Online Footprints of Family Change
A Study Based on Twitter

Nicolò Cavalli | Nuffield College, University of Oxford, United Kingdom
Francesco C. Billari | Bocconi University, Milan, Italy
Eric Qian | Duke University and University of North Carolina, Chapel Hill, NC, USA
Ingmar Weber | Qatar Computing Research Institute, Doha, Qatar

WITTGENSTEIN CENTRE CONFERENCE
3rd Human Fertility Database Symposium
Fertility across time and space: Data and research ad
5-7 December 2018

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement n° 694262), project DisCont - Discontinuities in Household and Family Formation.
Fertility decline, emergence of non-standard life courses and family forms

This is a theory about culture shifts and ideational change

Main Intuition → Decoupling between parenthood and traditional (gendered) family roles
The Presentation of Self in Everyday Online Life

• Updating Goffman in the era of “network connectivity”

• Fluidity between social and digital lives/identities

• Expectation of accuracy and its challenges

1956

2017
• 300M monthly active users worldwide doing this:

Why Twitter?
Using Twitter to capture family change

- We exploit the website Twiangulate (through API) to explore the Twitter user space

- Twiangulate allows to use boolean operators to implement complex queries on Twitter users’ bios

- We single out individual users whose include keywords (mother, father, wife, husband and related colloquialisms), by location (US states)
A look at the code

First dataset retrieved on June 14th, 2016

Since June 15th, 2017 we have run the code monthly

Here we present a dataset spanning from June to October 2017

```
"keyword" AND("US state postal code") OR("full US state name")
```

```r

```
How does our code work?

- Crawls through the (generally stable) bio section (left hand side) and searches for correspondences with our keywords.. Here we have a match!
What does the output look like?
Dealing with false positives

James Read
@IvorBDick

Professional Sugar Daddy ....seeks Sugar Baby .... Business interests in Russia Ukraine NYC LA Vegas and London.Homes in NYC Spain and London....
Filtering

• 1 Step (Scraping): We obtain a dataset of 5.4M

• 2 Step (Filtering): Clean by location and individual description (*only partially automated*)

• 3 Step (Qualitative check): Look at the resulting dataset and identify additional false positives/negatives

→ Final dataset of 2.5M observations (500K users/month)
Descriptives
(1)

<table>
<thead>
<tr>
<th>term_id</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>dad</td>
<td>212,462</td>
<td>8.26</td>
<td>8.26</td>
</tr>
<tr>
<td>father</td>
<td>369,003</td>
<td>14.34</td>
<td>22.60</td>
</tr>
<tr>
<td>husband</td>
<td>407,759</td>
<td>15.85</td>
<td>38.45</td>
</tr>
<tr>
<td>mom</td>
<td>706,194</td>
<td>27.45</td>
<td>65.91</td>
</tr>
<tr>
<td>mother</td>
<td>382,398</td>
<td>14.86</td>
<td>80.77</td>
</tr>
<tr>
<td>wife</td>
<td>494,689</td>
<td>19.23</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>2,572,505</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

- twice as many mothers as fathers
- comparable sizes for husbands vs. wives
**Descriptives (2)**

1. \( \text{PrevalenceFather}_j = \frac{\text{#"Father"}_j}{\text{#"Husband"}_j} \)

2. \( \text{PrevalenceMother}_j = \frac{\text{#"Mother"}_j}{\text{#"Wife"}_j} \)

3. \( \text{PrevalenceParent}_j = \frac{\text{#"Father"}_j + \text{#"Mother"}_j}{\text{#"Husband"}_j + \text{#"Wife"}_j} \)
Results (3) *Twitter prevalences and SDT measures*

- Percent non-Hispanic white women 25–29 without children in household, 2000
- Percent non-Hispanic white women 25–29 never married, 2000
- Percent non-Hispanic white ever-married women 25–29 without own children in household, 2000
- Abortion rate per 1,000 live births, 1992
- Percent non-Hispanic white women 30–34 never married, 2000
- Abortion rate per 1,000 women 15–44, 1996
- Non-Hispanic white fertility postponement ratio (fert. 30+/fert. 20–29), 2002
- Same-sex households per 1,000 households, 2000
- Non-Hispanic white total fertility rate, 2002
- Non-Hispanic white fertility rate 15–19, 2002
- Percent households that are “families,” 1990
- Percent households with cohabiters of same or different sex, 2000
- Divorce rate per 1,000 population, 1990
- Total fertility rate, all races, 2002
- Percent nonmarital births, 1990
- Percent births to teenagers, 1986
- Divorce rate per 1,000 population, 1962
- Percent population 30+ living with and responsible for grandchildren, 2000
- Percent nonmarital births, 2000

---

**Scores for component 1**

- parent_ratio
- Fitted values

R-squared=.504; Corr=.710
Operationalizing cultural change

(1) How do identities change as a function of the SDT?

Tim Flowers
@Tim_Flowers
Christ Follower, Husband, Worship Leader and musician

joanna scott
@JozieLocks
Mommy; Blogger; Creator of Mocha Parents Awesome Kids; Racial Equity Consultant; Educator; Researcher; Believer in the Brilliance of Black Children.

mochaparents.com

Joined August 2009

Joined April 2009
Operationalizing cultural change

Cultural correlates for mothers

(2) How do identities change as a function of the SDT?
Operationalizing cultural change

**Weakening** of the probability of observing a religious mother as a function of the SDT

**Strengthening** of the probability of observing a mother working in sectors filled by men as a function of the SDT

(1) Employing self-descriptions as “cultural items” in quantitative analyses

\[ E(Y|Mother=1) = \text{Constant} + \text{SDT} + \text{FE} + \text{error} ; \text{observations} = 1,088,592 \]
Operationalizing cultural change

(2) Employing self-descriptions as “cultural items” in quantitative analyses

Increased probability of observing a feminist parent as the SDT unfolds (across genders!)

\[ E(Y|Mother=1) = \text{Constant} + \text{SDT} + FE + \text{error} \; ; \; \text{observations} = 1,670,057 \]
“We focus especially on the endogenous dynamics of the process, which, we argue, depend on the conditions that favor rapid diffusion. Our core argument is that the turnaround is driven by the diffusion of gender-equalitarian norms”.
Mechanisms

- Role of networks
- How to estimate diffusion on twitter?
  I. Spread of cultural correlates
  II. Demographics of new users and user filiation (e.g. analyzing identities of new followers of existing users)

<table>
<thead>
<tr>
<th>Average # followers</th>
<th>Low SDT</th>
<th>High SDT</th>
<th>first_diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>fathers</td>
<td>1328.740</td>
<td>1916.046</td>
<td>587.306</td>
</tr>
<tr>
<td>husbands</td>
<td>1161.823</td>
<td>1813.780</td>
<td>651.957</td>
</tr>
<tr>
<td>mothers</td>
<td>718.5506</td>
<td>1326.631</td>
<td>608.080</td>
</tr>
<tr>
<td>wives</td>
<td>743.0628</td>
<td>1508.725</td>
<td>765.662</td>
</tr>
</tbody>
</table>
Closing remarks (Recap)

- We claim that web-data may be useful in analysing fast, culture-based demographic change without imposing strong a priori theory (e.g. through survey questions).
- We use Twitter to track differences in the unrolling of the SDT within the United States.
- We show that Twitter-based prevalences related to “new” vs “traditional” family role are statistically correlated to complex, low-frequency measures of demographic change.
- We propose a way to operationalise and assess cultural change via quantitative text analysis to show how the enrolling of the SDT correlates to different concepts of parenthood.
Closing remarks 

(Next steps)

• By lengthening our time-series, we can better assess the model against repeated SDT measures, clear out regional fixed effects, and also forecast

• Generate ways to smooth the filtering of false positives and effectively minimize false negatives

• Machine learning for cultural analysis, mining networks and applying demographic methods to their evolution

• Potential for cross-platform validation (Google correlates, Facebook ads)

• Extend the project at a global scale, while broadening the scope in the identification of cultural change (behaviours, sexual orientations, etc)