Discussion: The Economics of Social Data
(Bergemann, Bonatti, and Gan)

Joseph Abadi

February 21, 2020
Introduction

- Rise of data intermediation in several applications:
  1. Product platforms (e.g. Amazon, Uber)
  2. Social platforms (e.g. Facebook, Google)
  3. Data brokers (e.g. Oracle, Equifax)

- Benchmark economic theories: **Under-provision** of data
  - **Non-rivalry** of data → Social value exceeds private value
  - With privacy concerns, data ownership can undo negative effects (c.f. Posner/Weyl 2018)

- This paper: Data has a **social** dimension!
  - Externality: Potential **over-provision** of data
  - Study how data intermediation affects **downstream market**
Result 1: The data externality

- Consider a setting with common values only \((w_i = \theta)\) and information structure

\[ s_i = w_i + \epsilon_i \]

- With \(k\) signals, the firm infers

\[ \mathbb{E}[\theta|S] = (1 - x(k))\mu_\theta + x(k)\bar{s} \]

(where \(1 - x(k) = \) weight on prior).

- For large \(k\), consumers get expected utility

\[ \mathbb{E}[u_i] = \frac{1}{2} \text{Var}(\theta - p) \approx \frac{1}{2}(1 - \frac{1}{2}x(k))^2\sigma_\theta^2 \]

- Two main insights:
  1. Same \(k\) enters into all consumers' utility functions  
     \(\Rightarrow\) Externality
  2. Concavity of \(x(k)\)  \(\Rightarrow\) Diminishing marginal value of info.
Implications of data externalities

- Whether data is sold does not depend on whether it is used to create or extract value
  - Extension with firm that can chooses both price/quality
- Compensation for data goes to zero in large markets
- Access to more signals per consumer ⇒ More data acquisition
  - Collection of signals gives more info. about aggregate demand
Result 2: Optimality of privacy

- Second result: Data broker preserves consumers’ **anonymity**
  - Anonymized data **minimizes** surplus loss, holding data externality fixed

- Simple ex.: In a setting with only private values, social surplus is decreasing in
  \[ \text{Cov}(w_i, p_i) \]

  \[ \Rightarrow \] Broker and consumer can never agree to trade data
  - Is this why we don’t see much personalized pricing?
  - Interesting extension to group pricing (e.g. based on location)

- Even **without** personalized pricing, data intermediation can cause inefficiencies
  - Prop. in paper: Anonymized data \( \Leftrightarrow \) Inefficiency!
Relation to Acemoglu et al.

<table>
<thead>
<tr>
<th></th>
<th>Data market</th>
<th>Privacy concerns</th>
<th>Info. structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergemann et al.</td>
<td>“Data intermediary”</td>
<td>Endogenous (downstream mkt)</td>
<td>Designed by intermediary</td>
</tr>
<tr>
<td>Acemoglu et al.</td>
<td>“Platform”</td>
<td>Exogenous</td>
<td>Given</td>
</tr>
</tbody>
</table>

Authors can ask:

- How is downstream inefficiency affected by data intermediation?
- When will privacy be preserved? (Information design)
- Inefficiency beyond lack of privacy?
Comment 1: Policy implications

- What does this model teach us about policy?
  - Externality (good or bad) ⇒ Data intermediation at low price

- Key feature of benchmark environment: **Downstream** monopolistic pricing
  - Anonymized data ⇔ Social inefficiency

- Competition in downstream market? Examples:
  - Competition for user attention/ad targeting
  - Competitive credit/insurance market (Rothschild-Stiglitz)

- Distortion from monopolistic **data provision**?
  - Regulate data provider or downstream market?
Comment 2: Model uncertainty

- What if there were uncertainty about the model?
  - Tech companies hire data scientists exactly for this reason!
- Key tradeoff: Value **creation** vs. **extraction**
  - Broker may privately know which force dominates
  - Implications for use of personal vs. aggregate data?
- Requires *dynamic* model of data collection/externalities
  - Possible to think about this question in a two-period version of the model?
Comment 3: Privacy and market segmentation

- How can inefficient market segmentation be prevented?
  - E.g., Hirshleifer effect in insurance markets
- This paper: Benchmark model resembling direct sale of data
  - Examples of direct data use without personalized prices (Amazon, Uber)
  - Value extraction pushes towards privacy
- Greater scope for violation of privacy in some markets?
  - Social networks use extremely specific data in ad targeting (indirect sales)
- Limits of market segmentation with data ownership?
  - Important for benefits of data ownership
Conclusion

- Two key points:
  1. Social data externality: Main determinant of data provision
  2. Privacy: May be preserved even with downstream inefficiencies

- Lots of interesting work to be done beyond benchmark:
  - Policy implications?
  - What if agents face model uncertainty?
  - Implications for privacy in different settings?
    (E.g. indirect sales)

- Overall: Great paper that both answers and raises many stimulating questions