SOCIAL MEDIA AS PASSIVE POLLING: USING TWITTER TO MAP ISLAMIST SENTIMENT IN EGYPT

Todd Mostak
Harvard University
TWITTER AND THE NEW EGYPT

• According to the World Bank, 20.9% of Egyptians used the internet in 2009
• Internet social media played an oversized role in recent revolution
• But what about Twitter as a data source?
  • Idea: We can use what people say on Twitter to quantitatively estimate their positions on a number of issues
• How about testing long-standing debate on whether Islamism is associated with poverty?
THE DATA

- 19.5 million Egyptian tweets from October 2011 to present.
- All posts from ikhwan.net (Arabic-language forum affiliated with Muslim Brotherhood) from January 2004 to present.
- Data collected with Python scripts and placed in PostgreSQL database.
- 2005 Egyptian census data, aggregated by village district (shiyakha).
WHAT ARE PEOPLE SAYING ON TWITTER?

- We are praise God Muslims and united in our faith. Let rule those who protect our beautiful country from the extremists that exploit religion.

- I agree with you... if liberalism had every truly been adopted a foreign state then that state would have protected the rights of its Muslims


- The revolution against Mubarak happened on January 25, 2011. January 25, 2012 is the revolution against military rule. Someone warn the (Muslim) Brotherhood that January 25, 2013 is fast approaching.
METHODOLOGY

• Use Information Retrieval (IR) methods to quantify “discourse”
• Generate a baseline corpus to model Islamist modes-of-discourse
  • Forum posts taken from Ikhwan.net
  • Document-term matrix generated from these posts over same time period as tweet data
• Tweets aggregated for Egyptian twitter users over a period of 6 months
• Non-Arabic removed
• Arabic is stemmed to reduce noise
  • Necessary due to highly inflected nature of language
• A Term-Frequency / Inverse Document Frequency (TF/IDF) vector is computed for each Twitter user, which is compared to baseline ‘Islamist’ corpus using cosine similarity measure
• The Islamist scores for Twitter users are aggregated by the administrative districts they are tweeting from
• The district scores are then regressed against census data for those districts
PRELIMINARY RESULTS

• Islamist sentiment seems to be positively correlated with male unemployment, illiteracy, and percentage of land used in agriculture and negatively correlated with percentage of men in their youth aged 15-25 (the shabab)
• Note that female variables for unemployment and age were statistically insignificant
• Caveats with weighting, some census variables unreliable?

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>.827323789</td>
<td>4</td>
<td>.086830847</td>
</tr>
<tr>
<td>Residual</td>
<td>.057238214</td>
<td>711</td>
<td>.0000008504</td>
</tr>
<tr>
<td>Total</td>
<td>.084562083</td>
<td>715</td>
<td>.000118269</td>
</tr>
</tbody>
</table>

Number of obs = 716
F( 4, 711) = 84.85
Prob > F = 0.0000
R-squared = 0.3231
Adj R-squared = 0.3193
Root MSE = 0.0897

| isl_scr   | Coef.  | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|-----------|--------|-----------|-------|-------|---------------------|
| pct_male_unemployed | 0.0018827 | 0.000229 | 8.22  | 0.000 | 0.0014331 - 0.0023324 |
| pct_illiterate       | 0.0001562 | 0.0000475 | 3.33  | 0.001 | 0.0000649 - 0.0002515 |
| pct_male_15_to_25    | -0.0028666 | 0.0001969 | -14.53 | 0.000 | -0.0032471 - -0.002474 |
| pct_ag               | 0.0005407 | 0.0001499 | 3.61  | 0.000 | 0.000264 - 0.0008351 |
| _cons                | 0.1712246 | 0.0057751 | 29.65 | 0.000 | 0.1598862 - 0.1825629 |
GPU-RENDERED TIME-LAPSE HEATMAPS