Mining Democracy

A data-driven exploration of the Swiss political landscape

Vincent Etter, Julien Herzen, Patrick Thiran and Matthias Grossglauser
EPFL, Switzerland

COSN’14 - Dublin, Ireland
October 1st, 2014
Motivation

- **Open government** initiatives adopted worldwide
  - Datasets about multiple aspects of state affairs released
- **Voting advice applications** (VAA) set up in several countries
  - Candidates advertise their opinion by answering questions on several political aspects
  - Citizens can answer the same questions and get personalized voting recommendations
- Gives an **unprecedented view** of political opinions
Many questions

- Such data allow to answer many interesting questions:
  - Do politicians and citizens share similar concerns?
  - Could a candidate abuse a VAA?
  - On the contrary, can you use VAAs to monitor politicians?
  - How do voting behaviors change across a country?
  - …
Our laboratory: Switzerland

- **Diversified** party landscape
- Four official **languages**
- **smartvote**: VAA available since 2003
- Direct democracy with **frequent issue votes** on various subjects
  - at both parliamentary and citizen levels
smartvote dataset

- **smartvote pre-electoral opinions** of the 2011 parliamentary elections
  - 2,985 candidates (82.4% of all candidates)
  - 229,133 citizens (~9% of total turnout)
- **Examples** of questions:
  - Should Switzerland embark on negotiations in the next four years to join the EU?
  - How much should the public transport budget be?
- **Possible answers** (mapped to \{0.0, 0.25, 0.5, 0.75, 1.0\}):
  - strongly disagree - disagree - agree - strongly agree
  - less - no change - more
Discriminative questions

- What questions **discriminate** best the opinion of candidates?
- Is the **traditional left/right view** meaningful?
- Use **dimensionality reduction** to find out
- Use SVD on the matrix of candidates’ responses $C$

\[
\begin{pmatrix}
0.5 & 0.25 & \ldots & 0.0 \\
0.75 & 0.5 & \ldots & 1.0 \\
\vdots & \vdots & \ddots & \vdots \\
1.0 & 0.25 & \ldots & 0.75
\end{pmatrix}
\]
Ideological space of candidates

<table>
<thead>
<tr>
<th>Singular vector</th>
<th>First two questions</th>
</tr>
</thead>
</table>
| 1st             | 1. Would you support foreigners who have lived for at least ten years in Switzerland being given voting and electoral rights at municipal level?  
2. Are you in favour of legalizing the status of illegal immigrants? |
| 2nd             | 1. Are you in favour of the complete liberalization of shop opening times?  
2. Should Switzerland conclude an agricultural free trade agreement with the EU? |
| 3rd             | 1. Should Switzerland legalize the consumption of hard and soft drugs?  
2. Should same-sex couples who have registered their partnership be able to adopt children? |
Ideological space of candidates

<table>
<thead>
<tr>
<th>Singular vector</th>
<th>First two questions</th>
</tr>
</thead>
</table>
| 1st             | 1. Would you support foreigners who have lived for at least ten years in Switzerland being given voting and electoral rights at municipal level?  
2. Are you in favour of legalizing the status of illegal immigrants? |
| 2nd             | 1. Are you in favour of the complete liberalization of shop opening times?  
2. Should Switzerland conclude an agricultural free trade agreement with the EU? |
| 3rd             | 1. Should Switzerland legalize the consumption of hard and soft drugs?  
2. Should same-sex couples who have registered their partnership be able to adopt children? |
Ideological space of candidates

<table>
<thead>
<tr>
<th>Singular vector</th>
<th>First two questions</th>
</tr>
</thead>
</table>
| 1st             | 1. Would you support foreigners who have lived for at least ten years in Switzerland being given voting and electoral rights at municipal level?  
2. Are you in favour of legalizing the status of illegal immigrants? |
| 2nd             | 1. Are you in favour of the complete liberalization of shop opening times?  
2. Should Switzerland conclude an agricultural free trade agreement with the EU? |
| 3rd             | 1. Should Switzerland legalize the consumption of hard and soft drugs?  
2. Should same-sex couples who have registered their partnership be able to adopt children? |
Densities

- The density of candidates and citizens in the ideological plane varies
Abusing VAAs

- VAAs are **beneficial** on several aspects
  - Citizens can get **personalized recommendations**, and get to know candidates better
  - Data extracted from VAAs give **great insights** on the political landscape of a country
- Could this data be **misused**?
  - Candidate profiles are public, and used for recommendations
  - Could a **new candidate** use this to his advantage?
Crafting a profile

• smartvote (as most VAAs) simply uses the **Euclidean distance** to compute voting recommendations

• the **50 closest candidates** are recommended, in increasing order of distance to the citizen’s answer

• A **malicious** candidate could thus **tailor** his answers, such that he is:
  • far away from other candidates
  • close to many citizens
Empirical solution

• **Manually** pick your location in the ideological space

• Use the **inverse transformation** to find the answers that get you there
Effect of crafted profile

- We **crafted** the profile corresponding to the star in the previous plot.
- Then, we **re-computed the recommendations** for all 229,133 citizens.
- We checked how many times each candidate appears in the top $R$ recommendations, for $R \in \{1, \ldots, 50\}$. 
Recommendation results

- The crafted profile appears in the 50 closest candidates of nearly half of the citizens!
Quantifying opinion shifts

• Is it possible to detect whether a politician crafted his profile, given the way he votes once elected?

• Parliament votes (2,494 since the 2011 elections) are public

• Requires a mapping VAA answers $\leftrightarrow$ parliament votes

• Learning problem:

  Training data: all VAA responses $C$ and votes $\nu$ on a particular issue

  Predict: vote $\nu_c \in \{\text{yes, no}\}$ of candidate $c$
VAA responses can be used to predict parliament votes

Using only a linear regression, one can predict >= 50% of the votes with >= 95% accuracy.
Opinions shifts

Comparison between votes expected from VAA responses and actual votes cast in parliament (using votes predicted with accuracy > 95%)
Voting patterns at municipality level

- **Dataset**: outcome (% yes) of 245 votes since 1981 in 2,398 municipalities
- Dimensionality reduction highlights **linguistic/cultural** contrasts
Voting patterns at municipality level

- **Dataset**: outcome (% yes) of 245 votes since 1981 in 2,398 municipalities
- Dimensionality reduction highlights **linguistic/cultural** contrasts

"Röstigraben"
Voting patterns at municipality level

- **Dataset**: outcome (% yes) of 245 votes since 1981 in 2,398 municipalities
- Dimensionality reduction highlights linguistic/cultural contrasts
Voting patterns at municipality level

- **Dataset:** outcome (% yes) of 245 votes since 1981 in 2,398 municipalities
- Dimensionality reduction highlights **linguistic/cultural** contrasts
Voting patterns at municipality level
Voting patterns at municipality level

“Röstigraben”
Voting patterns at municipality level

“Röstigraben”

Basel

Zürich

Bern

Lausanne

Geneva

www.predikon.ch/eigenmap
Prediction of national results

• Knowing the result of one municipality in advance (e.g., from polling/survey), can we predict the final result?

• Answer: Yes, but it depends on which municipality!
Prediction of national results

• Knowing the result of **one** municipality in advance (e.g., from polling/survey), can we predict the **final result**?

• Answer: Yes, but it depends on which municipality!

 Ebikon (accuracy 95.9% on test set)
Conclusions

- New massive VAA / open government datasets
- Systematic data-mining highlights ideological/cultural idiosyncrasies
- VAAs can be significantly abused by candidates
- Municipality results allow to uncover interesting patterns and are useful to predict national outcomes

Future/ongoing work:
- Predict vote results for all municipalities
- Formalize/optimize candidate placement in VAAs
Thank you for listening!

www.predikon.ch
Partisan Sharing: Facebook Evidence and Societal Consequences

Jisun An

Social Computing Team, Qatar Computing Research Institute (QCRI)

with Daniele Quercia (Yahoo Labs Barcelona), Jon Crowcroft (Univ. of Cambridge)

COSN 2014
Which one would you choose to read in the waiting room?
Republicans/conservatives spend more time with National Review; Democrats/liberals spend more time with The Nation.
The theory of **selective exposure** holds that people tend to seek out political information confirming their beliefs and avoid challenging information.
SELECTIVE EXPOSURE: EXIST OR NOT?

It exists: people tend to preferentially choose, read, and enjoy partisan news.

Measure what people select to read through survey & experiment on newspaper, magazine, TV/radio programs.
SELECTIVE EXPOSURE: EXIST OR NOT?

It exists: people tend to preferentially choose, read, and enjoy partisan news.

Measure what people select to read through survey & experiment on newspaper, magazine, TV/radio programs.

It does not exist: no evidence for selective exposure in their news diet.

Measure actual exposure through web log & recording & Twitter on TV/radio programs and online news.
What happens after exposure?
How do people react to such exposure?
Sharing is one indication that reveals *what people act on* after the exposure.
Sharing captures how people perceive news information in an unobtrusive way.
Partisan sharing: people tend to share like-minded political information and avoid challenging ones.
OUR GOAL

By looking at news sharing in social media, we aim to examine whether partisan sharing exists or not.
HYPOTHESIS

Factors
- Political leaning
- Amount of political news consumption
- Time

Partisan sharing

Consequences
- Perceived bias of news outlets
- Political knowledge
- Voting probability
HYPOTHESIS

Factors
- Political leaning
- Amount of political news consumption
- Time

Consequences
- Perceived bias of news outlets
- Political knowledge
- Voting probability
HYPOTHESIS

Factors

- Political leaning
- Amount of political news consumption
- Time

Consequences

- Perceived bias of news outlets
- Political knowledge
- Voting probability
HYPOTHESIS

Factors
- Political leaning
- Amount of political news consumption
- Time

Consequences
- Perceived bias of news outlets
- Political knowledge
- Voting probability

Partisan sharing

Factors influence the pathway to consequences through partisan sharing.
NEWS SHARING IN FACEBOOK

Subset of myPersonality users
228,064 US based Facebook users (4.9M URLs)
NEWS SHARING IN FACEBOOK

Subset of myPersonality users
228,064 US based Facebook users (4.9M URLs)
MEASURING PARTISAN SHARING

We measure how balance a user’s news sharing is.

Net Partisan skew reflects partisanship and is based on which news article a user shares on the political leanings (liberal & conservative) of those news articles.
MEASURING PARTISAN SHARING

We measure how balance a user’s news sharing is.

Net partisan skew:

\[ \ln(\text{#conservative news}) - \ln(\text{#liberal news}) \]

- Share only liberal news articles
- 0 Share articles equally on both political leaning
- + Share only conservative news articles

It is ±2 if, for every 7.4 (∼ e²) conservative (liberal) articles, the user shares 1 liberal (conservative) article.

Net Partisan skew reflects partisanship and is based on which news article a user shares on the political leanings (liberal & conservative) of those news articles.
METHODOLOGY

1. Facebook news consumption
   From a list of 100 news papers in USA
   61,977 news articles posted by 12,495 users
   News articles comes from 37 news sites
METHODOLOGY

1. Facebook news consumption
   From a list of 100 news papers in USA
   61,977 news articles posted by 12,495 users
   News articles comes from 37 news sites

2. Determining media slant
   Classify news outlets into liberal, conservative, or center
   http://mondotimes.com & Shapiro’s classification
METHODOLOGY

1. Facebook news consumption
   From a list of 100 news papers in USA, 61,977 news articles posted by 12,495 users.
   News articles come from 37 news sites.

2. Determining media slant
   Classify news outlets into liberal, center, and conservative using http://mondotimes.com & Shapiro's classification.

<table>
<thead>
<tr>
<th>News Outlet</th>
<th>Liberal</th>
<th>Facebook</th>
<th>Conservative</th>
</tr>
</thead>
<tbody>
<tr>
<td>huffingtonpost</td>
<td>11,236</td>
<td>45.2%</td>
<td>3,774</td>
</tr>
<tr>
<td>nytimes</td>
<td>10,083</td>
<td>41.3%</td>
<td>2,767</td>
</tr>
<tr>
<td>msnbc.msn</td>
<td>4,892</td>
<td>30.7%</td>
<td>1,433</td>
</tr>
<tr>
<td>abcnews.go</td>
<td>2,752</td>
<td>23.2%</td>
<td>695</td>
</tr>
<tr>
<td>washingtonpost</td>
<td>2,468</td>
<td>58.3%</td>
<td>670</td>
</tr>
<tr>
<td>time</td>
<td>2,104</td>
<td>34.9%</td>
<td>565</td>
</tr>
<tr>
<td>cbsnews</td>
<td>1,868</td>
<td>22.8%</td>
<td>332</td>
</tr>
<tr>
<td>bostonglobe</td>
<td>1,260</td>
<td>29.3%</td>
<td>426</td>
</tr>
<tr>
<td>latimes</td>
<td>1,239</td>
<td>37.3%</td>
<td>288</td>
</tr>
<tr>
<td>salon</td>
<td>967</td>
<td>64.8%</td>
<td>239</td>
</tr>
<tr>
<td>slate</td>
<td>923</td>
<td>44.9%</td>
<td>118</td>
</tr>
<tr>
<td>sfgate</td>
<td>779</td>
<td>34.1%</td>
<td>130</td>
</tr>
<tr>
<td>wnd</td>
<td>441</td>
<td>72.1%</td>
<td>93</td>
</tr>
<tr>
<td>newyorker</td>
<td>368</td>
<td>33.9%</td>
<td>60</td>
</tr>
</tbody>
</table>
METHODOLOGY

1. Facebook news consumption
   From a list of 100 news papers in USA
   61,977 news articles posted by 12,495 users
   News articles comes from 37 news sites

2. Determining media slant
   Classify news outlets into liberal, conservative, or center
   http://mondotimes.com & Shapiro’s classification

3. Only news articles about politics
   Topic classification using Alchemy API
   12 topics: Arts Entertainment, Business, Computer Internet, Culture Politics,
   37% of news articles (22,929) are classified: 42% in Culture Politics
1. Facebook news consumption
   From a list of 100 news papers in USA
   61,977 news articles posted by 12,495 users
   News articles comes from 37 news sites

2. Determining media slant
   Classify news outlets into liberal, conservative, or center
   http://mondotimes.com & Shapiro’s classification

3. Only news articles about politics
   Topic classification using Alchemy API
   12 topics: Arts Entertainment, Business, Computer Internet, Culture Politics,
   37% of news articles (22,929) are classified: 42% in Culture Politics

4. Measuring partisanship: Net partisan skew
   \[ \ln(\text{#conservative news}) - \ln(\text{#liberal news}) \]
NEWS SHARING IN TWITTER

Build a VotingTime to recruit Twitter users
71 UK based Twitter users (1K political news articles)
METHODOLOGY

1. Twitter news consumption
   From a list of 100 news papers in USA
   5,714 news articles posted by 71 users
   News articles comes from 10 news sites

2. Determining media slant
   Classify news outlets into liberal, conservative, or center
   Manual coding by three UK political journalists (kappa = 0.92)

3. Only news articles about politics
   Topic classification using Alchemy API
   12 topics: Arts Entertainment, Business, Computer Internet, Culture Politics,
   17.6% in Culture Politics

4. Measuring partisanship: Net partisan skew
   $ln(#\text{conservative news}) - ln(#\text{liberal news})$
Does selective tendency exist in political news sharing?
PARTISAN SHARING (POLITICS)
PARTISAN SHARING (POLITICS)

Political News Sharing

Sharing Liberal news

Sharing conservative news

Density

4+ posts

8+ posts
PARTISAN SHARING (POLITICS)

Sharing Liberal news

Sharing conservative news

YES!
Users consume a considerable number of like-minded political news and avoid cross-cutting news.
PARTISAN SHARING (POLITICS)

YES!

Users consume a considerable number of like-minded political news and avoid cross-cutting news.
Does partisan sharing influence consumption of other type of news?
PARTISAN SHARING (ENTERTAINMENT)
NO!

When sharing entertainment news, people tend to select outlets regardless their political beliefs.
CHANGES ACROSS INDIVIDUALS

Does partisan sharing change depending on users’ characteristics? Does an individual net partisan skew depend on: 1) political leaning 2) amount of news consumption
CHANGES ACROSS INDIVIDUALS

Does partisan sharing change depending on users’ characteristics? Does an individual net partisan skew depend on:

1) political leaning  
2) amount of news consumption

\( t(166.54) = 5.805, \quad p < 0.0005 \)

**YES!** Conservatives share 43% less like-minded articles than liberals.
Does partisan sharing change depending on users’ characteristics? Does an individual net partisan skew depend on: 1) political leaning 2) amount of news consumption

YES! Conservatives share 43% less like-minded articles than liberals.

YES! As users share more news, they also share more partisan news.

(t(166.54) = 5.805, p < 0.0005)

\(r = 0.46\) (p < 0.001)
CHANGES OVER TIME

Is partisan sharing more or less prevalent in politically salient periods (e.g., during elections)?

Net partisanship skew is minimum in the election month and tends to increase to a stable point outside that period.

States with primary election in May and June.
VALIDATION WITH TWITTER DATA

Existence: Partisan sharing exists.

Changes across individual

1. Political leaning: Liberals tend to be more partisan (with net skew of 1.5).

2. News consumption: Those who share more are the ones with stronger partisanship.
SOCIAL CONSEQUENCES

Factors
- Political leaning
- Amount of political news consumption
- Time

Partisan sharing

Consequences
- Perceived bias of news outlets
- Political knowledge
- Voting probability
Will partisan sharing lead to polarized political attitudes?
PERCEIVED BIAS OF NEWS OUTLETS

Ask respondents to which extent they thought four news outlet – BBC (N), Telegraph (C), Guardian (L), The Sun (C) – were politically biased

('0' mean 'neutral' and '100' means 'strongly biased')
PERCEIVED BIAS OF NEWS OUTLETS

Liberal and conservative users significantly differ in their perceptions of the media’s leaning.
Ask respondent 4 small political knowledge quizzes about general UK political facts
POLITICAL KNOWLEDGE

Ask respondent 4 small political knowledge quizzes about general UK political facts

Those **politically knowledgeable** tended to be **more partisan** than those less knowledgeable.
POLITICAL KNOWLEDGE

Ask respondent 4 small political knowledge quizzes about general UK political facts

Those **politically knowledgeable** tended to be **more partisan** than those less knowledgeable.

VOTING PROBABILITY

**Average Net Partisan Skew**

<table>
<thead>
<tr>
<th># correct answers</th>
<th>Average Net Partisan Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=2</td>
<td><img src="image1" alt="Graph" /></td>
</tr>
<tr>
<td>&gt;2</td>
<td><img src="image2" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Absolute Net Partisan Skew**

<table>
<thead>
<tr>
<th>Decision to vote for next election</th>
<th>Absolute Net Partisan Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decided</td>
<td><img src="image3" alt="Graph" /></td>
</tr>
<tr>
<td>Haven't decided</td>
<td><img src="image4" alt="Graph" /></td>
</tr>
</tbody>
</table>

\[t(4.558) = 4.566, \ p < 0.01\]

UK people who have decided whether to vote are **more partisan** than those who remain undecided.
Partisan sharing exists.
Can we predict an individual net partisan skew?
Partisan sharing exists.

Can we predict an individual net partisan skew?

- News filtering system
- Expose people to diverse information
Partisan sharing exists.
Can we predict an individual net partisan skew?

- News filtering system
- Expose people to diverse information

- Knowledgeable in politics
- Participating to political events
PREDICTING AN INDIVIDUAL NET PARTISAN SKEW

|Net partisan skew| Three Facebook variables: # of Facebook friends, # of postings, and # of likes

Three personal attributes: sex, age, and size of the city she lives in

Five personality traits (OCEAN): openness, conscientiousness, extraversion, agreeableness, neuroticism
Predicting political leaning is far easier than predicting partisanship, which appears to be quite challenging.

None of them was correlated for conservatives, while only sex was correlated for liberals.
PREDICTING AN INDIVIDUAL NET PARTISAN SKEW

with perceived bias of news outlets
BBC (N), Telegraph (C), Guardian (L), The Sun (C)

Ask respondents their partisanship
0 (Labour) – 25 (liberal) -- -- 75 (conservative) – 100 (BNP, UKIP)
PREDICTING AN INDIVIDUAL NET PARTISAN SKEW

with perceived bias of news outlets
BBC (N), Telegraph (C), Guardian (L), The Sun (C)

Ask respondents their partisanship
0 (Labour) – 25 (liberal) -- -- 75 (conservative) – 100 (BNP, UKIP)

Linear regression

coefficients:

partisanship = \alpha + \beta_1 \text{bias}_\text{BBC} + \beta_2 \text{bias}_\text{Telegraph} + \beta_3 \text{bias}_\text{Guardian} + \beta_4 \text{bias}_\text{TheSun}

R^2 = 0.44

The Guardian (bias\text{Guardian} = 0.62) and
group-leaning The Sun (bias\text{TheSun} = -0.61).
SUMMARY

Factors
- Political leaning
- Amount of political news consumption
- Time

Consequences
- Perceived bias of news outlets
- Political knowledge
- Voting probability

Partisan sharing

Factors lead to Partisan sharing, which in turn leads to the consequences listed.
Partisan sharing exist, but selectivity is limited to political news.
Summary

Factors
- Political leaning
- People who are interested in politics tend to have stronger partisanship
- Amount of political news consumption
- Time

Consequences
- Perceived bias of news outlets
- Political knowledge
- Voting probability

Partisan sharing exist, but selectivity is limited to political news
Factors

- Political leaning
- People who are interested in politics tend to have stronger partisanship
- Perceived bias of news outlets
- Political diversity increases during the election period.

Consequences

- Partisan sharing exist, but selectivity is limited to political news
- Perceived bias of news outlets
- Political knowledge
- Voting probability

SUMMARY

People who are interested in politics tend to have stronger partisanship, but selectivity is limited to political news. Political diversity increases during the election period.
SUMMARY

Factors:
- Political leaning
- People who are interested in politics tend to have stronger partisanship
- Political diversity increases during the election period.

Partisan sharing exist, but selectivity is limited to political news.

Consequences:
- Negative: related to distorted perceptions of media bias
- Political knowledge
- Voting probability
SUMMARY

**Factors**
- Political leaning
  - People who are interested in politics tend to have stronger partisanship
- Political diversity increases during the election period

**Partisan sharing exist, but selectivity is limited to political news**

**Consequences**
- **Positive:** associated with people who are knowledgeable about politics and are actively engaged in political life
- **Negative:** related to distorted perceptions of media bias

**Factors**
- Amount of political news consumption
- Political leaning
- Partisan sharing
- Perceived bias of news outlets
- Political knowledge
- Voting probability

**Consequences**
- Positive: associated with people who are knowledgeable about politics and are actively engaged in political life
- Negative: related to distorted perceptions of media bias
Partisan Sharing: Facebook Evidence and Societal Consequences

Factors:
- Political leaning: People who are interested in politics tend to have stronger partisanship.
- Political diversity increases during the election period.

Consequences:
- Positive: associated with people who are knowledgeable about politics and are actively engaged in political life.
- Negative: related to distorted perceptions of media bias.

Partisan sharing exist, but selectivity is limited to political news.
Measurement and Analysis of OSN Ad Auctions

Yabing Liu†  Chloe Kliman-Silver‡
Robert Bell§  Balachander Krishnamurthy§
Alan Mislove†

†Northeastern University
§AT&T Labs–Research
‡Brown University
Online advertising networks are everywhere
Google earned over $50 billion in advertising in 2013

How is online advertising implemented?
Through Auctions!
Advertisers pick keywords, search terms and bid on ads
Advertising networks select the winning bidders and present ads to users
Two ways to pay
  CPM: Cost Per Mille, the cost of 1,000 ad impressions
  CPC: Cost Per Click
Motivation

The new OSN-based ad services became **popular**

Facebook had over **$7.8 billion** in advertising in **2013**

Ask users to fill in their information

**Significant** data about the users

- Personal information (demographics, interests, educational history, relationship status, etc)
- Identities of friends
- User activity

Target **users** directly (not keywords, or search terms)

- Atlas to serve ads on **non-OSN sites**
- across **multiple devices**
What has been studied?

Web-search-based advertising networks

"Estimated prices" from Google's Traffic Estimator Tool [Manage.Sci.'11]
Analytical models to predict the clicks, prices, CTR [WWW'14]
New models for conducting online auctions [EC'12]

User Value

Influential users in OSNs [EC'12]
The contribution of users to advertising revenue is skewed [IMC'13]
65% of ad categories received by users are targeting interests [HotNets'13]

Unfortunately

Little academic study of the OSN-based ad networks
OSNs have released little data about their advertising markets
This paper

Goal

Develop techniques to **measure and understand** OSN ad markets
Bring **visibility** to OSN ad markets, focusing on **Facebook**
Research problem **meaningful** for advertisers, users, and other researchers

Assumption

No current **tool** to measure Facebook ad market
No **visibility** to Facebook internal system (as external researchers)
Outline

Motivation

Exploring suggested bid mechanism

How are suggested bids calculated?

Exploring user value
Facebook advertising

AUDIENCE

Locations
- United States
  - All United States

Age
- 13 - No max

Gender
- All
- Men
- Women

Languages
- Enter a language...

More Demographics

Audience Definition
- Specific
- Broad
- Your audience selection is fairly broad.

Potential Reach: 82,000,000 people

Your ad targets people:
- Who live in United States
- Who are male

Suggested Bid
- $0.04–$1.23 USD
Facebook's targeting parameters

<table>
<thead>
<tr>
<th>Basic Fields</th>
<th>Parameters/Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Country, State, City, Postal code</td>
</tr>
<tr>
<td>Gender</td>
<td>Male, Female, All</td>
</tr>
<tr>
<td>Age</td>
<td>Range (from 13–65)</td>
</tr>
<tr>
<td>Precise Interest</td>
<td>Travel, Science, Music, ...</td>
</tr>
<tr>
<td>Broad Category</td>
<td>Cooking, Gardening, iPhone 5, ...</td>
</tr>
<tr>
<td>Interested In</td>
<td>Male, Female, All</td>
</tr>
<tr>
<td>Relationship Status</td>
<td>All, Single, In a relationship, Married, Engaged, Not specified</td>
</tr>
<tr>
<td>Language</td>
<td>English, Spanish, French, ...</td>
</tr>
<tr>
<td>Education</td>
<td>Anyone, In high school, In College, College Grad</td>
</tr>
<tr>
<td>Workplaces</td>
<td>Google, Facebook, AT&amp;T, ...</td>
</tr>
</tbody>
</table>

Notes

Target any combination of these parameters
Required to specify at least one country
Facebook advertising

![Audience settings for Facebook advertising](image)
Methodology

What are suggested bids?

Facebook undocumented feature

"The suggested bid range you see when creating your ads is based on the bids that are currently winning the ad auction for the users you've chosen to target."

How to collect suggested bids in scale?

Programmatically send HTTP GET requests to the Facebook Ad Creation URL

Query:

```
https://graph.facebook.com/reachestimate?targeting_spec=
{"countries":["US"],"age_min":21,"age_max":30,genders=[1]}&currency=USD&accountId=XXX&access_token=XXXX
```

Response:

```
{"data": {"users":62984500,"bid_estimations":
{"location":3,"cpc_min":54,"cpc_median":82,"cpc_max":144,
"cpm_min":3,"cpm_median":14,"cpm_max":83}}}
```
Suggested bid data

Example Dataset

1,000 suggested bids
Each of the 204 countries that Facebook supports
Queries were roughly spaced 35 milliseconds apart
U.S.: 159M; New Zealand: 2.2M; Antigua and Barbuda: 29K users

CPM ad prices
reasoning about CPC requires knowing an advertiser's CTR
CTR (click-through rate): the fraction of impressions that result in a click
1 Skewed distribution
Suggested bid observations

1 Skewed distribution

2 Significant variance

Antigua and Barbuda (29,580 users)
Suggested bid observations

1 Skewed distribution

2 Significant variance

3 Variance independent of audience size

204 Countries
Suggested bid observations

1 **Skewed distribution**

2 **Significant variance**

3 **Variance independent of audience size**

4 **Variance across accounts**
Suggested bid observations

1 Skewed distribution

2 Significant variance

3 Variance independent of audience size

4 Variance across accounts

5 Non-persistence of min or max
Outline

Motivation

Exploring suggested bid mechanism

How are suggested bids calculated?

Exploring user value
Reverse-engineering suggested bids

Goal:
The suggested bid algorithm is a black box
Look for the most reasonable explanation

Hypothesis 1: Winning bids change rapidly
Derived from the most-recent-k winning bids for the target users

Not true: significant variance observed on very short timescales
Reverse-engineering suggested bids

Hypothesis 2: Adding random noise

In order to obfuscate the true value

Statistical tests: if the data matched a number of common statistical distributions (Uniform random, Gaussian, Cauchy, Log-Normal or Logistic)

Example probability distribution function of CPM max values for 20,000 suggested bids.

Fails statistical tests.

Not true: poor fit for all distributions, with a p-value of less than $10^{-16}$
Reverse-engineering suggested bids

Hypothesis 3: Sampling winning bids
Sampling from the recent-k winning bids
Reporting the min, median, and max of the sample

Logical mechanism for calculating suggested bids
Consistent with all the five properties of suggested bids

Suggested bid is most likely sampled from the recent-k winning bids
How real auctions affect suggested bids?

Changes to the market

Actively participate in the advertising market

See how quickly we can affect the ad market

Chose a small country (Seychelles, 26K users) with low suggested CPM

Bid a higher CPM ($1.00) than the suggested CPM max ($0.16) from 3 accounts

Ran campaigns concurrently for 8 hours

Changes to the ad market are reflected in the suggested bids.
How suggested bids correlate with revenue?

Comparison with Facebook's revenue

The ground truth: Facebook's SEC filings
Average Revenue Per User (ARPU) at the granularity of regions
Aggregate our CPM median data into the same regions

<table>
<thead>
<tr>
<th>Region</th>
<th>Facebook ARPU</th>
<th>Suggested Bid</th>
</tr>
</thead>
<tbody>
<tr>
<td>US, Canada</td>
<td>$3.50</td>
<td>1.00</td>
</tr>
<tr>
<td>Europe</td>
<td>$1.60</td>
<td>0.45</td>
</tr>
<tr>
<td>Asia</td>
<td>$0.64</td>
<td>0.18</td>
</tr>
<tr>
<td>Rest of World</td>
<td>$0.50</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Notes

Rank the regions in the same order
Europe and Rest of the World regions at approximately the same ratios

The suggested bid data at least correlates with the distribution of Facebook's revenue.
How researchers use suggested bids

Suggested bid data is most likely calculated by sampling from the recent winning bids for the target users.

Multiple samples

- Extract the overall min, median, and max from the collated samples

Convergence

- How many samples to collate together?

We use 25 collated suggested bids
Outline

Motivation

Exploring suggested bid mechanism

How are suggested bids calculated?

Exploring user value
Location

How the location influences the ad auction winning bids?

GDP (Gross Domestic Product) per capita for 204 countries

The output of a country's economy per person

![Graph Showing GDP vs CPM for Different Countries]

Note

Observed a correlation of 0.37 between the GDP per capita and CPM values

Dramatic differences in ad auction prices across different countries
Age

How is CPM median price correlated with user age?

Select the same three countries (US, NZ, AG)

The smallest age is 13, while targeting age 65 encompasses all users 65 and over.

Notes

For U.S. and NZ, as age increases, the CPM median price increase as well.

Less clear trend for AG.
How is CPM median price correlated with user age?

Select the same three countries (US, NZ, AG)

The smallest age is 13, while targeting age 65 encompasses all users 65 and over.

Notes

For U.S. and NZ, as age increases, the CPM median price increase as well.

Less clear trend for AG

Less differences in ad auction prices across different ages
Price stability

How stable are the prices for different target demographics over time?
Select four different sets of targeting parameters
Retrieve 25 suggested bids each hour for a period of 3 weeks (Apr. 3~23, 2013)

Notes
G1 shows a periodic increase per week
G2 and G3 shows a multi-day increase starting on 04/16
G4 does not vary much over the study period
Significant long-term dynamics present in Facebook's ad auctions.
Summary

Identify the suggested bid mechanism to measure Facebook ad market

Validate and show how researchers can use the suggested bid data

Analyze how different users contribute to Facebook's revenue
  
  Dramatic differences in ad auction prices across different locations, interest
  
  Significant long-term dynamics present in Facebook's ad network
Questions?

Our suggested bid collection code and collected data available at http://osn-ads.ccs.neu.edu