MULTI-DIMENSIONAL POLITICAL ANALYTICS

JAKOB AABLING-THOMSEN
TOINE BOGERS
About this talk

• Work in progress
  – Open to any and all collaboration ideas
  – Please come with questions, comments, suggestions!
Outline

• What? What is the project about?
• Why? What’s our motivation?
• How? What combination of methods will we use?
• When? What is the project timeline?
• Who? Who is a part of this?
What / Breaking it down

• **Multi-dimensional**
  - Analysis based on multiple dimensions and data sources

• **Political**
  - Focused on the 2015 Danish general elections

• **Analytics**
  - Combination of quantitative and qualitative methods for data collection and analysis
What / Analysis dimensions

- **Topical**: What topics are most commonly discussed?
- **Social**: How do interactions between actors influence the discourse?
- **Temporal**: How does political discourse on Twitter develop?
- **Geographical**: Where is the interaction coming from?
- **Emotional**: What sentiments are expressed about topics and politicians?
- **Causal**: How does social media influence politicians and vice versa?
- **Ethical**: What are the ethics regarding political discourse/analytics?
What / Data sources

• Twitter
  – Main data source in the project
  – Popular communication channel among politicians (390 politicians, #dkpol)
Kampen om at blive Danmarks største parti

Alle Megafonmålinger siden valget i 2011.

- DF
- Venstre
- Soc.dem.

20.0%

2012, 2013, 2014

Interaction functionality (reply, re-tweet, favorite)
Så starter forhandlingerne om at droppe reklameafgiften. Mon ikke vi finder pengene og miljøkompensationen et andet sted.

#dkpol

Replies to this tweet

Jørgen Zinck Jensen @Zinckjo · 9h
@PerClausen3 @dkpolparrot Mon ikke vi finder pengene og miljøkomp. et andet sted. Det var sandelig et synspunkt !!!

Michael Jul Jensen @JulJenssen · 8h
@PerClausen3 BOLIGJÆRNE!!!

Torben Magnusson @torben_ironfoot · 8h
@PerClausen3 der er masser af eksempler på fråde og spild i offentlig sektor, så de midler finder i sagtens
What / Data sources

• Twitter
  – Main data source in the project
  – Popular communication channel among politicians (390 users 11/12-14)

• Online newspapers
  – Both news articles and the online discussion

• Election & opinion polls

• Prediction markets
  – Online markets created for the purpose of making predictions
  – Example: betting on political events

• Other social media
**Why / Motivation**

- Current research topic (with general elections expected in February 2015)
- Popular communication channel among politicians
- Relatively little related work on
  - How the different dimensions compare & interact
  - Interactions between different media sources
  - How this applies to Denmark and Danish politics
- Combines big data with deep data
- Potential for developing new methods of data analysis
- Interesting results could generate lots of media attention
How / Methods

• Collection
  - Crawling tweets and user profiles from Twitter (main data source)
    ▪ Continuous adjustments based on developing election themes
  - Collecting other data sources
    ▪ Election/opinion polls & prediction markets
    ▪ Online media & related discussions
  - Supplemented with qualitative methods
    ▪ Content analysis
    ▪ Interviews
How / Methods

• Analysis
  - Data mining on tweets, user profiles, and interaction patterns
  - Data visualization
  - Content analysis (automatic & manual)
  - Sentiment analysis
  - Linking data from Twitter, polls, prediction markets and media coverage
How / Topical

• What?
  - What political topics are most commonly discussed?
  - Are there differences in which topics are discussed where? And how?
  - Can we predict the political orientation of a user based on their tweets?

• How?
  - Identify political hot-button issues from the public debate
  - Identifying & crawling relevant hashtags
  - Combining content analysis with automatic topic modeling
  - Identifying & visualizing popularity of topics for different domains
  - Machine learning for predicting political orientation
How / Social

• What?
  − How do politicians interact with each other and other Danes through Twitter?
  − Do these interactions between users influence the political discourse?
  − How does the social network influence information propagation?

• How?
  − Tracking re-tweets along the social networks of Twitter users
  − Visualizing & analyzing replies, re-tweets, and friendship relations on Twitter
  − Identifying the properties that make tweets more likely to be retweeted
Example: Interactions between red and blue Twitter users on Ferguson riots
How / Temporal

• What?
  – How does political discourse on Twitter develop over time?
  – How does politicians’ use of Twitter change over time (both short-term and long-term)?
  – How do external events reverberate on Twitter (and vice versa)?
  – What influences the expiration date of political stories?

• How?
  – Identifying and tracking events in political tweets
  – Visualizing time series of Twitter interaction
  – Predicting the popularity of stories/events
How / Geographical

• What?
  - How are tweets distributed geographically?
  - Can we predict where a tweet originated based on its content?

• How?
  - Extracting location information from user profiles and geo-tagged tweets
  - Analyzing the geographic distribution of tweets
Tweet #Utøya or #Utoya, and shine a light for democracy all over the world.

Example: Geographic visualization of support in the wake of the Utøya tragedy.
How / Emotional

• What?
  – What sentiment/emotions are expressed about topics and politicians?
  – How objective/subjective are political tweets?
  – Do politicians refer to own emotions or appeal to those of their voters?
  – How are stereotypes (e.g., ‘blå Bjarne’) used to appeal to voter groups?

• How?
  – Automatic sentiment & subjectivity analysis
  – Analyzing the pronouns related to expressed sentiment
  – Content analysis of use of stereotypes in tweets
How / Causal

• What?
  − How does social media influence the political discourse (and vice versa)?
  − How do offline events influence the online world (and vice versa)?
  − What causes politicians to be (un)successful on social media?

• How?
  − Correlating social media events with the political discourse
  − Interviews about possible influences on the political agenda & strategic use of social media
How / Ethical

• What?
  – Is hate speech an increasing problem in the (online) public domain?
  – Which types of hate speech are most common and for which domains?
  – How big of a problem is paid online slandering/bullying?
  – How well is netiquette observed in the online political discourse (compared to the offline discourse)?
  – What are the ethical consideration of our data collection process?

• How?
  – Automatically identify & crawl tweets containing hate speech
  – Content analysis of hate speech in tweets
How / Collecting data from Twitter

• **Twitter search results**
  - Downloads 100 most recent tweets relevant for a query
    ▪ Any kind of query: ‘#dkpol’, ‘gucci helle’, ‘@larsloekke’, ...
  - Can set crawling interval/alarm for individual queries
    ▪ Every hour, at midnight, etc…

• **User profiles & timelines**
  - Downloads 200 most recent tweets on a user’s timeline
  - Downloads user profile information (including profile images)
  - Can set crawling interval/alarm for individual users
How / Collecting data from Twitter

• Individual tweets
  – Given a tweet ID, download all data for that tweet
  – Can be used to download original tweets in case we only have re-tweets in our data set

• Retweets of a tweet
  – Given a tweet ID, download the 100 most recent re-tweets
How / Example of crawled data

• Collecting data from Twitter
<table>
<thead>
<tr>
<th>retweet_count</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>retweet_of_tweet_id</td>
<td>-</td>
</tr>
<tr>
<td>retweet_of_screen_name</td>
<td>-</td>
</tr>
<tr>
<td>screen_name</td>
<td>CasperDall</td>
</tr>
<tr>
<td>source</td>
<td>Twitter for iPhone</td>
</tr>
<tr>
<td>text</td>
<td>Dagens grafer i #dkpol: DF er største parti... via @tv2politik #fv15 <a href="http://t.co/nzdsYMcKUx">http://t.co/nzdsYMcKUx</a> <a href="http://t.co/TxsdcoH9sj">http://t.co/TxsdcoH9sj</a></td>
</tr>
<tr>
<td>tweet_url</td>
<td><a href="https://twitter.com/CasperDall/status/538008410361696256">https://twitter.com/CasperDall/status/538008410361696256</a></td>
</tr>
<tr>
<td>user_id</td>
<td>25153213</td>
</tr>
</tbody>
</table>
When / Tentative timeline

- Testing Twitter crawler
- Crawling Twitter data
- Expected general elections
- Literature review
- Student projects/theses
- Data analysis
Who / Researchers

• Jakob Aabling-Thomsen
  – Original idea, research setup, PhD-project?
• Toine Bogers
  – Data collection & analysis
• You? 😊
Who / Students

- Master’s thesis (KOM)
  - Investigate how politicians use Twitter as a channel for political communication
- 4th semester course “Kommunikation og individ” (Humanistisk Informatik)
  - Examining the prevalence of hate speech on Twitter
- Possible collaboration:
  - 4th semester course “Brugerdreven innovation og formidling af ny teknologi” (Techno-Anthropology)
Questions, comments, suggestions?

• Example questions
  - Problems with/additions to the data collection setup?
  - Other analysis methods we should use?
  - Suggestions for a catchy acronym?
  - Is this project relevant/interesting (to you)?
  - Do you have any additional ideas?
  - Would you like to collaborate?
  - Suggestions for money (Please give us money! 😊)
  - …