Extending the Demography of Sexuality with Digital Trace Data

Connor Gilroy\textsuperscript{1}  Ridhi Kashyap\textsuperscript{2}

IUSSP Workshop on Demographic Research in the Digital Age
April 25, 2018
Denver, CO

\textsuperscript{1}Department of Sociology
University of Washington

\textsuperscript{2}Department of Sociology and Nuffield College
University of Oxford
The problem

We lack estimates of sexuality at the level of granularity we might want
The problem

Because of stigma and selectivity, we can only measure disclosure, not underlying prevalence.
The problem

ACS: same-sex couples, not sexual orientation

GSS & NSFG: too small for subnational estimates

Gallup: limited data, low response rate

England, Mishel, and Caudillo 2016
Hatzenbuehler, Flores, and Gates 2017
Schnabel 2018
Digital trace data: a solution?
Digital trace data

Facebook

- large N—210 million in US
- detailed data about users
### Basic Information

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth Date</td>
<td></td>
</tr>
<tr>
<td>Birth Year</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
</tr>
<tr>
<td>Interested In</td>
<td>Men</td>
</tr>
</tbody>
</table>
Facebook ads

Facebook users see targeted ads, like this one: Want this delivered? We got you. (The ice cream, not the guy.)
INCLUDE people who match at least ONE of the following:

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Relationship</th>
<th>Interested In</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Add demographics, interests or behaviors | Suggestions | Browse

**Audience Size**

Your audience selection is fairly broad.

Specific | Broad

Potential Reach: 25,000 people
Researchers have studied health, migration, and gender.

We extend the method to sexuality.

Araujo et al 2017
Zagheni, Weber, and Gummadi 2017
Fatehkia, Kashyap, and Weber 2018
Overview of data

Interested in:

• men
• women
• men and women
• not specified

cross-tabbed by gender, age, location, relationship status

(collected Sept 2017)
Overview of data

Sexuality

- interested in same gender: 0.8%
- interested in men and women: 1.2%
- interested in different gender: 26.9%
- not specified: 71.1%
Overview of data

Sexuality

- Interested in same gender: 2.9%
- Interested in men and women: 4%
- Interested in different gender: 93.1%
4.2 million Facebook users disclose as LGB

compared with 10 million LGBT Americans, estimated by Gallup 2016
Disclosed prevalence of LGB Facebook users
For what groups might disclosure be a good measure of prevalence?
Disclosure by age

LGB Facebook users

disclosed prevalence

age (18-64)
Disclosure by age

LGB Facebook users

women

men

disclosed prevalence

age (18-64)
Disclosure by age

LGB Facebook users

women

men

Age (18-64)

Disclosed prevalence

LG

B
Disclosure by age

non-LGB Facebook users

women

men

prevalence

age (18-64)

straight not specified
Conclusions

Particularly for younger ages, Facebook disclosure is a good proxy for prevalence.

Points to differences in prevalence by gender and sexual identity.
Conclusions

Aggregate Facebook estimates, with their scale and granularity, are a good complement to existing data sources about sexuality.
Thank you!

✉ cgilroy@uw.edu
🔗 @ccgilroy
🔗 ccgilroy

Partial support for this research came from a National Institutes of Health Big Data to Knowledge (BD2K) training grant, 3 T32 HD 7543-15 S1, to the Center for Studies in Demography & Ecology at the University of Washington.
Appendix
Privacy and targeting: information settings

Your information

About you

Your categories

Some of the ads you see are because advertisers are trying to reach people based on information they've provided on their profiles. Manage whether we can show you ads intended to reach people based on these profile fields.

Relationship status

Employer

Job title

Education

Interested in

These settings only affect how we determine whether to show certain ads to you. They don't change which information is visible on your profile or who can see it. We may still add you to categories related to these fields (see Your categories above).
Privacy and targeting: inferred interests

Choose an interest to preview examples of ads you might see on Facebook or remove it from your ad preferences.
Overview of data

Gender

women: 109M
men: 100M
Overview of data

Age distribution by age group:
- 18-24: 20.7%
- 25-34: 28.7%
- 35-44: 20.6%
- 45-54: 17%
- 55-64: 13%
How does Facebook compare to Gallup 2016?

(at state-level geography)
Validation

Facebook and Gallup

$L^2 : 0.49$

(District of Columbia removed as an outlier)
How does Facebook compare to the American Community Survey?

(at state-level geography)
Validation

Facebook and the American Community Survey

- Women: $R^2 : 0.38$
- Men: $R^2 : 0.46$
- Bisexual: $R^2 : 0.21$
- Lesbian/gay: $R^2 : 0.70$

(District of Columbia removed as an outlier for men)
DC and Facebook estimates: Gallup

Facebook and Gallup

$R^2: 0.57$
DC and Facebook estimates: ACS

Facebook and the American Community Survey

- Women: $R^2 : 0.39$
- Men: $R^2 : 0.49$
- Bisexual: $R^2 : 0.24$
- Lesbian/gay: $R^2 : 0.78$
Relationship statuses

- single: 18.9%
- in a relationship: 10.3%
- engaged: 2.4%
- married: 26.6%
- divorced: 0.7%
- widowed: 0.4%
- not specified: 39.4%

(all other categories: 1.4%)
### Relationship statuses

#### Sexuality by relationship status

<table>
<thead>
<tr>
<th>Status</th>
<th>Single</th>
<th>In a Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>interested in same gender</td>
<td>1.7%</td>
<td>1.6%</td>
</tr>
<tr>
<td>interested in men and women</td>
<td>1.9%</td>
<td>2%</td>
</tr>
<tr>
<td>interested in different gender</td>
<td>55.9%</td>
<td>49.1%</td>
</tr>
<tr>
<td>not specified</td>
<td>40.5%</td>
<td>47.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Status</th>
<th>Married</th>
<th>Not specified</th>
</tr>
</thead>
<tbody>
<tr>
<td>interested in same gender</td>
<td>0.6%</td>
<td>0.4%</td>
</tr>
<tr>
<td>interested in men and women</td>
<td>1.1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>interested in different gender</td>
<td>21.1%</td>
<td>9.7%</td>
</tr>
<tr>
<td>not specified</td>
<td>77.2%</td>
<td>89.4%</td>
</tr>
</tbody>
</table>