A methodology proposal to explore Facebook pages content using text mining

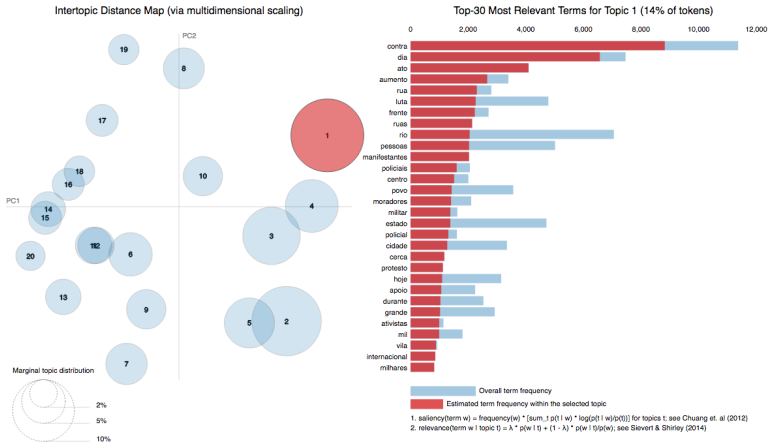
Objectives

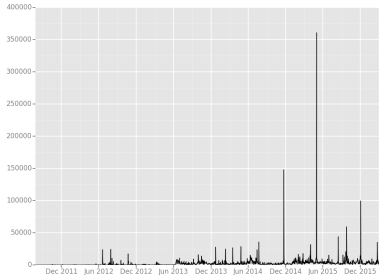
- To build a strong framework for a deep content analysis on Facebook pages, as a good portion of 2000's historical narratives on contentious politics are being told on Social Network platforms like Twitter and Facebook.
- To offer social researchers an intuitive tool to investigate current social processes online (e.g.: political mobilization) using digital methods.

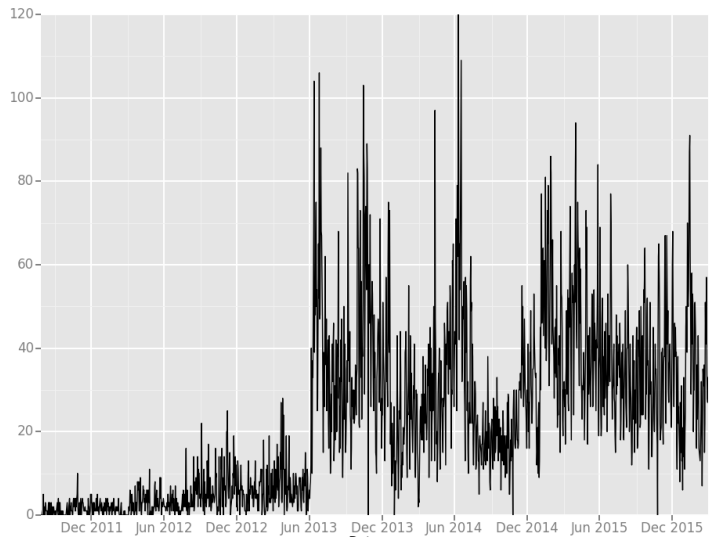
Methods

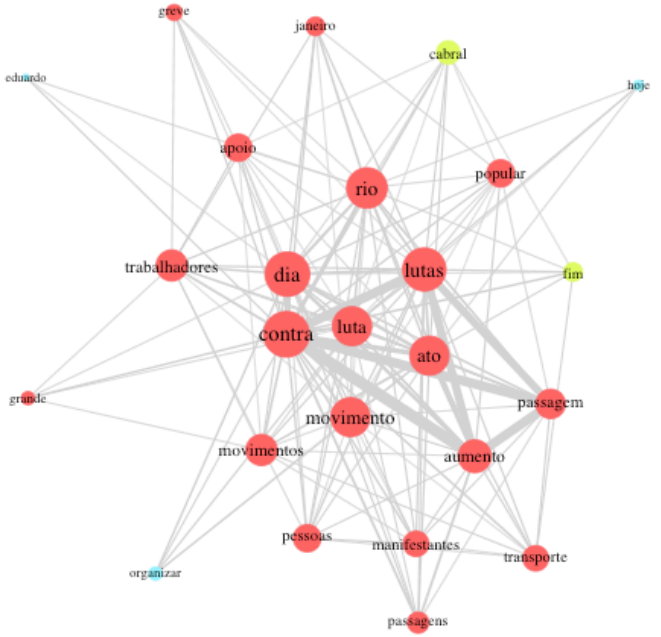
- Initial corpus = 99,921 posts from 46 pages of social movements and grassroots media in Brazil, for a core-period of 3 years (2013-2016)
- A set of procedures were applied to organize content by thematic, relevance and time:
  - ✓ identify the object of study and the problems to be investigated
  - ✓ 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

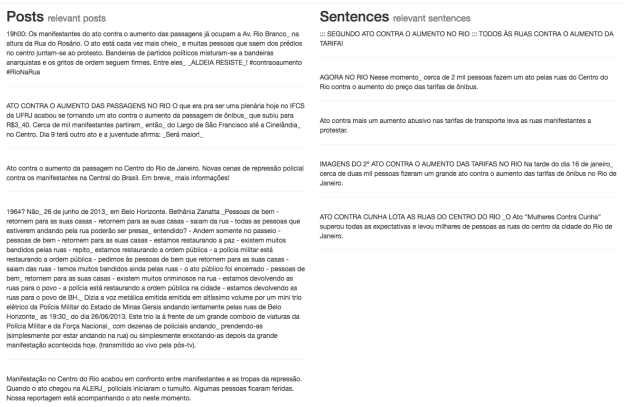
A methodology proposal to explore Facebook pages content using text mining











Posts relevant posts

19h00: Os manifestantes do ato contra o aumento das passagens já ocupam a Av. Rio Branco_, na altura da Rua do Rosário. O ato está cada vez mais cheio_ e muitas pessoas que saem dos prédios no centro juntam-se ao protesto. Bandeiras de partidos políticos misturam-se a bandeiras anarquistas e os gritos de ordem seguem firmes. Entre eles_ _ALDEIA RESISTE_! #contraoaumento #RioNaRua

ATO CONTRA O AUMENTO DAS PASSAGENS NO RIO O que era pra ser uma plenária hoje no IFCS da UFRJ acabou se tornando um ato contra o aumento da passagem de ônibus_ que subiu para R$3_40. Cerca de mil manifestantes partiram_ então_ do Largo de São Francisco até a Cinelândia_ no Centro. Dia 9 terá outro ato e a juventude afirma: _Será maior!_

Ato contra o aumento da passagem no Centro do Rio de Janeiro. Novas cenas de repressão policial contra os manifestantes na Central do Brasil. Em breve_ mais informações!

1964? Não_ 26 de junho de 2013_ em Belo Horizonte. Bethânia Zanatta _Pessoas de bem - retornem para as suas casas - retornem para as suas casas - saiam da rua - todas as pessoas que estiverem andando pela rua poderão ser presas_ entendido? - Andem somente no passeio - pessoas de bem - retornem para as suas casas - estamos restaurando a paz - existem muitos bandidos pelas ruas - repito_ estamos restaurando a ordem pública - a polícia militar está restaurando a ordem pública - pedimos às pessoas de bem que retornem para as suas casas - saiam das ruas - temos muitos bandidos ainda pelas ruas - o ato público foi encerrado - pessoas de bem_ retornem para as suas casas - existem muitos criminosos na rua - estamos devolvendo as ruas para o povo - a polícia está restaurando a ordem pública na cidade - estamos devolvendo as ruas para o povo de BH_. Dizia a voz metálica emitida emitida em altíssimo volume por um mini trio elétrico da Polícia Militar do Estado de Minas Gerais andando lentamente pelas ruas de Belo Horizonte_ as 19:30_ do dia 26/06/2013. Este trio ia à frente de um grande comboio de viaturas da Polícia Militar e da Força Nacional_ com dezenas de policiais andando_ prendendo-as (simplesmente por estar andando na rua) ou simplesmente arrastando-as depois da grande manifestação acontecida hoje. (transmitido ao vivo pela pós-tv).

Manifestação no Centro do Rio acabou em confronto entre manifestantes e as tropas da repressão. Quando o ato chegou na ALERJ_ policiais iniciaram o tumulto. Algumas pessoas ficaram feridas. Nossa reportagem está acompanhando o ato neste momento.

Sentences relevant sentences

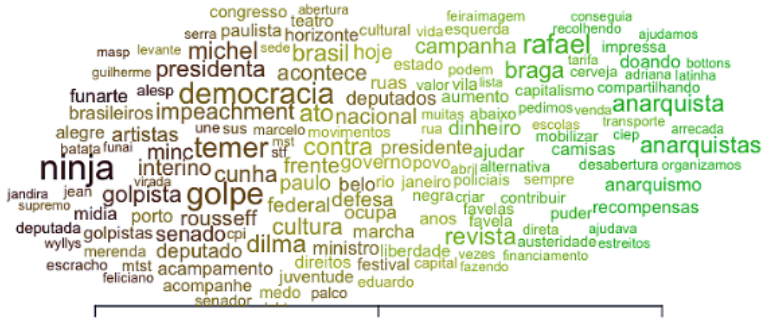
::: SEGUNDO ATO CONTRA O AUMENTO NO RIO ::: TODOS ÀS RUAS CONTRA O AUMENTO DA TARIFA!

AGORA NO RIO Nesse momento_ cerca de 2 mil pessoas fazem um ato pelas ruas do Centro do Rio contra o aumento do preço das tarifas de ônibus.

Ato contra mais um aumento abusivo nas tarifas de transporte leva as ruas manifestantes a protestar.

IMAGENS DO 2º ATO CONTRA O AUMENTO DAS TARIFAS NO RIO Na tarde do dia 16 de janeiro_ cerca de duas mil pessoas fizeram um grande ato contra o aumento das tarifas de ônibus no Rio de Janeiro.

ATO CONTRA CUNHA LOTA AS RUAS DO CENTRO DO RIO _O Ato "Mulheres Contra Cunha" superou todas as expectativas e levou milhares de pessoas as ruas do centro da cidade do Rio de Janeiro.



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 similiarity 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 in 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)