A methodology proposal to explore Facebook pages content using text mining

Objectives

• To build a strong framework for a deep content analysis on Facebook pages, as a good portion of 2000's historical narratives on contentious politics are being told on Social Network platforms like Twitter and Facebook.

• To offer social researchers an intuitive tool to investigate current social processes online (e.g.: political mobilization) using digital methods.

Methods

• Initial corpus = 99,921 posts from 46 pages of social movements and grassroots media in Brazil, for a core-period of 3 years (2013-2016)

• A set of procedures were applied to organize content by thematic, relevance and time:
  ✓ identify the object of study and the problems to be investigated
  ✓ define a method to collect data from Facebook and persist locally (users information is anonymized by default)
  ✓ proceed an exploratory analysis to understand basic characteristics within the data and to remove noise
  ✓ apply topic modeling to summarize the corpus
  ✓ identify topic's representative terms and explore its context
  ✓ identify key posts & events and explore
A methodology proposal to explore Facebook pages content using text mining
A methodology proposal to explore Facebook pages content using text mining

Conclusions
Automated topic modeling helps to reduce the sense of fragmentation by grouping a massive amount of posts into a limited set of thematics. The technique is the starting point for a set of visual graphics:

- Topics can be seen in accordance to the similarity among them or along the time.
- Timelines show peaks of activity, indicating where to focus a deeper content analysis.
- Semantic networks are like aerial views on the topic or page vocabulary, helping to understand which particular terms have been used in correlation to each others.
- Relative clouds of words are effective to compare general vocabulary between two pages.

Future work
This study is underway and is part of investigations at a PhD and a Masters levels. Next steps include to deepen the content analysis. The final outputs should include a web application where users can explore data sets from Facebook pages. The tool should automatize some procedures to support social researchers in the use of Facebook content as a source for social phenomena investigation.

Thank you!
Marcela Canavarro (mcanavarro@gmail.com)
Arian Pasquali (arrp@inesctec.pt)
Alípio Jorge (amjorge@fc.up.pt)