Using Twitter data to study the mood on migration

A. Righi, D.M. Bianco, M.M. Gentile

Palazzo Falletti, Rome
3 - 4 October 2018

1) Goal

2) Why Twitter data to perform Sentiment analysis

3) Previous studies on Sentiment analysis on migration based on Twitter data

4) Construction of Mood on migrations index of Italian-speaking twitterers

   ✓ Data & methods used

   ✓ Results

5. Discussion

❑ Since the refugee crisis that has affected the Mediterranean area after 2015, the media have highlighted links between the growth of arrivals and the emotional attitudes of public opinion

❑ The media seems having contributed to the increase of this emotionality that in some countries has resulted in negative or racist attitudes with tones that are not neutral

❑ Due to their accessibility and their ability to catch dynamic reactions, online social media are important sources for analysing events, their evolution, and their perceptions

❑ This study aims to verify whether Twitter data can help to understand how the phenomenon of migrations in Italy (a country particularly affected by refugees flows) is perceived and how the perception evolved during the time, trying also to verify whether the perceptions are influenced by particular occurred events and by the role of the media

Why Twitter data to perform Sentiment analysis

❑ Social media is an important data source about the opinions and the sentiment of their users because they allow to capture in real-time and in a not solicited way what the users think about a certain topic

❑ In Italy, Facebook and Twitter appear to be the most used social media and the latter has a greater accessibility and allows a more readily text analysis

❑ Twitter is a microblogging service, created in 2006, which lets users post texts and which has around 330MIL monthly worldwide active users at the end of 2017, according to Statista website and around 14MIL users in Italy

❑ Twitter allows access to data through APIs, and both the users' profile information and their tweets are downloadable (Twitter, 2018)

❑ Twitter messages are typically short and frequently express emotions, therefore, various researchers have exploited the special characteristics of Twitter for sentiment analysis, opinion mining and user polarization

- In Bartlett and Norrie (2015) machine-learning algorithms revealed that people do not express a generic sentiment about immigration on Twitter but the theme is faced through immediate reaction to related events

- Gualda and Rebollo (2016) studied 2MIL tweets (Dec.2015–Mar.2017) about term "refugees" in 6 languages performing SA and SNA in a comparative perspective. The discourses were emotional and highly polarized

- Pope and Griffith (2015) using SA explored the mood relating to the refugee crisis in Europe and to the Paris and Cologne attacks across two populations (English and German speaking) showing interesting trends and commonalities

- UN Global Pulse and UNHCR (2017) faced the analysis of host communities' sentiment towards refugees in reaction to terrorist attacks in 2015-2016 in Europe showing that only a small number of people connected the refugees and the incidents

- Coletto et al (2016) multidimensional analyses on content, locations and time applied to the refugees crisis of 2015 revealed that Europeans mostly express positive sentiments toward the refugees, but this attitude changes when a countries is more exposed to migration flows

- Hadgu et al (2016) confirmed a) the role of the News as a mediator between the actual and the perceived refugee situation, b) the diversity of countries in the perception of the refugee situation, c) an increase of polarized discussion in a country when the number of refugees arises

- Chung and Zeng (2016) operationalized a framework for social-media-based public policy on U.S. immigration and border security; the system iMood collected messages, extracted user sentiment and discovered the user networks in order to identify influential users and activists

- ❑ Sentiment analysis: different methods based on: 1) unsupervised learning techniques (Pak and Paroubek, 2010), 2) lexicon-based method (Taboada et al, 2011), 3) combinations of the two approaches (Kolchyna et al, 2015), 4) polarization, controversy and topic tracking in time (Garimella et al 2016), 5) hashtag classification through probabilistic models (Coletto et al, 2016)

- We applied a procedure to collect and process messages containing at least one keyword belonging to a specific 'filter', namely a definite set of Italian hashtags

  ✓ Filter should capture relevant messages and eliminate off-topic ones

  ✓ Migration filters have been designed by subject-matter experts or the base of top trend topics

  1) 'Mood on migration' filter is designed to measure the Italian sentiment on the topic and the related policies. It is made up of around 40 relevant # referring to concept as migration, immigration, boundary, frontiers, foreigners, clandestine, refugees, repatriation, nationality, asylum, citizenship regular/irregular presence, racism, but also to more specific topics related to the news (landings, reception, hotspots, walls, NGOs, Malta, Libya, Mediterranean) or even to political positions (pro or versus the migrants);

  2) It allows to collect ~2,400 tweets/day in the considered period since Jan 2015 to Oct 2018, showing a strong increasing trend in 2018

- **Data Collection**
  - ✓ The Twitter's REST API is exploited to get samples of public tweets, it allows to search terms or get tweets based on specific parameters
  - ✓ Only public tweets whose text matches at least one keyword # belonging to the filter
  - ✓ Only the textual content of the tweets is analysed

- **Text Cleaning and Normalization**
  - ✓ Pre-processing steps:
  - ✓ (i) convert to lowercase,
  - ✓ (ii) tokenize text into words,
  - ✓ (iii) apply basic orthographic repairs,
  - ✓ (iv) remove URLs, non-alphabetic characters and stop words,
  - ✓ (v) lemmatization of words to remove inflectional endings and to return the base form of a word
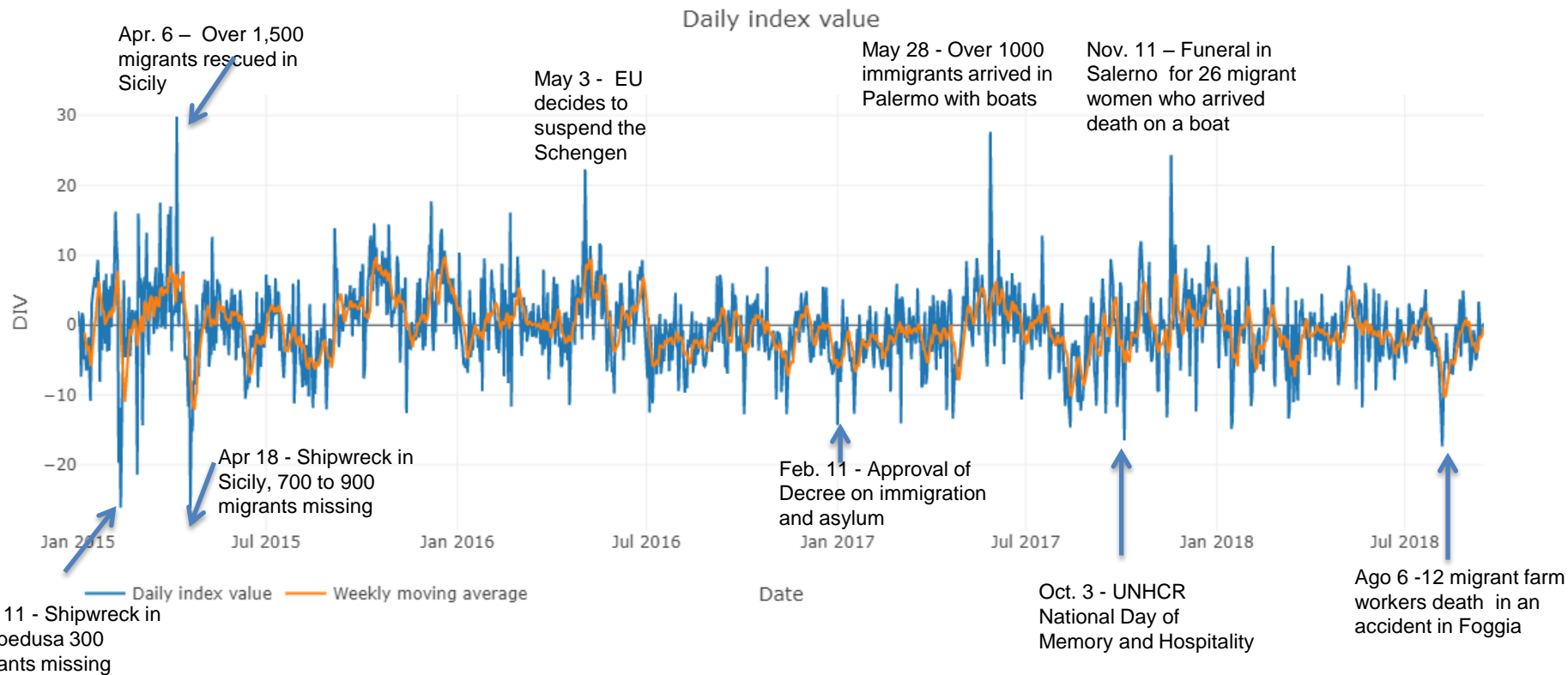
- **Sentiment Analysis Approach**
  - ✓ An unsupervised, lexicon-based approach has been adopted to classify tweets as Positive, Negative or Neutral

- Sentiment Analysis in two steps:
  1) Calculate sentiment scores for each tweet
  2) Use these sentiment scores to cluster tweets into three mutually exclusive classes: Positive (P), Negative (N) and Neutral (U)

- To attach sentiment scores to a tweet we used **Sentix**, the Italian Sentiment Lexicon whose lemmas are associated to pre-computed *positive* and *negative* sentiment scores

- As Sentix contains many duplicated lemmas we de-duplicated Sentix by averaging sentiment scores of duplicated lemmas

- Daily Sentiment scores are clustered (using a K-means algorithm) into Positive, Negative and Neutral

- The daily index value (M) is computed considering the distribution of tweets within the Positive, Neutral and Negative classes

$$M = \bar{\omega}_i = \frac{\sum_t i_t \, \omega_t}{\sum_t i_t} = \frac{\sum_{t \in P} i_t \, \omega_t + \sum_{t \in N} i_t \, \omega_t}{\sum_t i_t}$$

where $\quad \omega_t \overset{\text{def}}{=} 0 \;\; \forall t \in \text{Neutral}$

Daily index value

Apr. 6 – Over 1,500 migrants rescued in Sicily

May 3 - EU decides to suspend the Schengen

May 28 - Over 1000 immigrants arrived in Palermo with boats

Nov. 11 – Funeral in Salerno for 26 migrant women who arrived death on a boat

Apr 18 - Shipwreck in Sicily, 700 to 900 migrants missing

Feb. 11 - Approval of Decree on immigration and asylum

Feb. 11 - Shipwreck in Lampedusa 300 migrants missing

Oct. 3 - UNHCR National Day of Memory and Hospitality

Ago 6 -12 migrant farm workers death in an accident in Foggia

— Daily index value — Weekly moving average

- The index enables high-frequency (e.g. daily, weekly) measures of the Italian sentiment on migration
- The higher the value of the index, the better the mood and the observed peaks and valleys of the daily index can be related with specific events or declarations of politicians, measuring in this way their impact on public opinion
- The role of offline Media in relation to specific events occurred during the observed period can be analysed
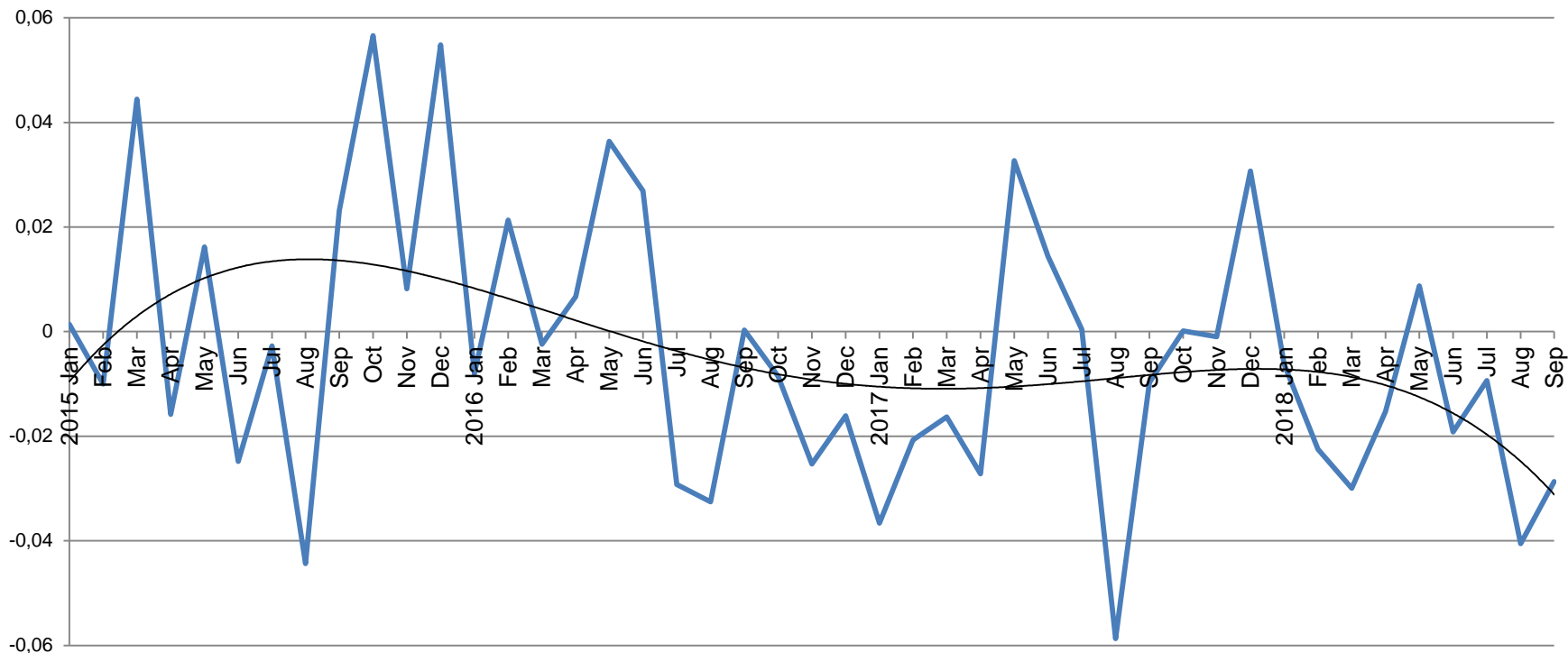
Daily number of tweets on the topic
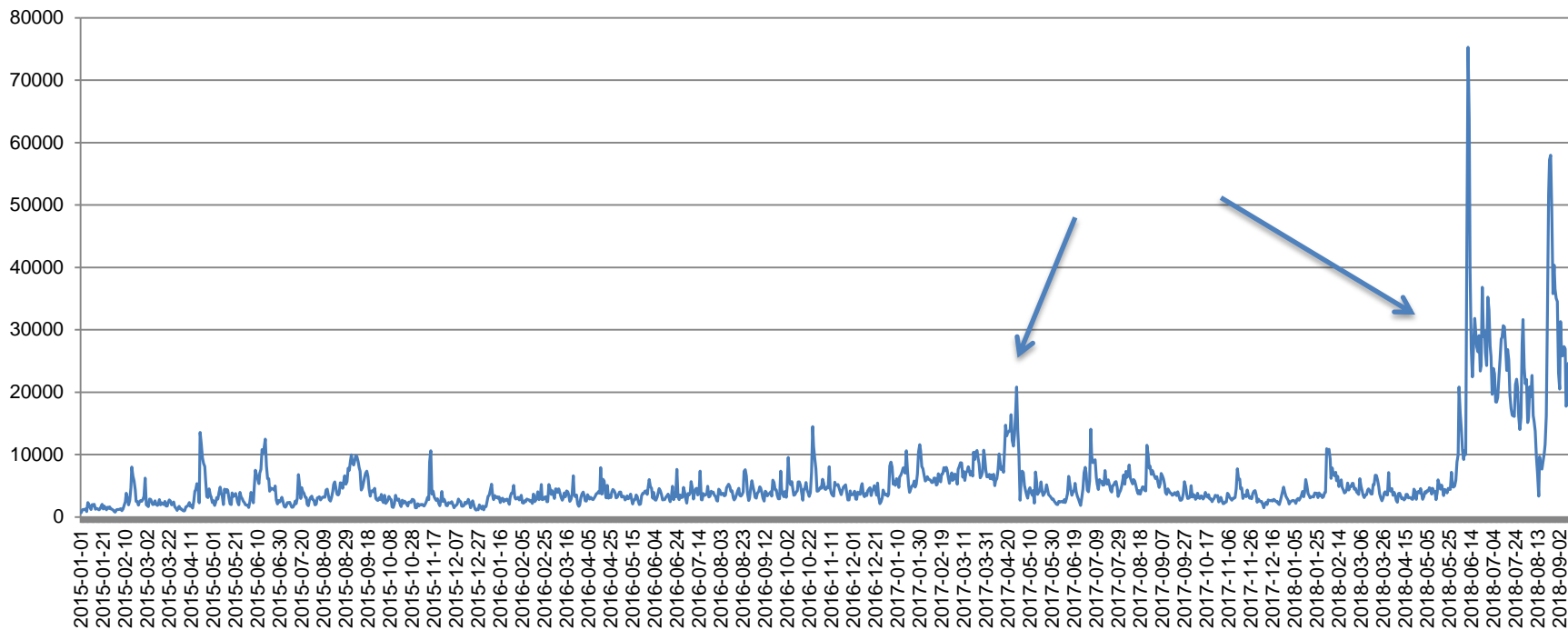
The number of daily messages, quite stable in the period jan 2015 dec 2016, slightly increased in the first months of 2017 and much more considerably after May 2018

Mood on Migration Index - Monthly averages

- The mood toward migration seems to move from an initial positive area to a negative one during the crisis of summer 2016 when the arrivals of migrants consistently increased,
- The negative sentiment seems deepening after March 2018

- The results confirm that Twitter data can help to understand how the phenomenon of migrations in Italy is perceived

- The new statistical tool analysing the text of filtered posted messages in order to measure the daily mood and to check how this evolves over time, allows the study of the effect on public opinion of the strong tones used by the mass media regarding the often dramatic episodes (shipwrecks of boats, or the rejections or the polemics on the redistribution of migrants)

- Even if the tool is based on social media data not representing the entire Italian population views, the index could provide a useful guidance to better manage the migration policies and the communication on mass media

- Further analysis will investigate the underlying reasons for the relationship between the trend of the mood and the political or social events