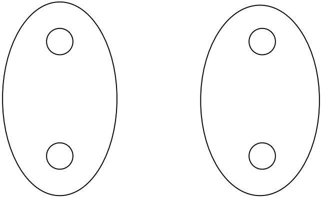
- Prices are decreasing in the set of users who share their information.
- $p_i$ is increasing in $\sigma_i^2$.

# Inefficiency I

- "low-value users": $\mathcal{V}^{(l)} = \{i \in \mathcal{V} : \ v_i \leq 1\}$ .
- "high-value users": $\mathcal{V}^{(h)} = \{i \in \mathcal{V} : \ v_i > 1\}$.

Lemma

*All low-value users share their data in equilibrium.*



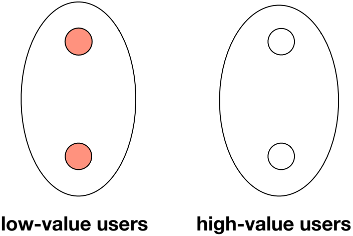
**low-value users** **high-value users**

# Inefficiency I

- "low-value users": $\mathcal{V}^{(l)} = \{i \in \mathcal{V} : \ v_i \leq 1\}$ .
- "high-value users": $\mathcal{V}^{(h)} = \{i \in \mathcal{V} : \ v_i > 1\}$.

Lemma

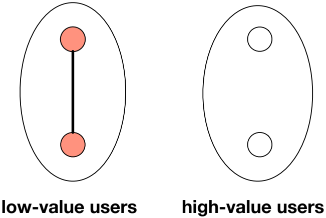
*All low-value users share their data in equilibrium.*



# Inefficiency II

**Theorem**

1. *Suppose high-value users are uncorrelated with others. Then the equilibrium is efficient.*
2. *Suppose at least one high-value user is correlated with a low-value user. Then there exists $\bar{\mathrm{v}} \in \mathbb{R}^{|\mathcal{V}^{(h)}|}$ such that for $\mathrm{v}^{(h)} \geq \bar{\mathrm{v}}$ the equilibrium is inefficient.*
3. *Suppose high-value users $\tilde{\mathcal{V}}^{(h)} \subseteq \mathcal{V}^{(h)}$ are correlated with at least one other high-value user (and no high-low correlation). Then for each $i \in \tilde{\mathcal{V}}^{(h)}$ there exists $\bar{v}_i > 1$ such that if for any $i \in \tilde{\mathcal{V}}^{(h)}$ $v_i < \bar{v}_i$, the equilibrium is inefficient*

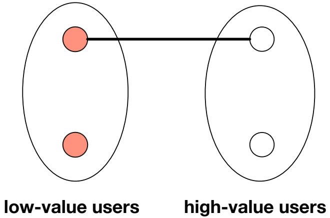


# Inefficiency II

## Theorem

1. *Suppose high-value users are uncorrelated with others. Then the equilibrium is efficient.*
2. *Suppose at least one high-value user is correlated with a low-value user. Then there exists $\bar{v} \in \mathbb{R}^{|\mathcal{V}^{(h)}|}$ such that for $v^{(h)} \geq \bar{v}$ the equilibrium is inefficient.*
3. *Suppose high-value users $\tilde{\mathcal{V}}^{(h)} \subseteq \mathcal{V}^{(h)}$ are correlated with at least one other high-value user (and no high-low correlation). Then for each $i \in \tilde{\mathcal{V}}^{(h)}$ there exists $\bar{v}_i > 1$ such that if for any $i \in \tilde{\mathcal{V}}^{(h)}$ $v_i < \bar{v}_i$, the equilibrium is inefficient*

Part 2 captures inefficiencies from externalities

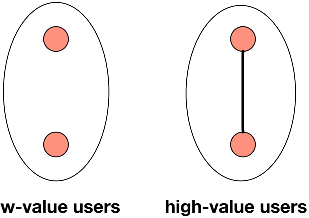


# Inefficiency II

Theorem

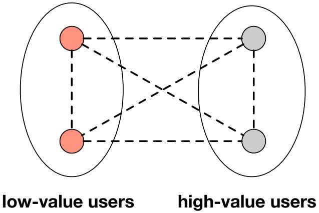
1. *Suppose high-value users are uncorrelated with others. Then the equilibrium is efficient.*
2. *Suppose at least one high-value user is correlated with a low-value user. Then there exists* $\bar{\mathrm{v}} \in \mathbb{R}^{|\mathcal{V}^{(h)}|}$ *such that for* $\mathrm{v}^{(h)} \geq \bar{\mathrm{v}}$ *the equilibrium is inefficient.*
3. *Suppose high-value users* $\tilde{\mathcal{V}}^{(h)} \subseteq \mathcal{V}^{(h)}$ *are correlated with at least one other high-value user (and no high-low correlation). Then for each* $i \in \tilde{\mathcal{V}}^{(h)}$ *there exists* $\bar{v}_i > 1$ *such that if for any* $i \in \tilde{\mathcal{V}}^{(h)}$ $v_i < \bar{v}_i$*, the equilibrium is inefficient*

Part 3 captures inefficiencies from depressed prices



# Are Data Markets Beneficial?

Social surplus$(\mathsf{a}^{\mathrm{E}})$



Corollary

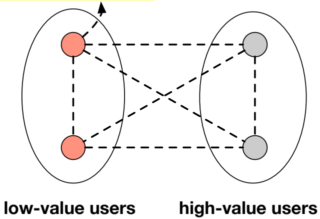
*If* $\sum_{i\in\mathcal{V}^{(h)}}(v_i-1)\mathcal{I}_i(\mathcal{V}^{(l)}) > \sum_{i\in\mathcal{V}^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V})$*, then welfare improves when data markets are shut down. In terms of primitives:*

$$\sum_{i\in\mathcal{V}^{(h)}}\left((v_i-1)\frac{\sum_{j\in\mathcal{V}^{(l)}}\Sigma_{ij}^2}{||\Sigma^{(l)}||_1+1}\right) > \sum_{i\in\mathcal{V}^{(l)}}\sigma_i^2(1-v_i)$$

# Are Data Markets Beneficial?

$$\text{Social surplus}(\mathsf{a}^{\mathrm{E}}) \leq \sum_{i \in \mathcal{V}^{(l)}} (1 - v_i)\mathcal{I}_i(\textcolor{red}{\mathcal{V}})$$

$$(1 - v_i)\mathcal{I}_i(\mathcal{V}^e) \leq (1 - v_i)\mathcal{I}_i(\mathcal{V})$$



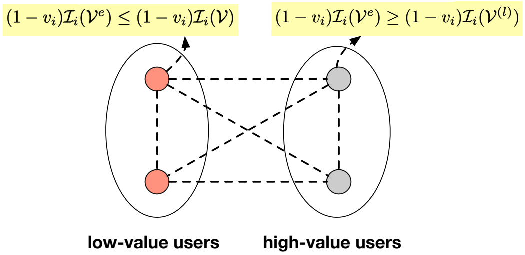
**Corollary**

*If $\sum_{i \in \mathcal{V}^{(h)}} (v_i - 1)\mathcal{I}_i(\mathcal{V}^{(l)}) > \sum_{i \in \mathcal{V}^{(l)}} (1 - v_i)\mathcal{I}_i(\mathcal{V})$, then welfare improves when data markets are shut down. In terms of primitives:*

$$\sum_{i \in \mathcal{V}^{(h)}} \left( (v_i - 1) \frac{\sum_{j \in \mathcal{V}^{(l)}} \Sigma_{ij}^2}{||\Sigma^{(l)}||_1 + 1} \right) > \sum_{i \in \mathcal{V}^{(l)}} \sigma_i^2 (1 - v_i)$$

# Are Data Markets Beneficial?

$$\text{Social surplus}(\mathsf{a}^{\mathrm{E}}) \leq \sum_{i \in \mathcal{V}^{(l)}} (1 - v_i)\mathcal{I}_i(\mathcal{V}) + \sum_{\mathcal{V}^{(h)}} (1 - v_i)\mathcal{I}_i(\mathcal{V}^{(l)}).$$



$(1 - v_i)\mathcal{I}_i(\mathcal{V}^e) \leq (1 - v_i)\mathcal{I}_i(\mathcal{V})$

$(1 - v_i)\mathcal{I}_i(\mathcal{V}^e) \geq (1 - v_i)\mathcal{I}_i(\mathcal{V}^{(l)})$

Corollary

*If* $\sum_{i \in \mathcal{V}^{(h)}} (v_i - 1)\mathcal{I}_i(\mathcal{V}^{(l)}) > \sum_{i \in \mathcal{V}^{(l)}} (1 - v_i)\mathcal{I}_i(\mathcal{V})$*, then welfare improves when data markets are shut down. In terms of primitives:*

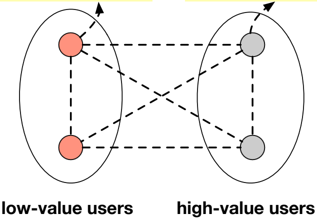
$$\sum_{i \in \mathcal{V}^{(h)}} \left( (v_i - 1) \frac{\sum_{j \in \mathcal{V}^{(l)}} \Sigma_{ij}^2}{||\Sigma^{(l)}||_1 + 1} \right) > \sum_{i \in \mathcal{V}^{(l)}} \sigma_i^2 (1 - v_i)$$

# Are Data Markets Beneficial?

$$\text{Social surplus}(\mathsf{a}^{\mathrm{E}}) \leq \sum_{i\in\mathcal{V}^{(l)}} (1-v_i)\mathcal{I}_i(\mathcal{V}) + \sum_{\mathcal{V}^{(h)}}(1-v_i)\mathcal{I}_i(\mathcal{V}^{(l)}).$$

$(1-v_i)\mathcal{I}_i(\mathcal{V}^e) \leq (1-v_i)\mathcal{I}_i(\mathcal{V})$

$(1-v_i)\mathcal{I}_i(\mathcal{V}^e) \geq (1-v_i)\mathcal{I}_i(\mathcal{V}^{(l)})$



### Corollary

*If $\sum_{i\in\mathcal{V}^{(h)}}(v_i-1)\mathcal{I}_i(\mathcal{V}^{(l)}) > \sum_{i\in\mathcal{V}^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V})$, then welfare improves when data markets are shut down. In terms of primitives:*

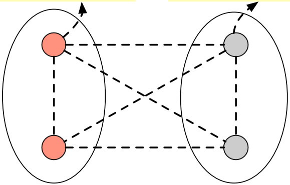
$$\sum_{i\in\mathcal{V}^{(h)}}\left((v_i-1)\frac{\sum_{j\in\mathcal{V}^{(l)}}\Sigma_{ij}^2}{||\Sigma^{(l)}||_1+1}\right) > \sum_{i\in\mathcal{V}^{(l)}}\sigma_i^2(1-v_i)$$

# Are Data Markets Beneficial?

$$\text{Social surplus}(\mathsf{a}^{\mathrm{E}}) \leq \sum_{i \in \mathcal{V}^{(l)}} (1 - v_i)\mathcal{I}_i(\mathcal{V}) + \sum_{\mathcal{V}^{(h)}} (1 - v_i)\mathcal{I}_i(\mathcal{V}^{(l)}).$$

$(1 - v_i)\mathcal{I}_i(\mathcal{V}^e) \leq (1 - v_i)\mathcal{I}_i(\mathcal{V})$

$(1 - v_i)\mathcal{I}_i(\mathcal{V}^e) \geq (1 - v_i)\mathcal{I}_i(\mathcal{V}^{(l)})$



**low-value users**

**high-value users**

## Corollary

*If* $\sum_{i \in \mathcal{V}^{(h)}} (v_i - 1)\mathcal{I}_i(\mathcal{V}^{(l)}) > \sum_{i \in \mathcal{V}^{(l)}} (1 - v_i)\mathcal{I}_i(\mathcal{V})$*, then welfare improves when data markets are shut down. In terms of primitives:*

$$\sum_{i \in \mathcal{V}^{(h)}} \left( (v_i - 1) \frac{\sum_{j \in \mathcal{V}^{(l)}} \Sigma_{ij}^2}{||\Sigma^{(l)}||_1 + 1} \right) > \sum_{i \in \mathcal{V}^{(l)}} \sigma_i^2 (1 - v_i)$$

# Generalization I: Beyond Normal Signals and MSE

- Relax the functional form restrictions and consider the following general conditions:
  1. Monotonicity
  2. Submodularity
- Our baseline setup satisfies these two conditions.
- Assume Properties 1-2 hold. All the results (including inefficiencies we identified) continue to hold.

# Generalization I: Beyond Normal Signals and MSE

- Relax the functional form restrictions and consider the following general conditions:
  1. Monotonicity
  2. Submodularity
- Our baseline setup satisfies these two conditions.
- Assume Properties 1-2 hold. All the results (including inefficiencies we identified) continue to hold.

# Other Generalizations

1. The same results generalize when the platform does not know the value of users.
2. The same results generalize when the platform does not know the correlation structure but has beliefs over it.
3. Similar results hold when there are competing platforms.
    - Competition does not necessarily improve efficiency, and may worsen it as the next example shows.

# Competition Between Platforms

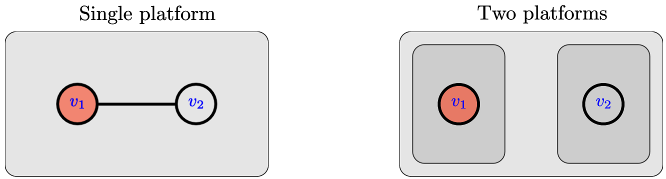
- Consider two competing platforms.
  1. Users simultaneously decide which platform, if any, to join.
     - $c_i : 2^{\mathcal{V}} \to \mathbb{R}^+$: joining value of user $i$ and is monotone in the set of joined users.
     - $b_i \in \{0, 1, 2\}$: the joining decision of user $i$.
     - $J_1 = \{i \in \mathcal{V} \ : \ b_i = 1\}$ and $J_2 = \{i \in \mathcal{V} \ : \ b_i = 2\}$.
  2. The platforms simultaneously offer price vectors $\mathrm{p}^{J_1}$ and $\mathrm{p}^{J_2}$.
  3. Users simultaneously make their data sharing decisions.
- Pure strategy equilibrium for joining decision may not exist.
- There exists a mixed strategy joining equilibrium in which users join each platform with probability 1/2.
- The results generalize when there are also information leakages between platforms.

# Competition Between Platforms

- Consider two competing platforms.
  1. Users simultaneously decide which platform, if any, to join.
     - $c_i : 2^{\mathcal{V}} \to \mathbb{R}^+$: joining value of user $i$ and is monotone in the set of joined users.
     - $b_i \in \{0, 1, 2\}$: the joining decision of user $i$.
     - $J_1 = \{i \in \mathcal{V} \ : \ b_i = 1\}$ and $J_2 = \{i \in \mathcal{V} \ : \ b_i = 2\}$.
  2. The platforms simultaneously offer price vectors $\mathrm{p}^{J_1}$ and $\mathrm{p}^{J_2}$.
  3. Users simultaneously make their data sharing decisions.
- Pure strategy equilibrium for joining decision may not exist.
- There exists a mixed strategy joining equilibrium in which users join each platform with probability $1/2$.
- The results generalize when there are also information leakages between platforms.

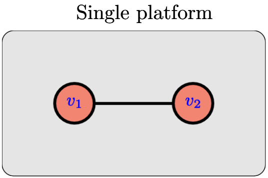
# Does Competition Help Efficiency?

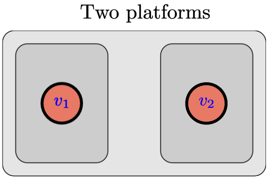
- Two users with correlated data, $v_1 < 1$, and constant joining value $c$.
- *Competition improves equilibrium surplus:* If $v_2 \gg 1$:
  - Under monopoly, only user 1 shares.
  - With competition, users join different platforms and user 1 shares.
  - $\Rightarrow$ Equilibrium surplus improves because the data of user 1 does not leak information about user 2.
- *Competition reduces equilibrium surplus:* If $v_2 < 1$:
  - Under monopoly, both users share.
  - With competition, users join different platforms and they both share.
  - $\Rightarrow$ Equilibrium surplus reduces because the platforms do not gain from the data externality.



# Does Competition Help Efficiency?

- Two users with correlated data, $v_1 < 1$, and constant joining value $c$.
- *Competition improves equilibrium surplus:* If $v_2 \gg 1$:
  - Under monopoly, only user 1 shares.
  - With competition, users join different platforms and user 1 shares.
  - $\Rightarrow$ Equilibrium surplus improves because the data of user 1 does not leak information about user 2.
- *Competition reduces equilibrium surplus:* If $v_2 < 1$:
  - Under monopoly, both users share.
  - With competition, users join different platforms and they both share.
  - $\Rightarrow$ Equilibrium surplus reduces because the platforms do not gain from the data externality.





# Inefficiency with competition

**Theorem**

1. *Suppose high-value users are uncorrelated with others. Then the equilibrium is efficient if and only if* $c_i(\mathcal{V}) - c_i(\{i\}) \geq v_i \mathcal{I}_i(\mathcal{V}^{(l)} \setminus \{i\})$ *for all* $i \in \mathcal{V}^{(l)}$.
2. *Suppose there is high-low correlation. Then there exist* $\bar{\mathrm{v}} \in \mathbb{R}^{|\mathcal{V}^{(h)}|}$ *and* $\underline{\mathrm{v}} \in \mathbb{R}^{|\mathcal{V}^{(l)}|}$ *such that when* $\mathrm{v}^{(h)} \geq \bar{\mathrm{v}}$ *and* $\mathrm{v}^{(l)} \geq \underline{\mathrm{v}}$ *the equilibrium is inefficient.*
3. *Suppose high-value users* $\tilde{\mathcal{V}}^{(h)} \subseteq \mathcal{V}^{(h)}$ *are correlated with at least one other high-value user (and no high-low correlation). Then for each* $i \in \tilde{\mathcal{V}}^{(h)}$ *there exists* $\bar{v}_i > 0$ *such that if for any* $i \in \tilde{\mathcal{V}}^{(h)}$ $v_i < \bar{v}_i$, *the equilibrium is inefficient*

- Efficiency is now even harder to achieve.

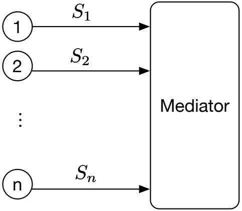
# Taxation

- What can be done about inefficiency?
- Person-specific taxes can decentralize the first best (not surprisingly)
- But such taxes require a social planner to have too much information about each individual.
- Also uniform taxes do not always improve efficiency of economic surplus.

# Mediated Data Sharing

- We next investigate an alternative architecture of data markets.
- Consider the following

"de-correlation" scheme: $\hat{S} = \Sigma^{-1}\mathsf{S}$ for $\mathsf{S} = (S_1, \ldots, S_n)$

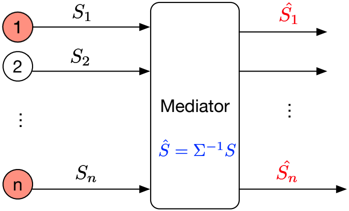


- With this linear transformation of $S$, we have:
  1. $X_i$ and $\hat{S}_{-i}$ have zero correlation
  2. $X_i$ and $\hat{S}_i$ are fully correlated

# Mediated Data Sharing

- We next investigate an alternative architecture of data markets.
- Consider the following

"de-correlation" scheme: $\hat{S} = \Sigma^{-1}S$ for $S = (S_1, \ldots, S_n)$

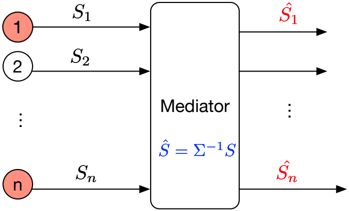


- With this linear transformation of $S$, we have:
  1. $X_i$ and $\hat{S}_{-i}$ have zero correlation
  2. $X_i$ and $\hat{S}_i$ are fully correlated

# Mediated Data Sharing

- We next investigate an alternative architecture of data markets.
- Consider the following

“de-correlation” scheme: $\hat{S} = \Sigma^{-1}S$ for $S = (S_1, \ldots, S_n)$



- With this linear transformation of $S$, we have:
  1. $X_i$ and $\hat{S}_{-i}$ have zero correlation
  2. $X_i$ and $\hat{S}_i$ are fully correlated

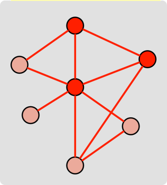
# Mediated Data Sharing (II)

**Lemma**

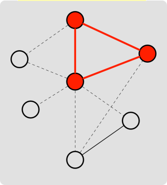
*With de-correlation, leaked information about user $i$ is*

$$\widehat{\mathcal{I}}_i(\mathsf{a}) = \sigma_i^2 - \min_{\hat{x}_i} \mathbb{E}\left[\left(X_i - \hat{x}_i\left(\hat{\mathsf{S}}_\mathsf{a}\right)\right)^2\right] = \begin{cases} 0, & a_i = 0, \\ \mathcal{I}_i(a_i, \mathsf{a}_{-i}), & a_i = 1. \end{cases}$$

- De-correlation removes the correlation between any user who does not wish to share her data and all other users, while maintaining the correlation among users sharing their data.



Without decorrelation



With decorrelation

# Efficiency with De-correlation

Theorem

*Let* $(\hat{\mathsf{a}}^{\mathrm{E}}, \hat{\mathsf{p}}^{\mathrm{E}})$ *and* $(\mathsf{a}^{\mathrm{E}}, \mathsf{p}^{\mathrm{E}})$ *denote the equilibrium with and without the de-correlation scheme, respectively. Then*

$$\text{Social surplus}(\hat{\mathsf{a}}^{\mathrm{E}}) \geq \mathsf{max} \left\{ \text{Social surplus}(\mathsf{a}^{\mathrm{E}}), 0 \right\}.$$

- Intuitively, with de-correlation:
  - High-value users never contribute negative value. Hence social surplus is always non-negative.
  - Negative externalities are lessened, so social surplus always improves.
- But de-correlation does not guarantee first best.

# Conclusion

- A contribution to our understanding of the effects of externalities in data markets.
- Main results:
  - Depressed data prices.
  - Potentially too much data being transacted.
  - Shutting down data markets may be socially beneficial.
  - Introducing mediated data transactions may improve welfare and in the presence of such interactions it is never optimal to shut down data markets.
- Much to be done!

**Thank You!**

# Generalization: Unknown Correlations

- The platform does not know the (realized) correlation among users, but knows its distribution.

Theorem

1. *Suppose every high-value user is uncorrelated with all other users almost surely, i.e.,* $\mathbb{P}_{\Sigma \sim \mu}\left(\Sigma_{ij} = 0\right) = 1$*, for all* $i \in \mathcal{V}^{(h)}, j \in \mathcal{V}^{(l)}$*. Then the equilibrium is efficient.*
2. *Suppose there exists high-value* $i \in \mathcal{V}^{(h)}$ *and low-value users* $j \in \mathcal{V}^{(l)}$ *who are correlated, i.e.,* $\mathbb{P}_{\Sigma \sim \mu}\left(\Sigma_{ij} \neq 0\right) > 0$*. Then there exists* $\bar{\mathrm{v}} \in \mathbb{R}^{|\mathcal{V}^{(h)}|}$ *such that for* $\mathrm{v}^{(h)} \geq \bar{\mathrm{v}}$ *the equilibrium is inefficient.*
3. *Suppose every high-value users* $\tilde{\mathcal{V}}^{(h)} \subseteq \mathcal{V}^{(h)}$ *are correlated with at least one other high-value user with positive probability (and no high-low correlation). Then for each* $i \in \tilde{\mathcal{V}}^{(h)}$ *there exists* $\bar{v}_i > 1$ *such that if for any* $i \in \tilde{\mathcal{V}}^{(h)}$ $v_i < \bar{v}_i$*, the equilibrium is inefficient*

# Competition with Data Prices

- Consider the following timing:
  1. Platforms simultaneously offer price vectors $\mathrm{p}^1 \in \mathbb{R}^n$ and $\mathrm{p}^2 \in \mathbb{R}^n$.
  2. Users simultaneously decide which platform, if any, to join, i.e., $\mathrm{b} = \{b_i\}_{i \in \mathcal{V}}$ (which determines $J_1$ and $J_2$) and whether to share their data.
- Now data prices attract consumers to a platform.
- Now price competition leads to possible discontinuities in payoffs (as in standard Bertrand competition).
- In this setting, mixed equilibrium always exists. Moreover, similar inefficiency results holds (with extra conditions on the extreme values of the joining value function)

# Generalization: Unknown valuations

- So far we assumed that the platform knows the value of privacy of different users.
- The more realistic assumption:
    - The platform does not know the exact valuations of users.
    - But understands that $v_i$, has a distribution with cumulative distribution $F_i$ and density function $f_i$ (with upper support denoted by $v^{\max}$)
- Now equilibria have to be incentive compatible.

### Theorem

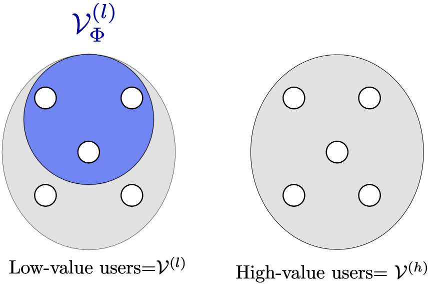
*Suppose for all $i \in \mathcal{V}$, the function $\Phi_i(v) = v + \frac{F_i(v)}{f_i(v)}$ is nondecreasing. For any reported* v, *the equilibrium is given by*

$$\mathrm{a}^{\mathrm{E}}(\mathrm{v}) = argmax_{\mathrm{a}\in\{0,1\}^n} \sum_{i=1}^{n} (1-\Phi_i(v_i))\mathcal{I}_i(\mathrm{a}) + \Phi_i(v_i)\mathcal{I}_i(\mathrm{a}_{-i}, a_i = 0),$$

*and* $p_i^{\mathrm{E}}(v_i) = \int_v^{v_{\max}} \left(\mathcal{I}_i(\mathrm{a}^{\mathrm{E}}(x, \mathrm{v}_{-i})) - \mathcal{I}_i(\mathrm{a}_{-i}^{\mathrm{E}}(x, \mathrm{v}_{-i}), a_i = 0)\right) dx +$
$v_i \left(\mathcal{I}_i(\mathrm{a}^{\mathrm{E}}(v_i, \mathrm{v}_{-i})) - \mathcal{I}_i(\mathrm{a}_{-i}^{\mathrm{E}}(v_i, \mathrm{v}_{-i}), a_i = 0)\right)$. *Moreover, all users report truthfully.*

# Generalization: Unknown valuations Inefficiency

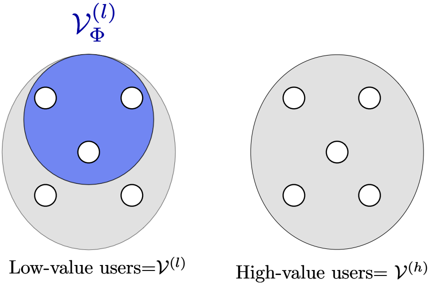
- Let $\mathcal{V}_{\Phi}^{(l)} = \{i \in \mathcal{V} \ : \ \Phi_i(v_i) \leq 1\}$ (i.e., low virtual value replaces low value).
- Now for efficiency we need high-value users to be uncorrelated with low-value users and $\mathcal{V}^{(l)} = \mathcal{V}_{\Phi}^{(l)}$.
    - If some low-value users have virtual value greater than one, then efficiency may fail even in this case.
- The rest of the inefficiency theorem applies as before, but again, by replacing low-value users with low virtual value users.


$\mathcal{V}_{\Phi}^{(l)}$
Low-value users=$\mathcal{V}^{(l)}$
High-value users= $\mathcal{V}^{(h)}$

- The results extend to the case in which correlations are unknown to the platform.

# Generalization: Unknown valuations Inefficiency

- Let $\mathcal{V}_\Phi^{(l)} = \{i \in \mathcal{V} \ : \ \Phi_i(v_i) \leq 1\}$ (i.e., low virtual value replaces low value).
- Now for efficiency we need high-value users to be uncorrelated with low-value users and $\mathcal{V}^{(l)} = \mathcal{V}_\Phi^{(l)}$.
  - If some low-value users have virtual value greater than one, then efficiency may fail even in this case.
- The rest of the inefficiency theorem applies as before, but again, by replacing low-value users with low virtual value users.


$\mathcal{V}_\Phi^{(l)}$
Low-value users=$\mathcal{V}^{(l)}$
High-value users= $\mathcal{V}^{(h)}$

- The results extend to the case in which correlations are unknown to the platform.