OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author's version published in: [http://oatao.univ-toulouse.fr/22434](http://oatao.univ-toulouse.fr/22434)

**To cite this version:**

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr
POLITICS ON TWITTER: A PANORAMA

July 9, 2018

Ophélie FRAISIER
• **CONTEXT**

• **POLARISATION**

• **STANCE DETECTION**

• **ELECTION PREDICTION**

• **STUDY OF POLITICAL ENGAGEMENT**
CONTEXT
One of the biggest social media worldwide

- 2018: 336 million monthly active users
- Majority of data is public and easily accessible
One of the biggest social media worldwide

- 2018: 336 million monthly active users
- Majority of data is public and easily accessible

Twitter Revolution: How the Arab Spring Was Helped By Social Media

By Saleem Kassim | July 3, 2012
One of the biggest social media worldwide

- 2018: 336 million monthly active users
- Majority of data is public and easily accessible

In presidential campaign, Twitter was a powerful political tool

Twitter reports 1 billion election-related tweets since August 2015

By Sharon Gaudin
Senior Writer, Computerworld | Nov 8, 2016 11:32 AM PT
Twitter has emerged as the single most powerful “socioscope” available to social scientists for collecting fine-grained time-stamped records of human behavior and social interaction at the level of individual events.”

(Golder & Macy, 2014)
Social positioning of a person, a thoughtful positioning, justified by a set of values and beliefs, put in relation with the other existing points of view on the given subject.
POLITICAL STANCES?

- Twitter data has important limits:
POLITICAL STANCES?

- Twitter data has important limits:
  - Hardly quantifiable quality
POLITICAL STANCES?

- Twitter data has important limits:
  - Hardly quantifiable quality
  - Limited depth in terms of arguments
POLITICAL STANCES?

- Twitter data has important limits:
  - Hardly quantifiable quality
  - Limited depth in terms of arguments
    - 280 (140) characters
POLITICAL STANCES?

- Twitter data has important limits:
  - Hardly quantifiable quality
  - Limited depth in terms of arguments
    - 280 (140) characters
    - "Difficult" interactions
POLITICAL STANCES?

- Twitter data has important limits:
  - Hardly quantifiable quality
  - Limited depth in terms of arguments
    - 280 (140) characters
    - "Difficult" interactions
POLITICAL STANCES?

- Twitter data has important limits:
  - Hardly quantifiable quality
  - Limited depth in terms of arguments
    - 280 (140) characters
  - "Difficult" interactions
POLITICAL STANCES?

- Twitter data has important limits:
  - Hardly quantifiable quality
  - Limited depth in terms of arguments
    - 280 (140) characters
  - "Difficult" interactions

How relevant is it to use this data to study complex political topics?
POLITICAL STANCES?

- Public opinion characteristics according to Allport (1937):
P OLITICAL STANCES?

- Public opinion characteristics according to Allport (1937):
  - Verbalisations produced by many profiles on subjects important to them
POLITICAL STANCES?

- Public opinion characteristics according to Allport (1937):
  - Verbalisations produced by many profiles on subjects important to them
  - Subjects rooted in common culture
POLITICAL STANCES?

- Public opinion characteristics according to Allport (1937):
  - Verbalisations produced by many profiles on subjects important to them
  - Subjects rooted in common culture
  - Other profiles may react to these topics without necessarily being in direct contact
POLITICAL STANCES?

- Public opinion characteristics according to Allport (1937):
  - Verbalisations produced by many profiles on subjects important to them
  - Subjects rooted in common culture
  - Other profiles may react to these topics without necessarily being in direct contact
  - Aware that their behaviour can enable them to reach a goal
POLITICAL STANCES?

- Public opinion characteristics according to Allport (1937):
  - Verbalisations produced by many profiles on subjects important to them
  - Subjects rooted in common culture
  - Other profiles may react to these topics without necessarily being in direct contact
  - Aware that their behaviour can enable them to reach a goal
  - Component of interpersonal conflict when different stances
POLITICAL STANCES?

- Public opinion characteristics according to Allport (1937):
  - Verbalisations produced by many profiles on subjects important to them
  - Subjects rooted in common culture
  - Other profiles may react to these topics without necessarily being in direct contact
  - Aware that their behaviour can enable them to reach a goal
  - Component of interpersonal conflict when different stances

Twitter can be an useful medium for studying stances
POLARISATION
Homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people. […] Homophily implies that distance in terms of social characteristics translates into network distance, the number of relationships through which a piece of information must travel to connect two individuals."

(McPherson et al., 2001)
“Homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people. [...] Homophily implies that distance in terms of social characteristics translates into network distance, the number of relationships through which a piece of information must travel to connect two individuals.”

(McPherson et al., 2001)

- Can lead to "echo chambers"

(Sunstein, 2009)
INFLUENCE OF RETWEETS

- Retweet largely used
  - Action of sharing a tweet
  - One of the most important interaction on the platform
Politics on Twitter: A Panorama

Influence of Retweets

- Retweet largely used
  - Action of sharing a tweet
  - One of the most important interaction on the platform

- Motivations for retweeting (boyd et al., 2010):
  - To publicly agree with someone
  - To validate others’ thoughts
OBSERVED ON VARIOUS POLITICAL LANDSCAPES

Highest level of polarization

(Barberá et al, 2015)
OBSERVED ON VARIOUS POLITICAL LANDSCAPES

- **2010 US midterm elections**
  (Conover et al, 2011)
  - Retweet network
  - 93% right-leaning profiles

- **Secular vs Islamist polarization in Egypt**
  (Weber et al, 2013)
  - Retweet network
  - 80% left-leaning profiles

Color = cluster assignment

Islamists
Secularists
Center
2017 French presidential election (Fraisier et al, 2018)

Retweet network
Average number of retweets by profile:
• Intra-party: 149
• Inter-party: 4

Mention network
Average number of mentions by profile:
• Intra-party: 281
• Inter-party: 14
STANCE DETECTION
AIM

- Detect profiles' political stance based on their activity
AIM

- Detect profiles' political stance based on their activity

  - Global political stance
    - Political parties
    - Conservatives vs Liberals
    - Left vs Right
AIM

- Detect profiles' political stance based on their activity

  - Global political stance
    - Political parties
    - Conservatives vs Liberals
    - Left vs Right

  - Specific political stance
    - Political figure
    - Abortion
    - Climate change
    - Feminism
    - Gun control
    - LGBT rights
    - Immigration
    - Israeli-palestinian conflict
BASED ON TWEETS' TEXTUAL CONTENT
BASED ON TWEETS' TEXTUAL CONTENT

- Supervised models (Naive Bayes & SVM) (Mohammad et al., 2017; Conover et al, 2011)
Based on Tweets' Textual Content

- Supervised models (Naive Bayes & SVM) (Mohammad et al., 2017; Conover et al, 2011)
  - n-grams, hashtags, punctuation, capitalization, emoticons, ...
Based on Tweets' Textual Content

- Supervised models (Naive Bayes & SVM) (Mohammad et al., 2017; Conover et al, 2011)
  - n-grams, hashtags, punctuation, capitalization, emoticons, ...
  - Sentiment lexicon
BASED ON TWEETS' TEXTUAL CONTENT

- Supervised models (Naive Bayes & SVM) \( \text{(Mohammad et al., 2017; Conover et al, 2011)} \)
  - n-grams, hashtags, punctuation, capitalization, emoticons, …
  - Sentiment lexicon

- Unsupervised method to reduce the need for annotated data
Based on tweets' textual content

- Supervised models (Naive Bayes & SVM) (Mohammad et al., 2017; Conover et al., 2011)
  - n-grams, hashtags, punctuation, capitalization, emoticons, ...
  - Sentiment lexicon

- Unsupervised method to reduce the need for annotated data
  - Topic modeling (Fang et al., 2015)
SUPervised models (Naive Bayes & SVM) (Mohammad et al., 2017; Conover et al., 2011)
- n-grams, hashtags, punctuation, capitalization, emoticons, …
- Sentiment lexicon

Unsupervised method to reduce the need for annotated data
- Topic modeling (Fang et al., 2015)
- Poisson's law modeling of the discourse (Boireau, 2014)
BASED ON PROFILES' SOCIAL INTERACTIONS
BASED ON PROFILES' SOCIAL INTERACTIONS

- Retweet network
BASED ON PROFILES' SOCIAL INTERACTIONS

- Retweet network
  - Label propagation (Conover et al., 2011)
BASED ON PROFILES' SOCIAL INTERACTIONS

- Retweet network
  - Label propagation (Conover et al., 2011)
  - Community detection (Cherepnalkoski & Mozetic, 2015; Guerrero-Solé, 2017)
Retweet network

- **Label propagation** *(Conover et al., 2011)*
- **Community detection** *(Cherepnalkoski & Mozetic, 2015; Guerrero-Solé, 2017)*

Figure 4. Composition of the 9 communities by political groups in the core network. Different colors indicate the 8 political groups in the EP.
Based on Profiles’ Social Interactions

- Retweet network
  - Label propagation (Conover et al., 2011)
  - Community detection (Cherepnalkoski & Mozetic, 2015; Guerrero-Solé, 2017)

- Friends / Followers network

Figure 4. Composition of the 9 communities by political groups in the core network. Different colors indicate the 8 political groups in the EP.
Based on profiles' social interactions

- Retweet network
  - Label propagation (Conover et al., 2011)
  - Community detection (Cherepnalkoski & Mozetic, 2015; Guerrero-Solé, 2017)

- Friends / Followers network
  - Bayesian modeling (Barberá, 2015)

Figure 4. Composition of the 9 communities by political groups in the core network. Different colors indicate the 8 political groups in the EP.
Figure 4. Composition of the 9 communities by political groups in the core network. Different colors indicate the 8 political groups in the EP.
Politics on Twitter: a Panorama

Based on profiles’ social interactions

- Retweet network
  - Label propagation (Conover et al., 2011)
  - Community detection (Cherepnalkoski & Mozetic, 2015; Guerrero-Solé, 2017)

- Friends / Followers network
  - Bayesian modeling (Barberá, 2015)
  - Use of friends interests (Volkova et al., 2016)
BASSED ON PROFILES' SOCIAL INTERACTIONS

- Retweet network
  - Label propagation (Conover et al., 2011)
  - Community detection (Cherepnalkoski & Mozetic, 2015; Guerrero-Solé, 2017)

- Friends / Followers network
  - Bayesian modeling (Barberá, 2015)
  - Use of friends interests (Volkova et al., 2016)

Figure 4. Composition of the 9 communities by political groups in the core network. Different colors indicate the 8 political groups in the EP.
POLITICS ON TWITTER: A PANORAMA

BASED ON TEXT AND SOCIAL INTERACTIONS
BASED ON TEXT AND SOCIAL INTERACTIONS

Textual content

Social interactions

Joint use
Topic modeling taking into account tweets and social graph
(Thonet et al., 2017)
Based on text and social interactions

- Topic modeling taking into account tweets and social graph (Thonet et al., 2017)
- SVM trained on tweets and social graph (Magdy et al., 2016)
BASED ON TEXT AND SOCIAL INTERACTIONS
Based on Text and Social Interactions

Textual content

Social interactions

Mutual reinforcement
Based on Text and Social Interactions

- Consistence between tweets and retweets (Wong et al., 2016)
Based on text and social interactions

- Consistence between tweets and retweets
  (Wong et al., 2016)
Based on text and social interactions

- Consistency between tweets and retweets (Wong et al., 2016)
- Supervised classification with possible corrections from social graph (Pennacchiotti & Popescu, 2011)
Based on text and social interactions

- Consistence between tweets and retweets (Wong et al., 2016)
- Supervised classification with possible corrections from social graph (Pennacchiotti & Popescu, 2011)
- Supervised classification on text followed by propagation on social graph (Rabelo et al., 2012)
Based on Text and Social Interactions

- Consistence between tweets and retweets (Wong et al., 2016)
- Supervised classification with possible corrections from social graph (Pennacchiotti & Popescu, 2011)
- Supervised classification on text followed by propagation on social graph (Rabelo et al., 2012)
- Social graph for a portion of tweets followed by supervised classification (Rajadesingan & Liu, 2014)
Politics on Twitter: a Panorama

Based on Text and Social Interactions

- Consistence between tweets and retweets (Wong et al., 2016)
- Supervised classification with possible corrections from social graph (Pennacchiotti & Popescu, 2011)
- Supervised classification on text followed by propagation on social graph (Rabelo et al., 2012)
- Social graph for a portion of tweets followed by supervised classification (Rajadesingan & Liu, 2014)
ELECTION PREDICTION
### MULTIPLES ATTEMPTS

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>US presidential election</td>
<td>(O’Connor et al. 2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Gayo-Avello 2011)</td>
</tr>
<tr>
<td>2009</td>
<td>German federal election</td>
<td>(Tumasjan et al. 2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Jungherr et al. 2011)</td>
</tr>
<tr>
<td>2010</td>
<td>US elections in various states</td>
<td>(Metaxas et al. 2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Livne et al. 2011)</td>
</tr>
<tr>
<td>2011</td>
<td>Irish general election</td>
<td>(Bermingham &amp; Smeaton, 2011)</td>
</tr>
<tr>
<td></td>
<td>Singaporean general election</td>
<td>(Skoric et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>Dutch senate election</td>
<td>(Sang &amp; Bos, 2012)</td>
</tr>
<tr>
<td>2013</td>
<td>Pakistani general election</td>
<td>(Razzaq et al., 2014)</td>
</tr>
<tr>
<td>2015</td>
<td>Venezuelan parliamentary election</td>
<td>(Castro et al., 2017)</td>
</tr>
</tbody>
</table>
### Politics on Twitter: a Panorama

#### Multiples Attempts

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
<th>References</th>
<th>Volume of tweets</th>
<th>Sentiment analysis</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>US presidential election</td>
<td>(O’Connor et al. 2010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Gayo-Avello 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>German federal election</td>
<td>(Tumasjan et al. 2010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Jungherr et al. 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>US elections in various states</td>
<td>(Metaxas et al. 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Livne et al. 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>Irish general election</td>
<td>(Bermingham &amp; Smeaton, 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Singaporean general election</td>
<td>(Skoric et al., 2012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>Dutch senate election</td>
<td>(Sang &amp; Bos, 2012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>Pakistani general election</td>
<td>(Razzaq et al., 2014)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Venezuelan parliamentary election</td>
<td>(Castro et al., 2017)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Good predictions & better than traditional polls*
BUT...

- Highly dependant on data collection
BUT...

- Highly dependant on data collection
- Rarely takes into account bias in Twitter data
BUT...

- Highly dependant on data collection
- Rarely takes into account bias in Twitter data
  - Demographics bias
BUT...

- Highly dependant on data collection
- Rarely takes into account bias in Twitter data
  - Demographics bias
  - Vocal minority vs silent majority
BUT...

- Highly dependant on data collection
- Rarely takes into account bias in Twitter data
  - Demographics bias
  - Vocal minority vs silent majority
- Data purity questionable
BUT...

- Highly dependant on data collection
- Rarely takes into account bias in Twitter data
  - Demographics bias
  - Vocal minority vs silent majority
- Data purity questionable
  - Not all collected profiles eligible to vote
BUT...

- Highly dependant on data collection
- Rarely takes into account bias in Twitter data
  - Demographics bias
  - Vocal minority vs silent majority
- Data purity questionable
  - Not all collected profiles eligible to vote
- For the time being, not better than traditional polls
STUDY OF POLITICAL ENGAGEMENT
COMMUNICATIONS OF GUN POLICY ORGANIZATIONS

(Merry, 2016)

Brady Campaign

<table>
<thead>
<tr>
<th></th>
<th>Tweets containing character</th>
<th>% of tweets with Twitter handle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ally</td>
<td>492</td>
<td>5.7</td>
</tr>
<tr>
<td>Hero</td>
<td>800</td>
<td>30.0</td>
</tr>
<tr>
<td>Opponent</td>
<td>25</td>
<td>4.0</td>
</tr>
<tr>
<td>Villain</td>
<td>730</td>
<td>9.0</td>
</tr>
</tbody>
</table>

NRA

<table>
<thead>
<tr>
<th></th>
<th>Tweets containing character</th>
<th>% of tweets with Twitter handle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ally</td>
<td>289</td>
<td>10.4</td>
</tr>
<tr>
<td>Hero</td>
<td>519</td>
<td>30.3</td>
</tr>
<tr>
<td>Opponent</td>
<td>259</td>
<td>3.9</td>
</tr>
<tr>
<td>Villain</td>
<td>508</td>
<td>5.1</td>
</tr>
</tbody>
</table>
2017 FRENCH PRESIDENTIAL CAMPAIGN

(Fraisier et al., 2018)

Profiles

<table>
<thead>
<tr>
<th>Party</th>
<th>No.</th>
<th>Male</th>
<th>Female</th>
<th>Other/Und.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FI</td>
<td>22%</td>
<td>6%</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>PS</td>
<td>8%</td>
<td>20%</td>
<td>17%</td>
<td>17%</td>
</tr>
<tr>
<td>EM</td>
<td>17%</td>
<td>17%</td>
<td>14%</td>
<td>14%</td>
</tr>
<tr>
<td>LR</td>
<td>19%</td>
<td>14%</td>
<td>24%</td>
<td>24%</td>
</tr>
<tr>
<td>FN</td>
<td>15%</td>
<td>31%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Und.</td>
<td>15%</td>
<td>22%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Rest</td>
<td></td>
<td>21%</td>
<td>11%</td>
<td>11%</td>
</tr>
</tbody>
</table>

Tweets

<table>
<thead>
<tr>
<th>Party</th>
<th>No.</th>
<th>Male</th>
<th>Female</th>
<th>Other/Und.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FI</td>
<td>22%</td>
<td>6%</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>PS</td>
<td>8%</td>
<td>20%</td>
<td>17%</td>
<td>17%</td>
</tr>
<tr>
<td>EM</td>
<td>17%</td>
<td>17%</td>
<td>14%</td>
<td>14%</td>
</tr>
<tr>
<td>LR</td>
<td>19%</td>
<td>14%</td>
<td>24%</td>
<td>24%</td>
</tr>
<tr>
<td>FN</td>
<td>15%</td>
<td>31%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Und.</td>
<td>15%</td>
<td>22%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Rest</td>
<td></td>
<td>21%</td>
<td>11%</td>
<td>11%</td>
</tr>
</tbody>
</table>

Retweets

<table>
<thead>
<tr>
<th>Party</th>
<th>No.</th>
<th>Male</th>
<th>Female</th>
<th>Other/Und.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FI</td>
<td>22%</td>
<td>6%</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>PS</td>
<td>8%</td>
<td>20%</td>
<td>17%</td>
<td>17%</td>
</tr>
<tr>
<td>EM</td>
<td>17%</td>
<td>17%</td>
<td>14%</td>
<td>14%</td>
</tr>
<tr>
<td>LR</td>
<td>19%</td>
<td>14%</td>
<td>24%</td>
<td>24%</td>
</tr>
<tr>
<td>FN</td>
<td>15%</td>
<td>31%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Und.</td>
<td>15%</td>
<td>22%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Rest</td>
<td></td>
<td>21%</td>
<td>11%</td>
<td>11%</td>
</tr>
</tbody>
</table>

Number of tweets:

- FI: 53%
- PS: 38%
- EM: 52%
- LR: 51%
- FN: 57%
- Und.: 42%
- Rest: 52%

Number of retweets:

- FI: 22%
- PS: 7%
- EM: 17%
- LR: 21%
- FN: 28%
- Und.: 20%
- Rest: 6%
ININVOLVEMENT IN OCCUPY WALL STREET

(Conover et al., 2013)
ITALIAN INTRA-PARTY POLITICS

- (Ceron, 2017)
COALITIONS IN THE EUROPEAN PARLIAMENT

Co-voting agreement within and between political groups

Average retweets within and between political groups

(Cherepnalkosk, 2016)
DETECTION OF SOCIAL UNREST

- Social unrest: public expression of discontent, including public protest that does not threaten the regime’s hold on power, and/or sporadic but low-level violence.

➡ Identifying tweets relevant to social unrest (Mishler et al., 2017)
➡ Identifying unstable countries based on tweets (Raja et al., 2016)
• Large body of work on Twitter and politics
  • Various tasks
  • Diversity of subjects, after being focused on US politics for some time

• Known limits
  • Need for caution when extrapolating

• Importance of quantitative & qualitative analysis
THANK YOU
FOR YOUR ATTENTION
Politics on Twitter: a Panorama

- Gaudin, S. (2016, novembre 8). In presidential campaign, Twitter was a powerful political tool. https://www.computerworld.com/article/3137261/social-media/in-presidential-campaign-twitter-was-a-powerful-political-tool.html
Politics on Twitter: A Panorama

- Volkova, S., Bachrach, Y., & Durme, B. V. (2016). Mining User Interests to Predict Perceived Psycho-Demographic Traits on Twitter (p. 36–43). IEEE. https://doi.org/10.1109/BigDataService.2016.28