CS 136: Economics and Computation

Lecture 16
Reputation Systems

David C Parkes
Fall 2018
Market platforms

- Amazon, eBay, Uber, Airbnb,…
- Some design features:
  - Matching and search
  - Dynamic pricing
  - Reputation systems
Market platforms

• Amazon, eBay, Uber, Airbnb,…
• Some design features:
  – Matching and search
  – Dynamic pricing
  – **Reputation systems**

• Enable safe transactions between strangers
• Trust can also established through certification, insurance.

Role of Reputation Systems

• **Moral Hazard**
  – **Low quality actions** because of incentive misalignment.
    • E.g., seller doesn’t ship

• **Adverse selection**
  – **Low quality entry** because of information asymmetry.
    • E.g., fraudulent sellers enter
Role of Reputation Systems

- **Moral Hazard**
  - *Low quality actions* because of incentive misalignment.
    - E.g., seller doesn’t ship
  - *Reputation system addresses via “sanctioning”*

- **Adverse selection**
  - *Low quality entry* because of information asymmetry.
    - E.g., fraudulent sellers enter
  - *Reputation system addresses via “signaling”*

A well functioning reputation system?

Three components:

1. The publicly reported information reflects the past (requires unbiased ratings).
2. Buyers correctly interpret the reputation information.
3. Leading sellers to change action because of this “shadow on the future.”

Also addresses adverse selection:

1. 2. 3. Low quality sellers no longer enter.
The Design Space

• **Who**: One-way or two-way feedback?
• **Who**: Restricted, Anyone, or Required?
• **Timing**: Sequential- or simultaneous report?
• Explicit *incentive* for feedback?
• What to *elicit*, how to aggregate/share?
• Allow *private messaging* between parties?

### Design Features

<table>
<thead>
<tr>
<th></th>
<th>Direction</th>
<th>Who</th>
<th>Elicit</th>
<th>Centr.?</th>
<th>Incentives?</th>
</tr>
</thead>
<tbody>
<tr>
<td>eBay (1.0)</td>
<td>2-way seq</td>
<td>restrict</td>
<td>{+, o, -}, text</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>eBay (2.0)</td>
<td>2-way seq</td>
<td>restrict</td>
<td>{+, o, -}, text</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>1-way</td>
<td></td>
<td>detailed seller rating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amazon</td>
<td>1-way</td>
<td>anyone</td>
<td>stars, text</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Yelp</td>
<td>1-way</td>
<td>anyone</td>
<td>stars, text</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Airbnb</td>
<td>2-way sim</td>
<td>restrict</td>
<td>stars, text</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Uber</td>
<td>2-way seq</td>
<td>restrict</td>
<td>stars</td>
<td>yes</td>
<td>nudge</td>
</tr>
<tr>
<td>Digg</td>
<td>1-way</td>
<td>anyone</td>
<td>vote</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Google Local Guides</td>
<td>1-way</td>
<td>anyone</td>
<td>stars, text</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>
eBay Feedback 1.0

700,000 transactions, 11/06-12/06
nPos / nRatings = 1.0 : 67%
nPos / nRatings > 0.99: 81%

A World Without Reputation

- e.g., Airbnb game

|       | Player 1 | Player 2
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>D (don’t clean)</td>
</tr>
<tr>
<td>C</td>
<td>3, 3</td>
<td>0, 5</td>
</tr>
<tr>
<td>D (party!)</td>
<td>5, 0</td>
<td>2, 2</td>
</tr>
</tbody>
</table>
A World Without Reputation

- e.g., Airbnb game

<table>
<thead>
<tr>
<th></th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0, 5</td>
</tr>
<tr>
<td>D (don’t clean)</td>
<td>2, 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Player 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3, 3</td>
</tr>
<tr>
<td>D (party!)</td>
<td>5, 0</td>
</tr>
</tbody>
</table>

- n players. Match every n-1 rounds. Grim trigger Nash eq:

\[
5 + \frac{2\delta^{n-1}}{1 - (\delta^{n-1})} \leq \frac{3}{1 - (\delta^{n-1})}
\]

Need \( \delta \geq \left(\frac{2}{3}\right)^{\frac{1}{n-1}} \)

For example, \( \delta \geq 0.9996 \) for 1000 players

A Simple Model of Reputation

- Everyone starts with “good reputation”
- (perfect monitoring) Gain a “bad reputation” if ever defect against good player
A Simple Model of Reputation

- Everyone starts with “good reputation”
- (perfect monitoring) Gain a “bad reputation” if ever defect against good player
- Reputational Grim Trigger:
  - play C if both players good reputation
  - else, play D

Nash eq: \[ 5 + \frac{\delta^2}{1 - \delta} \leq \frac{3}{1 - \delta} \] Need: \( \delta \geq \frac{2}{3} \)

Why Leave Feedback?

- Social experience (community)
- Explicit rewards (TB / 20% discount)
- “warm glow”
- Hedonistically rewarding
- Retaliate against seller
- Very easy to do (e.g., Uber)
Feedback

• On eBay, why might negative ratings be missing?
• On Orbitz, why might positive ratings be missing?
• On Yelp, why might neutral ratings be missing?
• On Airbnb, why might ratings generally be overly positive?

In general, lead to bias in the data
  – not missing at random
  – ratings may be false, or leave out info
percentage positive score

$$PP = \frac{n_{\text{pos}}}{n_{\text{ratings}}}$$

67% of sellers had PP = 1.0
81% of sellers had PP >= 0.99

---

**eBay F1.0 (pre 2007)**

- Both parties can leave {-1, 0, +1} feedback
- *Sequential-reveal*: can see feedback by other party before leaving own feedback.

- Concerns?
Lab Experiment (2007)

(Bolton, Grenier & Ockenfels'11)

- **Simultaneous**
  - Only share after both report, or after deadline
  - Less strategic behavior, but also reduced quantity

- **Detailed seller ratings (DSR)** (add B2S rating, after seller feedback or deadline for seller)
  - Less strategic behavior, detailed feedback correlated with quality, and did not reduce quantity
eBay Feedback 2.0 (2007+)

- DSR: additional feedback from buyer
- Available after seller provides feedback or at seller deadline. Public when >= 10 reports.

### eBay 2.0 vs eBay 1.0 ratings

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Average Rating</th>
<th>Number of Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item as described</td>
<td>★★★★★</td>
<td>288</td>
</tr>
<tr>
<td>Communication</td>
<td>★★★★★</td>
<td>284</td>
</tr>
<tr>
<td>Shipping time</td>
<td>★★★★★</td>
<td>289</td>
</tr>
<tr>
<td>Shipping and handling charges</td>
<td>★★★★★</td>
<td>288</td>
</tr>
</tbody>
</table>
Nobody is paying attention

• Click logs: only 1% buyers look at Detailed Seller Ratings
Nobody is paying attention

- Click logs: only 1% buyers look at Detailed Seller Ratings

- May 2008: Keep DSR, and modify CF.
  - Seller can only leave positive feedback, or choose to leave no feedback.

- Does this help?

Klein et al.'15 "Adverse selection and moral hazard in anonymous markets"

DSR feedback of May 2007 seller cohort

remove ability for sellers to leave neg feedback
Klein et al.'15 “Adverse selection and moral hazard in anonymous markets”

DSR feedback of May 2007 seller cohort

Klein et al.'15 “Adverse selection and moral hazard in anonymous markets”

DSR feedback of May 2007 seller cohort
Klein et al.’15 “Adverse selection and moral hazard in anonymous markets”

DSR feedback of May 2007 seller cohort

Exit of May 2007 seller cohort

- DSR introduced (May 2008)
- CF change

Klein et al.’15 “Adverse selection and moral hazard in anonymous markets”

DSR feedback of May 2007 seller cohort

Exit of May 2007 seller cohort

- DSR introduced (May 2008)
- CF change

Conclude: exit rate doesn’t change.
Dropping seller neg feedback helps with moral hazard.
2011 onwards
Feedback Bias Persists!
(Nosko and Tadelis’15)

• Analysis 2011 buyer cohort over 3 years: median CF percent positive 1.0, mean 0.99!
• Ground truth (Oct 2011 transactions):
  – 0.4% negative, conventional feedback
  – 1% dispute tickets opened with eBay
  – DSR suggests 3.4% unsatisfactory
  – ~3.3% post-transaction B2S messages negative sentiment (June 2011 data)

• Still seems to be a bias. Why?
What else could eBay do?

Effective Percent Positive
(Nosko and Tadelis’15)

$$PP = \frac{n\_pos}{n\_ratings}$$

Sellers that transact with 2011 cohort of eBay buyers
Effective Percent Positive
(Nosko and Tadelis’15)

\[ PP = \frac{n_{\text{pos}}}{n_{\text{ratings}}} \quad \text{EPP} = \frac{n_{\text{pos}}}{n_{\text{tx}}} \]

Sellers that transact with 2011 cohort of eBay buyers

Dec 2011 A/B experiment: for 10% buyers, promote sellers with high EPP.
Effective Percent Positive
(Nosko and Tadelis’15)

\[
PP = \frac{n_{\text{pos}}}{n_{\text{ratings}}}
\]

\[
EPP = \frac{n_{\text{pos}}}{n_{\text{tx}}}
\]

Sellers that transact with 2011 cohort of eBay buyers

Dec 2011 A/B experiment: for 10% buyers, promote sellers with high EPP. Result: buyers significantly more likely to return and purchase again within 180 days.

Fake reviews?

- Compare Expedia and TripAdvisor reviews
- Greater bunching at extreme ratings on TripAdvisor, and correlated with local competition.

Mayzlin et al ‘14.
Fake it Till You Make It

M. Luca and G. Zervas (2016)

• ~16% of restaurant reviews on Yelp are algorithmically filtered. Take proxy of review fraud

• Results:
  – more likely to commit review fraud when reputation is weak (few reviews, recent bad reviews)
  – chain restaurants – which benefit less from Yelp – are also less likely to commit review fraud.
  – when face increased competition, become more likely to receive unfavorable fake reviews.

• Corroborate with a second data set, a sting conducted by Yelp. Those soliciting fake reviews: no chains, lower Yelp ratings, fewer reviews.

Rating complexity @

Information on listing (quant ratings, stories). No quantitative info on guests.

From guest: stars, a story, and private host feedback.
To platform: binary up/down on host.

From host: post on guest page; private guest feedback.
To platform: rate, and up/down recommend.
Analysis of Airbnb RepSys
(Fradkin et al. 2016)

- Pre May’14: sequential-reveal (30 days, any review immediately posted)
- Working ok. How know?
  - conditioned on “thumbs down” to platform, <10% gave 5* ‘overall experience’
  - conditioned on “thumbs up” to platform, 74% gave 5* ‘overall experience’, and 20% gave 4*.
Analysis of Airbnb RepSys
(Fradkin et al. 2016)

• Pre May’14: sequential-reveal (30 days, any review immediately posted)
• Working ok. How know?
  – conditioned on “thumbs down” to platform, <10% gave 5* ‘overall experience’
  – conditioned on “thumbs up” to platform, 74% gave 5* ‘overall experience’, and 20% gave 4*.

• 2014 on-platform experiment: simultaneous- vs seq-reveal. Simultaneous design:
  – more feedback (+2% guests, +7% hosts)
  – 5* feedback on hosts down 1.6%, 5% feedback on guests down 0.4%

Disruption
Sesame Credit, also known as Zhima Credit, is a private credit scoring and loyalty program system developed by Ant Financial Services Group (AFSG), an affiliate of the Chinese Alibaba Group.

“Zhima Credit’s algorithm considers not only whether you repay your bills but also what you buy, what degrees you hold, and the scores of your friends.”
On my first day in Shanghai, I opened Zhima Credit to scan a yellow bike.

A scan of a bike’s QR code revealed a four-digit number that unlocked the back wheel, and a ride across town cost roughly 15 cents. Because of my middling score, however, I had to pay a $30 deposit before I could scan my first bike.

Nor could I get deposit-free hotel stays or GoPro rentals, or free umbrella rentals.

[Alipay knows that..]

at 1 pm on the afternoon of August 26, I rented an Ofo brand bike ... at 1:24 pm I bought a snack in the mall ... got in a Didi car bound for a neighborhood to the northwest... entered a supermarket (accepts only Alipay at checkout) at 3:36 pm I bought bananas, cheese, and crackers. ... got in a taxi, and that I arrived at my destination at 4:01 pm. ... at 4:19 pm I paid $8 for an Amazon delivery.

... I rented another Ofo bike outside a hotel in central Shanghai, cycled 10 minutes, and at 7:11 pm parked it outside a popular restaurant (Ant Financial is a strategic investor in Ofo, might know the route I took.)
Summary: Reputation

• Reputation systems have a crucial role in promoting trust in market platforms
• Stylized model: the “reputation game.”
• Watch out for biases in reports
• Data analytics based on private, on-platform messages also have a role
• Reputation systems are becoming more powerful