

# Opinion mining in social networks

Corrado Monti  
Università degli Studi di Milano



# Measuring phenomena on Twitter

- ▶ Sentiment analysis and textual classification can extract information from the huge amount of Twitter messages
- ▶ Links with real-world indicators were discovered<sup>1</sup>
- ▶ Ok, but can we predict elections with Twitter?
- ▶ (Quite obviously) not really.<sup>2</sup>

---

<sup>1</sup>O'Connor et al., *From tweets to polls: Linking text sentiment to public opinion time series*, ICWSM 11 (2010): 122-129.

<sup>2</sup>Chung and Mustafaraj, *Can collective sentiment expressed on twitter predict political elections?* In 25th AAAI Conf. on AI, 2011.

- ▶ For which phenomena is this possible?
- ▶ Apparently economic trust is one of them<sup>3</sup>
- ▶ Can **political disaffection** in Italy be measured through massive tweet classification?
  - ▶ It is a relevant phenomenon
  - ▶ Lot of interest, academic (sociology) and not

---

<sup>3</sup>Bollen, Mao, Pepe, *Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena*, ICWSM 2011

# Text classification

- ▶ “political disaffection” → political topic, negative sentiment, presence of some keywords
- ▶ We had a training dataset of 28'340 labelled tweets
- ▶ We developed *ad-hoc* re-usable classification techniques
  - ▶ We built robust classifiers, thanks to ontologies from DBpedia

# Experimental comparison

## Surveys

- ▶ Accepted way to measure collective sentiment
- ▶ We got fraction of italians that say they would not vote **for any party**
- ▶ One every  $\sim 10$  days in April-October 2012



# Experimental comparison

## Surveys

- ▶ Accepted way to measure collective sentiment
- ▶ We got fraction of italians that say they would not vote **for any party**
- ▶ One every  $\sim 10$  days in April-October 2012



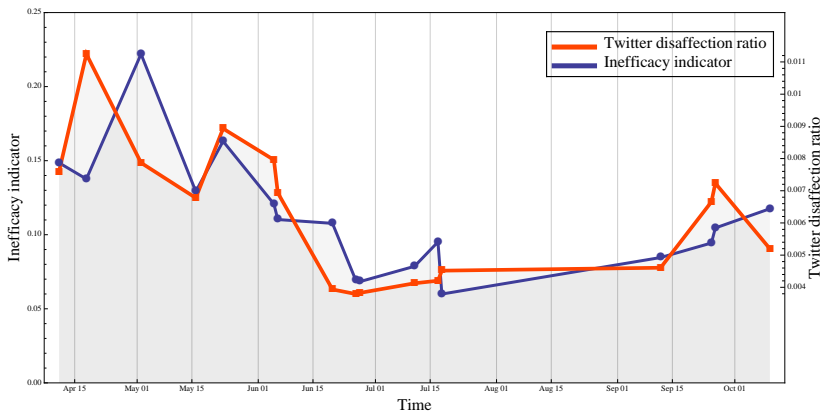
## Tweet sample

- ▶ 35'882'423 tweet
- ▶ For each survey, we compute the ratio of disaffected tweet volume over political tweet volume from  $\Delta = 14$  days before



# Results

Pearson correlation index for  $\Delta = 14$  days  $\rightarrow \rho = \mathbf{0.7860}$

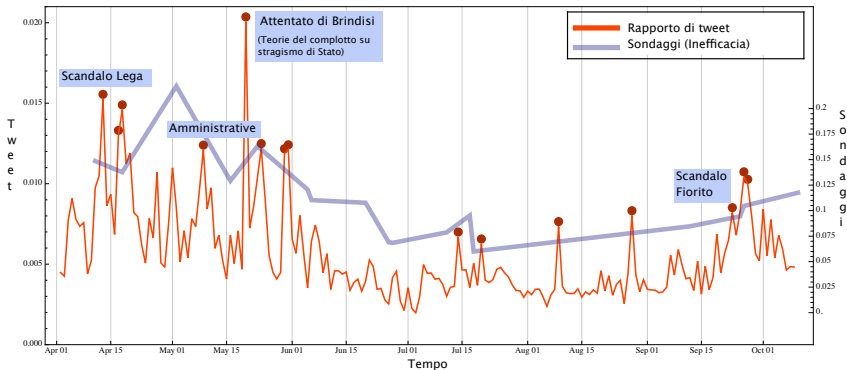


# Interpretation

- ▶ Data seem to indicate a good correlation between disaffected tweet and diffusion of the phenomena in society
- ▶ This does not mean that Twitter is a representative sample!
- ▶ We can guess that the quantity of discussion about *this phenomenon* is connected with how much it will spread



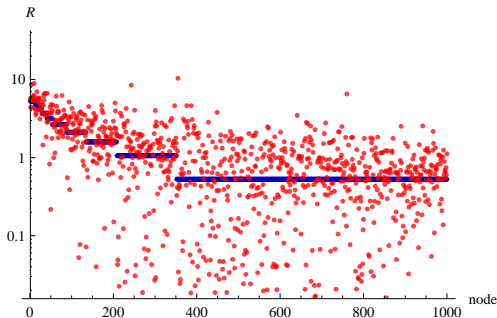
## We found peak causes from newspaper titles through text mining



# Ongoing work

- ▶ I plan to use these kind of data to better understand network centrality measures
- ▶ We are developing a social network model where every node is represented as a set of features
- ▶ Features can be also be opinions!

- ▶ In this model, every node has *a priori* ability to transmit feature
- ▶ We are more or less able to reconstruct the value of this ability through Gibbs Sampling



## Credits and References

- ▶ Corrado Monti, Matteo Zignani, Alessandro Rozza, Adam Arvidsson, Giovanni Zappella, and Elanor Colleoni. *Modelling political disaffection from twitter data*, Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining, p. 3. ACM, 2013.
- ▶ My supervisors are Paolo Boldi and Sebastiano Vigna
- ▶ Ongoing work with Irene Crimaldi (IMT Lucca)

# Thanks!

email: `corrado.monti@unimi.it`