

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

Todd Mostak  
Harvard University

# TWITTER AND THE NEW EGYPT

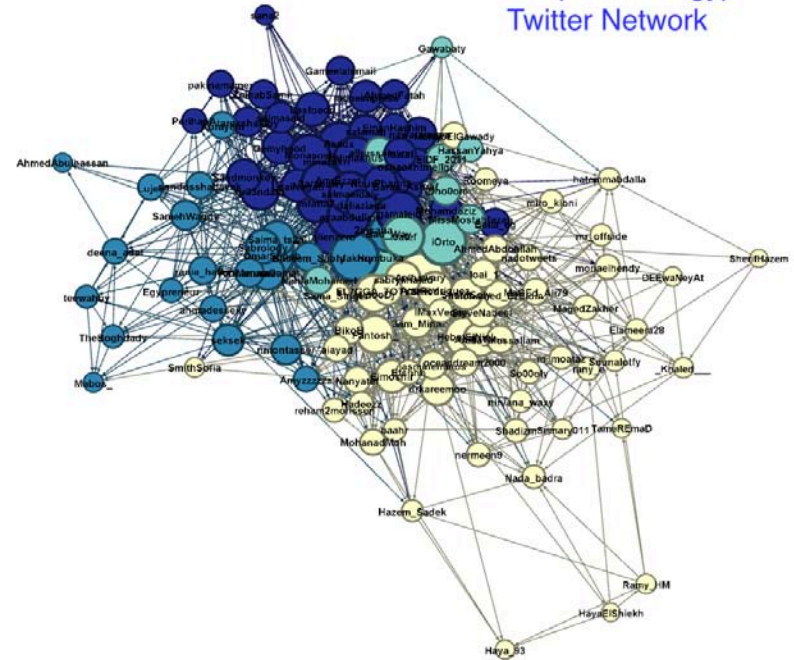
- According to the World Bank, 20.9% of Egyptians used the internet in 2009
- Internet social media played an oversized role in recent revolution
- But what about Twitter as a data source?
  - Idea: We can use what people say on Twitter to quantitatively estimate their positions on a number of issues
- 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)

Snapshot of Egypt's  
Twitter Network





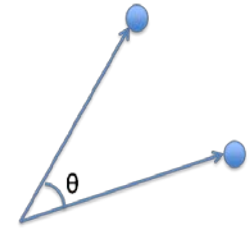


# METHODOLOGY (CONT.)

- 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

	row.names	رسالت	شباب	ثور	يناير	ايناي	اشد	ارض	ا
1	268190502	0	29	256	24	0	1	9	1
2	265376535	0	35	84	8	0	1	10	5
3	281983837	0	59	295	42	0	2	19	1
4	319840031	0	11	38	2	1	0	12	8
5	72326479	0	54	226	15	2	0	7	8
6	154241412	0	11	23	5	1	1	10	8
7	293993610	0	1	4	1	0	0	1	1
8	217841998	0	65	126	27	0	0	20	1
9	45831035	0	20	133	16	0	3	19	3
10	44336273	0	11	129	8	1	1	36	3
11	71792261	0	66	137	15	0	3	14	8
12	342449796	2	30	0	0	0	1	7	8
13	42196320	2	178	579	68	4	6	37	1
14	336767829	1	83	304	9	0	1	15	1
15	130273505	0	28	94	9	0	2	4	1
16	14315020	1	59	401	18	0	2	23	1
17	279890284	0	169	415	36	1	6	84	2
18	45101987	0	27	43	8	0	2	7	1
19	18633523	0	0	0	0	0	0	2	6
20	104191305	0	78	190	15	1	0	21	1

$$sim(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$





# 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?

Source	SS	df	MS			
Model	.027323789	4	.006830947	Number of obs =	716	
Residual	.057238214	711	.000080504	F( 4, 711) =	84.85	
Total	.084562003	715	.000118269	Prob > F =	0.0000	
				R-squared =	0.3231	
				Adj R-squared =	0.3193	
				Root MSE =	.00897	

isl_scr	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pct_male_unemployed	.0018827	.000229	8.22	0.000	.0014331	.0023324
pct_illiterate	.0001582	.0000475	3.33	0.001	.0000649	.0002515
pct_male_15_to_25	-.0028606	.0001969	-14.53	0.000	-.0032471	-.002474
pct_ag	.0005407	.0001499	3.61	0.000	.0002464	.0008351
_cons	.1712246	.0057751	29.65	0.000	.1598862	.1825629

