Persian Twitter: A Transforming Social Media Landscape

Emad Khazraee, Ph.D.
Berkman Klein Center for Internet & Society, Harvard University
School of Information, Kent State University
Outline

- Background
- Why Cartography?
- Comparing Persian Twitterverse between two elections 2013, 2017
  - Analysis of Information Diffusion Networks
  - Mapping Political Landscape
- Conclusion and future work
June 2009: Tehran

WHERE IS MY VOTE?!
Background: Twitter Revolution?

Background: Twitter Revolution?
Cartography
Evolving repertoires: Digital media use in contentious politics

Emad Khazraee
Kent State University, USA

James Losey
Stockholm University, Sweden

Abstract
The spread of the Internet coupled with knowledgeable users has led to the use of digital media as a tool for advocacy and activism. Building on theoretical foundations of eventful histories and digital formations, this article investigates the interrelated nature of contentious politics and digital technologies. Our analysis documents the eventful history of changing digital repertoires of contention in the context of messaging, blogging, and social networking sites in Iran. We argue that investigating single moments of protest offers only snapshots of how digital technologies are used in contentious politics, and entails the risk of focusing on a single platform rather than the mosaic of online and offline repertoires. We demonstrate that documenting event histories challenges the assumptions of the emancipatory nature of a specific technology by revealing the changing efficacy of repertoires during different moments of contention; therefore, we should avoid assigning stable causal relations between digital technologies and the democratization processes of societies.

Keywords
Collective action, contentious politics, digital formation, digital repertoires, eventful history

Introduction
During the 2000s, the Iranian blogosphere developed as a national and international platform for vibrant discussion and political dissent. “Blogestan,” as it is often called, emerged as a leading platform for contentious politics. Although this communicative space has declined in recent years, blogging is one of several platforms for challenging state power over information in Iran. For example, Blogestan offered a primary platform for contentious issues such as feminism in Iran and discussions have dispersed and grown on other platforms such as Facebook (Faris & Rahimi, 2015; Novak & Khazraee, 2015).
How to do that?

• A systematic program of study
  – Systematic data collection
  – Developing an understanding of context

• Developing tools and establishing methods for computational social science (digital humanities)
Goals

• Theoretical goals
  – Mapping Iran’s online public
  – Improve our understanding of political landscape of social media in Iran beyond the “Twitter Revolution”!!!
  – Fill the empirical gap
    • Repressive culture and restrictions
    • New technological developments

• Methodological goals
  – Developing computational tools to achieve above goals considering the scale of data
  – Find best methods for understanding different issues
Iran’s Twitter REvolution
Twitter in Iran

• It is not the most important platform
• Mostly used by younger and savvy demographics
• A good presence of elite (now)
• It is a good case to understand the impact of policy shifts in Iran
• Comparing Persian Twitter in 2013 & 2017
Objectives of the Case Study

• Understanding the information dissemination processes and power structure
  – Using social network analysis

• Understanding the political landscape
  – Using machine learning

• Develop tools to work on Persian text
2013 Case

• Data Collection
  – Using Twitter REST API
  – Total 47 keywords and hashtags
  – Data set cover tweets between May 14 and June 29 (Six weeks)

• Result
  • 3 Million Tweets (460K Persian tweets)
  • Studied Persian and English Networks
2017 Case

• Data Collection
  – Using Twitter Streaming API
  – Total 94 keywords and hashtags
  – Data set cover tweets between May 1 and May 25

• Result
  • 2.23 Million Tweets in Persian
    – 6 fold growth in comparison to the same period in 2103 (350K in 25 days)
    – Twitter global growth: 1.5 times
Overview of Network Analysis

Keyword search → Language ID → Different data sets → Data extraction → Processed tweets

Iran

Final results → Qualitative Analysis → SNA → Network data
Overview of Machine Learning

1. Keyword search
2. Language ID
3. Persian Tweets
4. Training tweets
5. Manual annotation
6. Annotated tweets
7. ML classifier
8. Processed tweets
9. Statistical analysis
10. Final results
Classifier Training and Data Processing

- 2013 Workflow
- Language ID
  - La-strings trained to recognize
- Classifier training
  - Stanford Classifiers
- Data processing workflow
  - Language identifier used to identify the language of each tweet
  - Identify subject and candidate(s) of Persian tweets
  - If the subject is politics and there is at least one candidate, detect sarcasm, sentiment, election participation, and political orientation

- 2017 Workflow (WIP)
- Language ID
  - Twitter Lang ID
  - fastText Lang ID
- Classifier training
  - fastText
- Data processing workflow
  - Language identifier used to identify the unknown tweets
  - Identify tweets with candidate(s) names
  - Detect sentiment towards the name entity
## Basic Stats

<table>
<thead>
<tr>
<th></th>
<th>2013*</th>
<th>2017</th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Tweets</td>
<td>351,223</td>
<td>2,236,023</td>
<td>537%</td>
</tr>
<tr>
<td>Number of Tweets with Links</td>
<td>81,846</td>
<td>1,083,730</td>
<td>1224%</td>
</tr>
<tr>
<td>Number of Retweets</td>
<td>81,554</td>
<td>1,240,670</td>
<td>1421%</td>
</tr>
<tr>
<td>Unique Users</td>
<td>17,236</td>
<td>80,395</td>
<td>366%</td>
</tr>
</tbody>
</table>

* Similar 25-day period
Most used platforms for Tweets

Platforms used for tweets
2013

Platforms used for tweets
2017
Distribution of Tweets 2017

**Total Tweets per Day**

**Total Tweets per Hour**
Information Diffusion

• We studied retweet network

• Identified the most influential (Network Centrality) and classified them

• Community detection
Classifying Most Influential Users

We classified the 100 most influential users (most retweeted and PageRank/eigenvector centrality) in four main categories:

• **Official news/media outlets:** (e.g. @cnnbrk: breaking news from CNN)

• **Journalists:** (e.g. @Gesfandiari: Radio Free Europe)

• **Politicians:** official twitter accounts of politicians

• **Social media (Twitter) celebrities:** Popular users who tweet from inside or outside Iran e.g., famous Iranian bloggers
Retweet network of Persian tweets 2013

Isolated MEK Community
Retweet network of Persian tweets 2017

MEK Community

Conservative Community

Reformist Community

Raminkarni1988

NiMaa62

Maryam_Rajavi_P

EvandHQ

Eshagh Jahanbakhsh

Shargh Daily

KhalatBaran

SaeidMehrazi

Eshaq_jahangiri

Mostafazadeh

Rouhani.ir

mah_sadeghi

election96
Transformation of Power Structure

Institutional Elites

Social Media Elites

2013  2017
Retweet network of Persian tweets 2017

MEK Community
Conservative Community
Reformist Community
The Follower Factory

Everyone wants to be popular online. Some even pay for it. Inside social media's black market.

By Nicholas Confessore, Gabriel J. Acosta, Richard Harris and Mark Hansen

Jan 27, 2019

Leave an answer
انتقاد شدید ایران از مواد مداختمویه، محراب و تنش‌های آمریکا

عضویت وزارت امور خارجه کشورمان در ارکشبا ادعاهای تکراری و مداختمویه ریپرایوری آمریکا در عربستان، این موضوع را مداختمویه، تکراری و متوالی از ادعاهای پیشنهادی علیه کشورمان ع...

isna.ir
Sentiment analysis: Distribution 2013

- Rouhani: 39%
- Jalili: 15%
- Qalibaf: 9%
- Aref: 9%
- Gharazi: 11%
- Rezaei: 7%
- Haddad Adel: 6%
- Velayati: 4%
Positive and Negative Sentiments

Reformist Candidates

Conservative Candidates

Aref
Rouhani
Qalibaf
Velayati
Rezaei
Jalili
Gharazi
Haddad Adel

0.00%
10.00%
20.00%
30.00%
40.00%
50.00%
60.00%
70.00%
80.00%

Positive
Negative

Positive
Negative

34
Positive and Negative Sentiments

Reformist Candidates

Conservative Candidates

Positive and Negative Sentiments

Rouhani

Jahangiri

Hashemitaba

Qalibaf

Raisi

Mirsalim

Positive and Negative
Conclusion

• Persian Twitter had a growth rate much faster than global trends
• The environment and practices are transforming
• The power structure has changed from social media celebrities to institutional elites
• Still predominantly in favor of reformists, but there is more critical perspectives
Future work

• Further exploration of text mining and machine learning for mapping the environment e.g., experimenting with Tweet2Vec

• Expanding training data and improving text mining results

• Investigating the role and the reach of automated accounts
Thank You!

Emad Khazraee
emad@kent.edu
@seakha
http://khazraee.com