

Progetti di Sentiment Analysis

Rachele Sprugnoli – rachele.sprugnoli@unicatt.it

Centro Interdisciplinare di Ricerche per la Computerizzazione
dei Segni dell'Espressione (CIRCSE)



UNIVERSITÀ
CATTOLICA
del Sacro Cuore



COVID-19: Insights & Feelings

- Progetto congiunto di Expert System (azienda di tecnologie semantiche e intelligenza artificiale) e Sociometrica (azienda che si occupa di analisi economica e sociale)
- Monitoraggio sentiment ed emozioni dai social media
- Classificazione in 80 emozioni
- <https://expertsystem.com/it/covid-19-insights-feelings/>



COVID-19: Insights & Feelings

26 maggio 2020

Un paese diviso dalle offese reciproche

22 maggio 2020

Più tranquillità, ma senza ottimismo

19 maggio 2020

Sentiment meno negativo, ma ancora avvolto nella tristezza

12 maggio 2020

L'economia abbatte il sentiment degli Italiani

8 maggio 2020

Cresce l'irritazione per le condizioni della ripresa

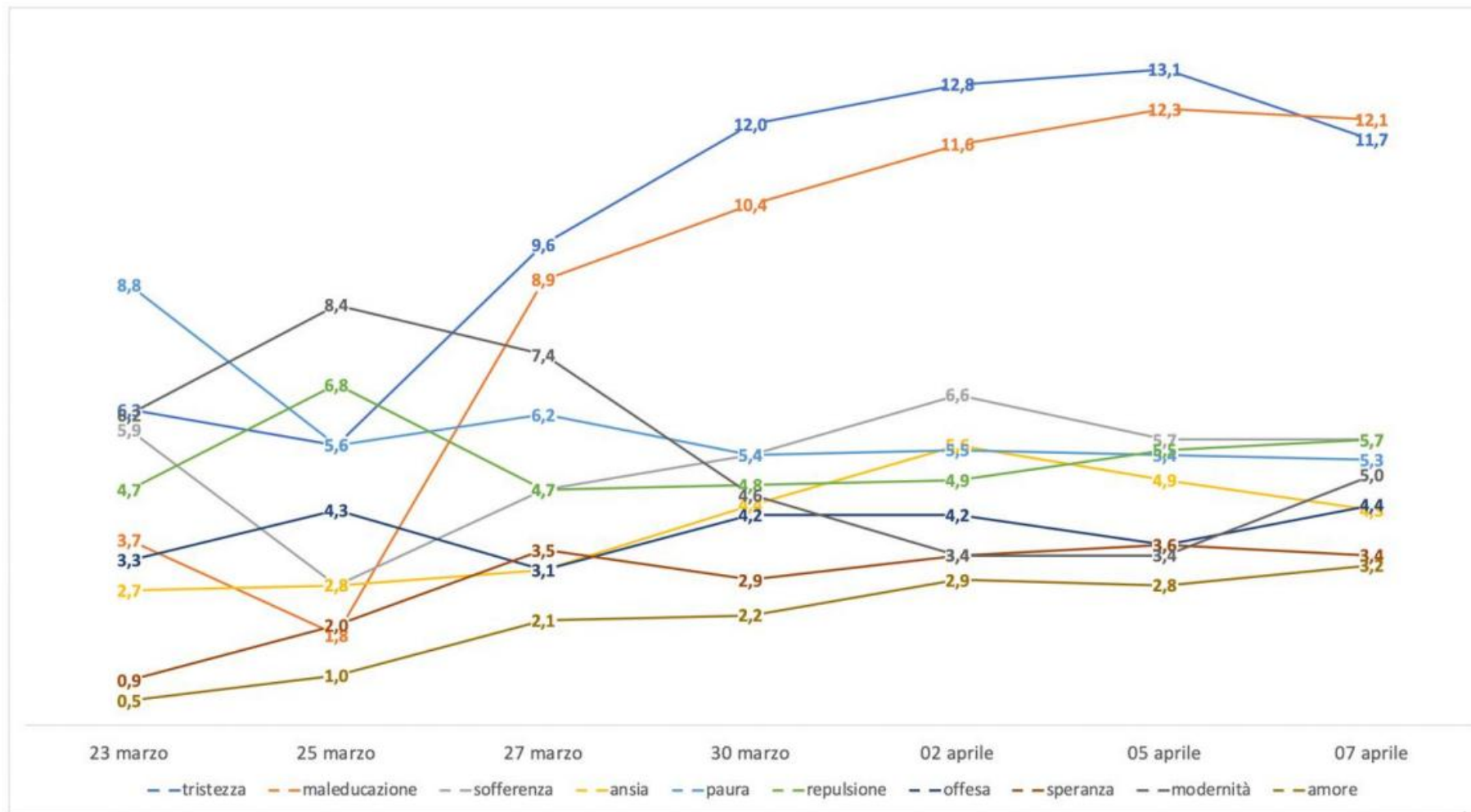
5 maggio 2020

Una finestra di ottimismo



COVID-19: Insights & Feelings

Andamento delle emozioni principali degli Italiani rispetto al Covid-19 (23 marzo – 7 aprile)

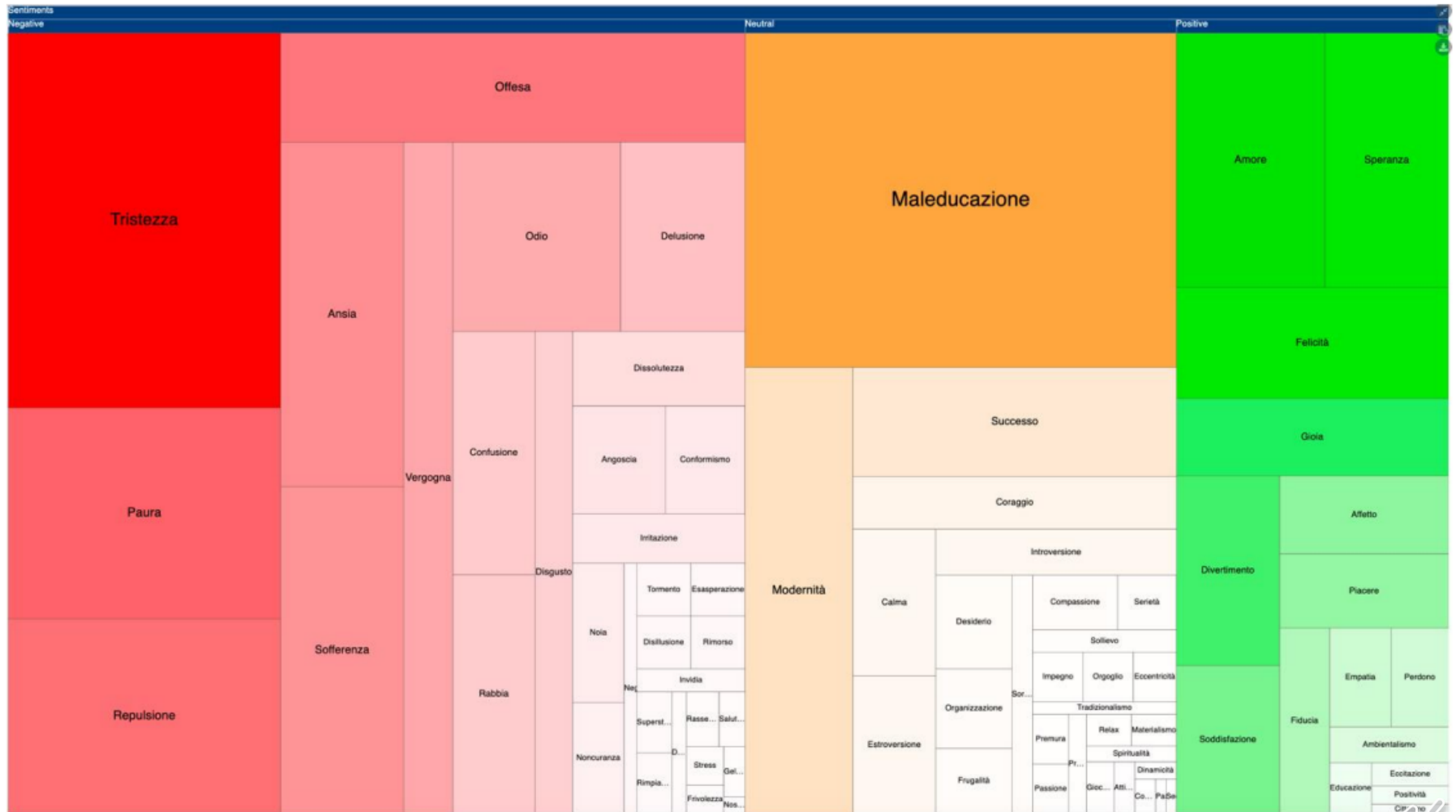


Fonte: Expert System, Sociometrica, 7 aprile 2020



COVID-19: Insights & Feelings

Mappa delle emozioni principali degli Italiani rispetto al Covid-19



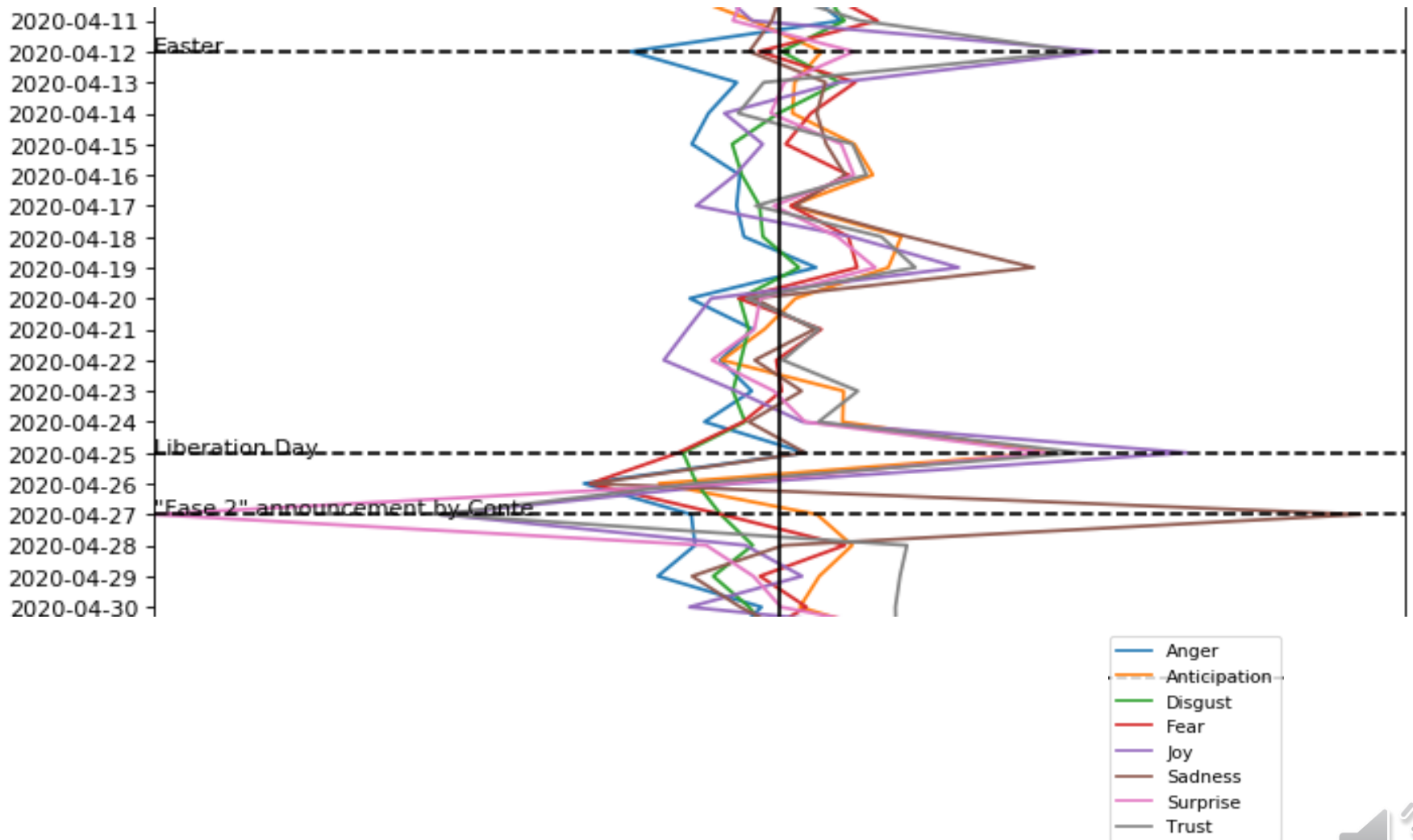
Fonte: Expert System, Sociometrica, 26 maggio 2020

Emoltaly

- Analisi automatica e visualizzazione delle emozioni dei tweet scritti dagli italiani dall'inizio del 2020
- 30.000-50.000 tweet in italiano al giorno presi casualmente
- Approccio lexicon-based usando l'NRC Emotion Lexicon
- Emozioni della ruota di Plutchik
- <http://corpora.ficlit.unibo.it/Emoltaly/>



Emoltaly



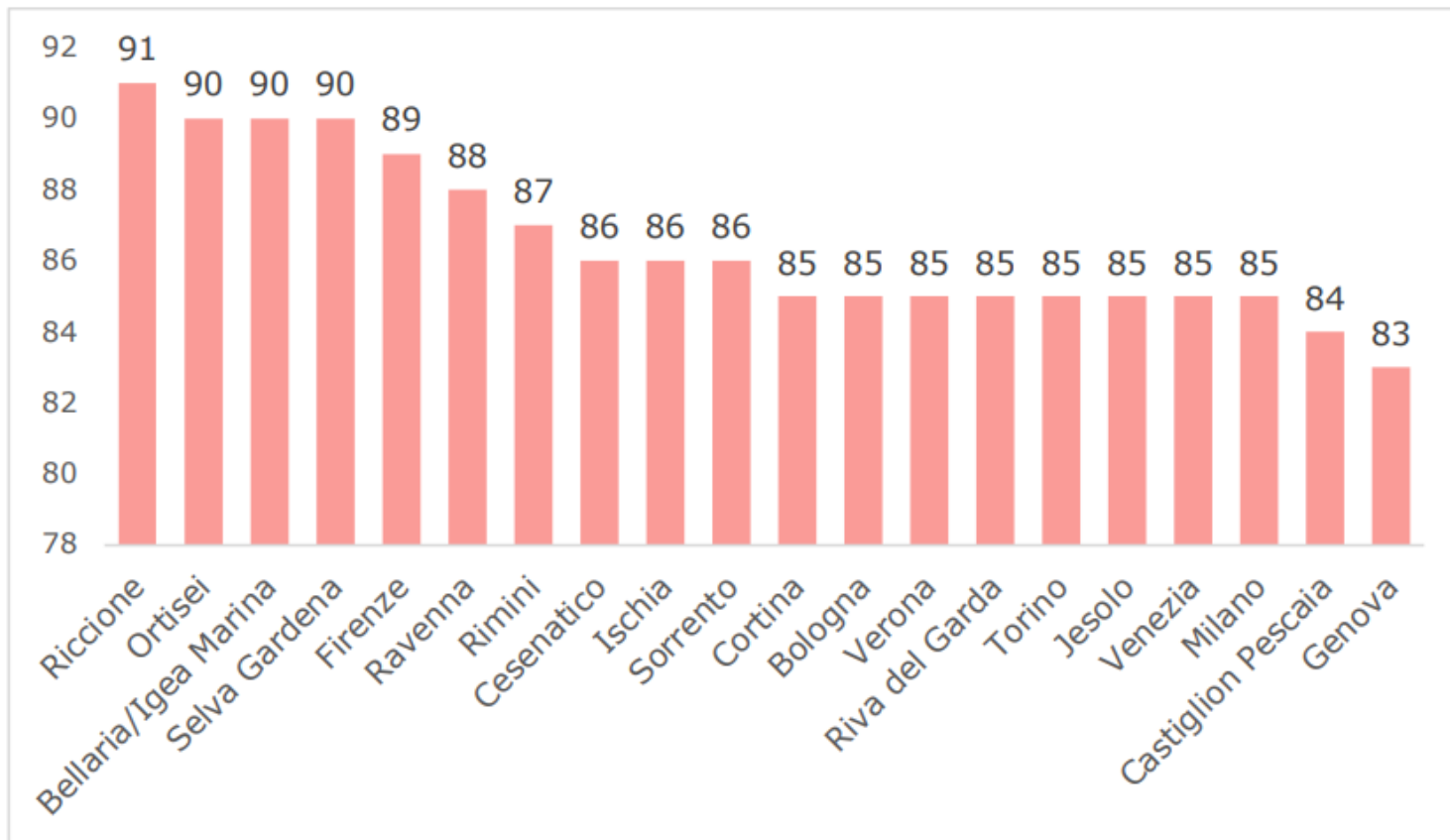
ENJOY: Rapporto sulla Percezione dell'Italia Turistica

- Progetto congiunto di Expert System (azienda di tecnologie semantiche e intelligenza artificiale), Sociometrica (azienda che si occupa di analisi economica e sociale) e Tripadvisor
- 135.000 recensioni in lingua inglese per hotel, ristoranti e attrazioni di 35 destinazioni turistiche italiane nel periodo tra gennaio e novembre 2019
- Aspect-based sentiment analysis
- 90 sentiment diversi
- <https://www.sociometrica.it/sites/default/files/enjoyReportFinale.pdf>



ENJOY: Rapporto sulla Percezione dell'Italia Turistica

Tab. 6 – Ranking delle destinazioni turistiche italiane

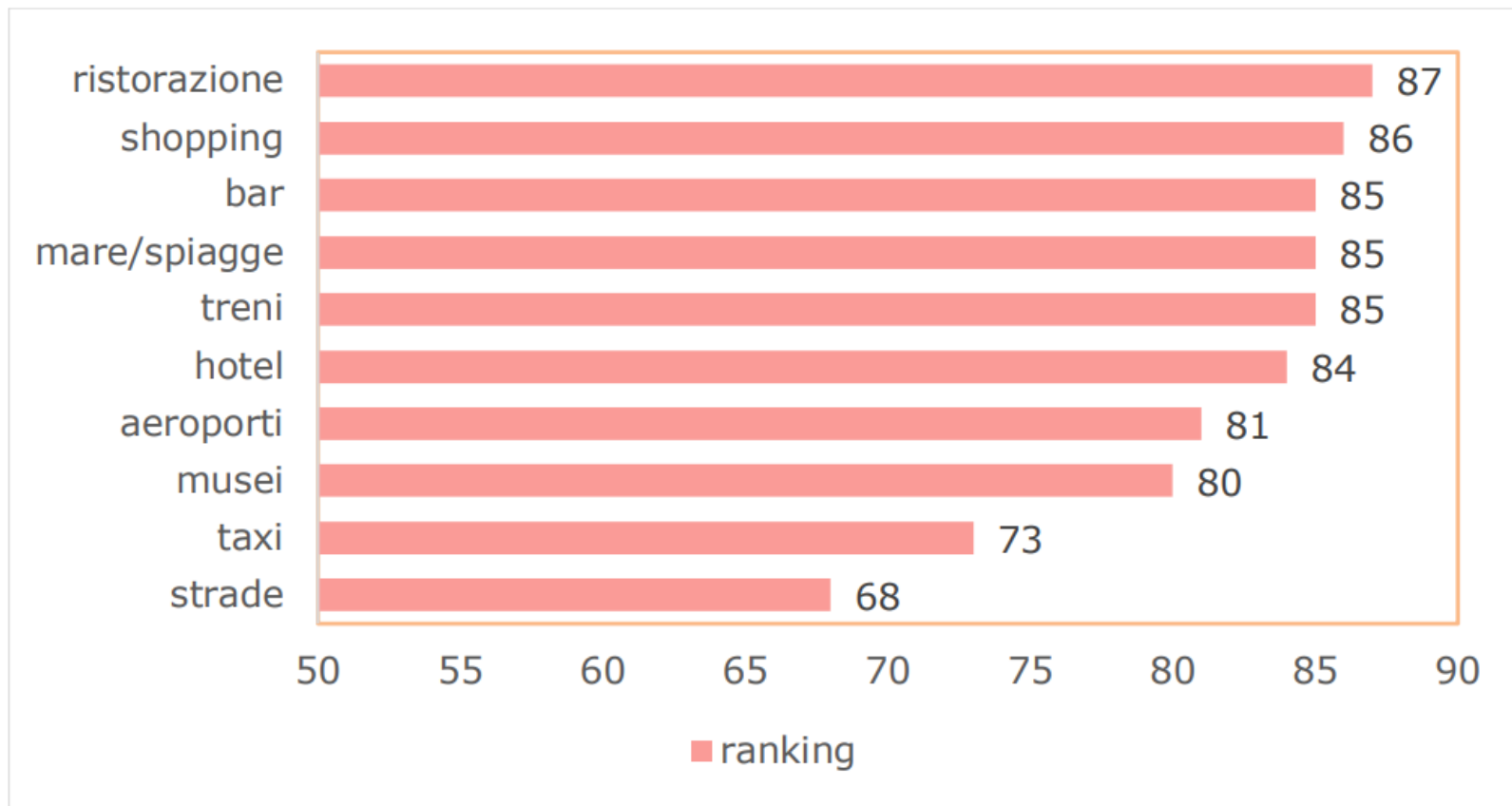


Fonte: semantica Expert System, elaborazione Sociometrica delle recensioni TripAdvisor, 2020



ENJOY: Rapporto sulla Percezione dell'Italia Turistica

Tab. 2 – Sentimento verso gli elementi principali dell'offerta



Fonte: semantica Expert System, elaborazione Sociometrica delle recensioni TripAdvisor, 2020



ENJOY: Rapporto sulla Percezione dell'Italia Turistica

Tav. 3 – sentimenti positivi e negativi prevalenti a Riccione

great good nice friendly
excellent recommend enjoy
amazing fantastic clean wonderful
perfect lovely love delicious
comfortable highly thank beautiful
special best fun fresh close like free
incredible quiet available spacious high
fabulous superb worth super tasty
amaze happy high quality enjoyable
knowledgeable exceptional improve
awesome greet reasonable pleasant
simple perfectly interesting outstanding too
trouble smile amenities ok safe ability
convenient hot easily

problem wait expensive
unfortunately terrific wrong
not recommend poor slow
difficult bad no recommending
cheap crowded miss not great
negative disappoint not good fall
not best not like downside can't
recommend grumpy low limited don't
like mistake not unique not better
can't highly not nice sad dirty
complain not efficient complaint
without best unsafe unpleasant
unfriendly rusty touristy too sweet
lose cheat didn't like injure spoiled
sort out sorry not highly

Fonte: semantica Expert System, elaborazione Sociometrica delle recensioni TripAdvisor, 2020

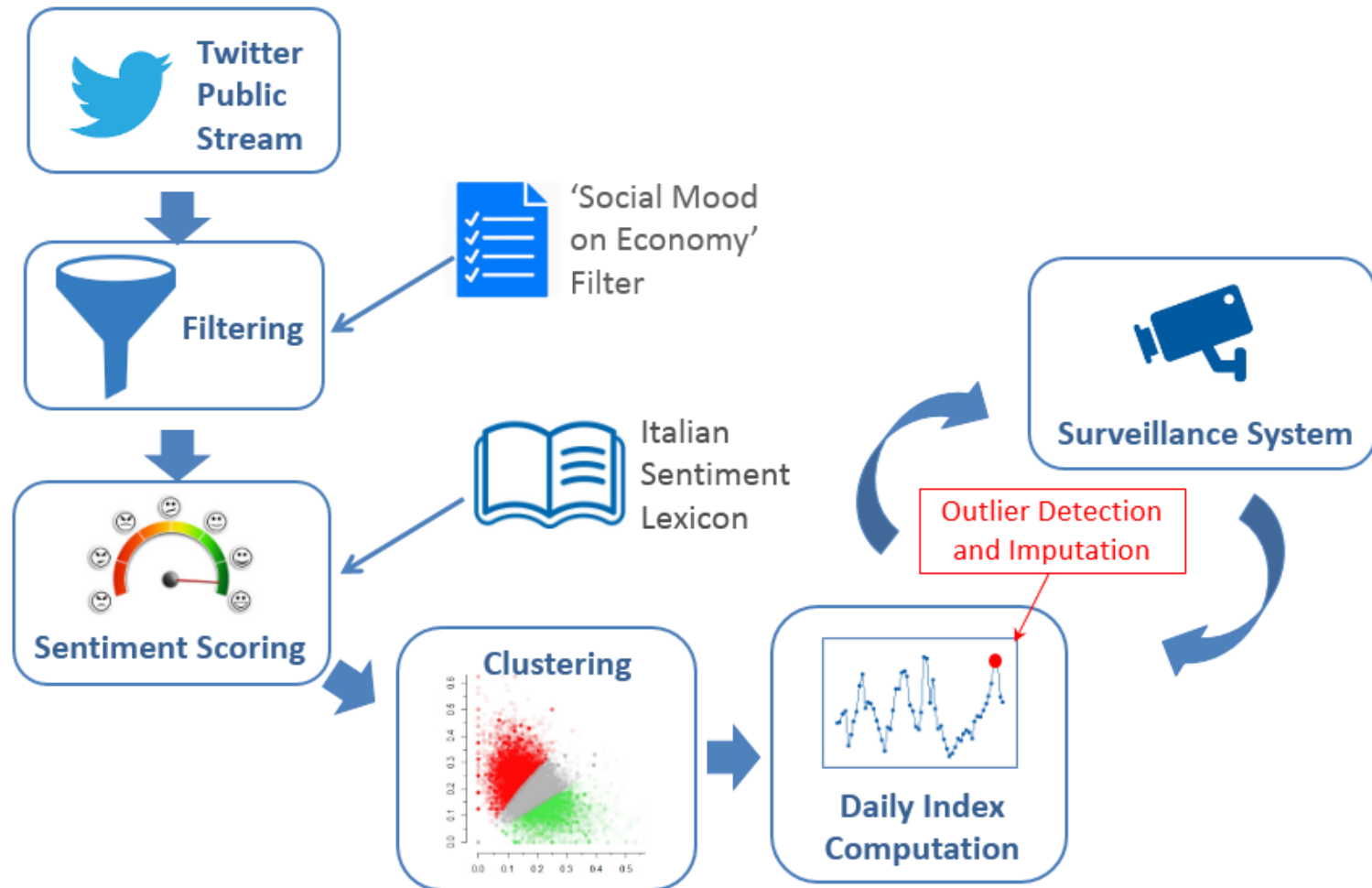


Social Mood On Economy Index

- ISTAT: misure giornaliere del sentiment italiano sull'economia, derivate da campioni di tweet pubblici in lingua italiana, catturati in streaming
- Scraping di circa 55.000 tweet al giorno in base a 60 keyword
- Pre-processing: conversione in minuscolo, correzione ortografica, rimozione URL, caratteri speciali, stop words, stemming
- Approccio lexicon-based: Sentix lexicon
- <https://www.istat.it/it/archivio/219585>



Social Mood On Economy Index

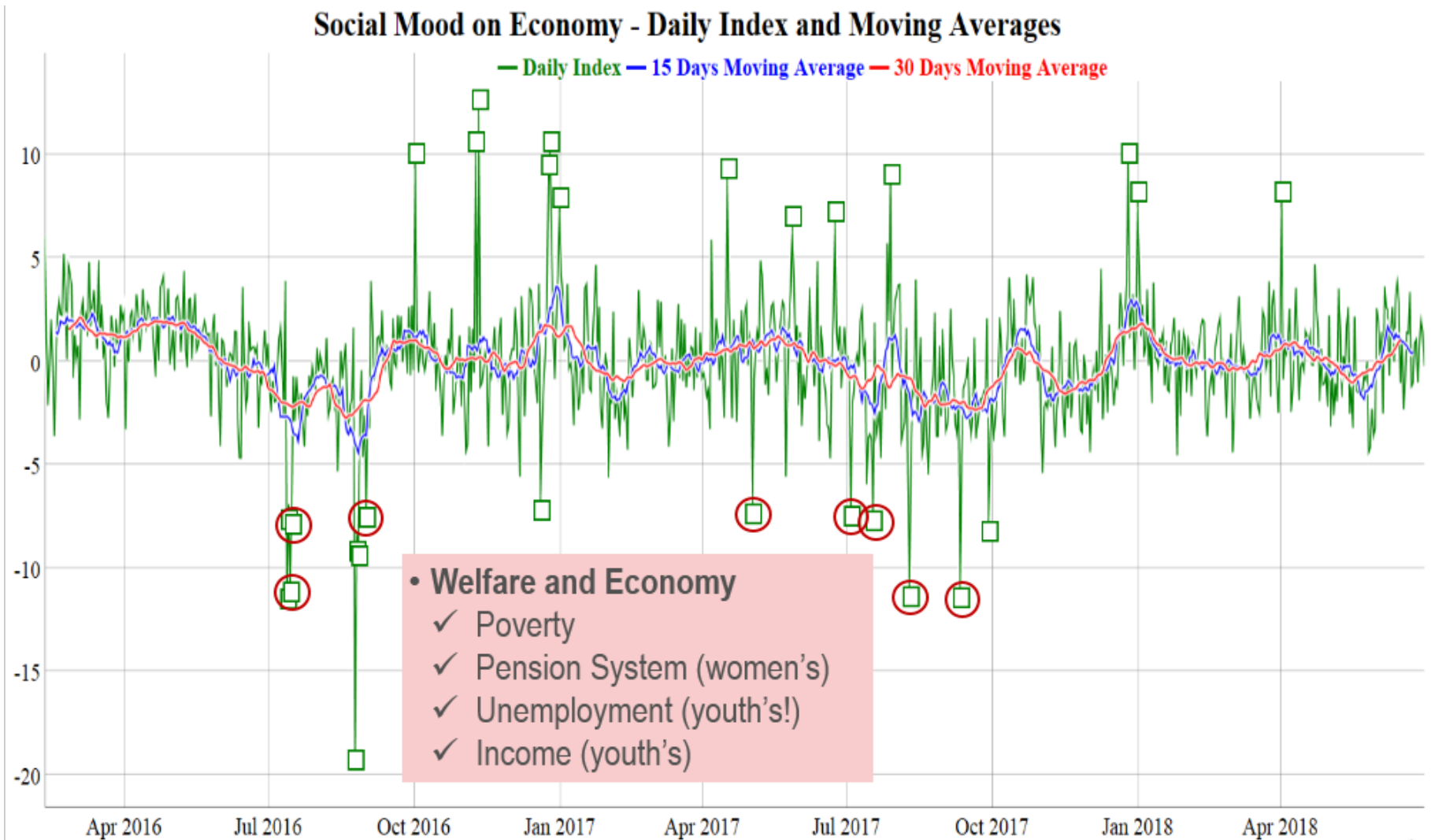


Fonte: Diego Zardetto



Social Mood On Economy Index

Social Mood on Economy - Daily Index and Moving Averages

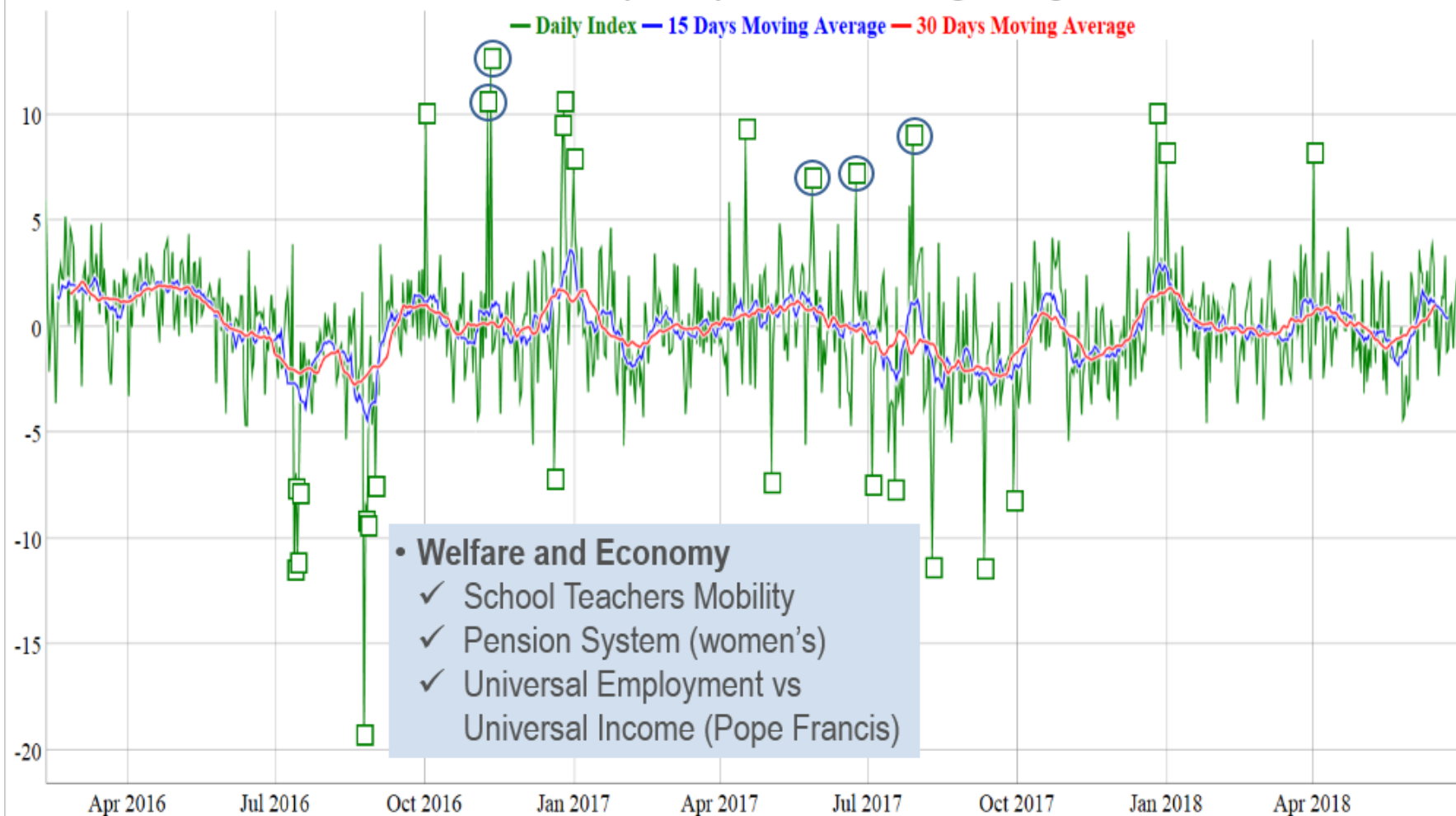


Fonte: Diego Zardetto



Social Mood On Economy Index

Social Mood on Economy - Daily Index and Moving Averages



Fonte: Diego Zardetto



Mappe dell'Intolleranza

- Progetto in collaborazione tra Vox Diritti (osservatorio italiano sui diritti) e l'Università Aldo Moro di Bari (Dipartimento di Informatica)
- Ogni anno per quattro anni, 4 mesi considerati, 200,000 tweet analizzati
- Hate-speech detection
- 6 argomenti: Omofobia, Razzismo, Antisemitismo, Sessismo, Disabilità, Islamofobia
- Ultima analisi (2019): <http://www.voxdiritti.it/la-nuova-mappa-dellintolleranza-4/>



Mappe dell'Intolleranza



XENOFOBIA, LA VERA PIAGA ITALIANA

CONTRO I MIGRANTI SI MOLTIPLICANO I TWEET INTOLLERANTI E DISCRIMINATORI. ODIANO DUE ITALIANI SU TRE TRA COLORO CHE TWITTANO

Un boato di odio, che non si placa, ma aumenta con il tam tam della politica. I migranti sono al primo posto tra le categorie di persone più colpite dallo hate speech, in netta crescita rispetto all'anno passato. E tra le città più intolleranti, spicca Milano.

Il grafico mostra la quantità di tweet raccolti per ogni singolo cluster:

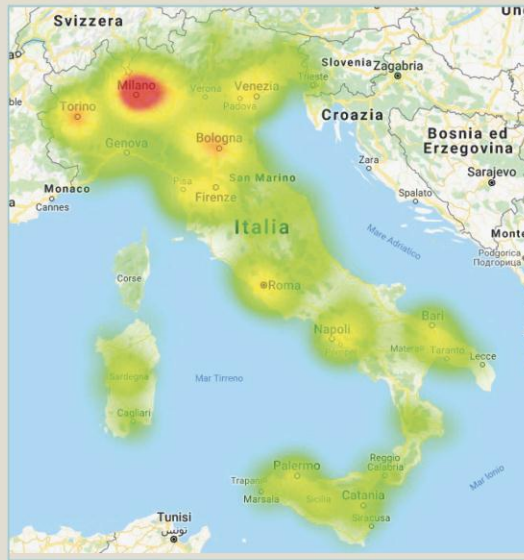
- ANTISEMITISMO
- RAZZISMO
- DISABILITÀ
- MISOGINIA
- ISLAMOFOBIA
- OMOFOBIA

Tweet sui migranti
74.451*

TWEET
215.377
TOTALI
ESTRATTI**

* Totale dei tweet rilevati, contenenti le parole sensibili, relative al cluster Razzismo

** Totale dei tweet estratti nei periodi di rilevazione con valenza sia positiva che negativa.



COME SI LEGGE LA MAPPA TERMOGRAFICA

Le mappe sono state costruite grazie al software Open StreetMap: sintetizzano la diffusione di tweet intolleranti attraverso delle termografie. Quanto più caldo (vicino al rosso) è il colore, tanto più concentrato è il numero di tweet intolleranti in quella zona.



I numeri del fenomeno

5,1 milioni

CITTADINI
STRANIERI
presenti in Italia
nel 2017

Circa l'**8,5%**
della popolazione
italiana

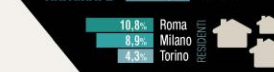
CONTRO
7,9 MILIONI della Francia
6,0 MILIONI della Spagna
12,2 MILIONI della Germania

Numero di migranti
nel mondo nel **2017**
258 MLN

RISPETTO ALLA POPOLAZIONE **8,5%**



ISCRITTI ALLA
ANAGRAFE



RESIDENTI

I RISULTATI

2019



32% sul totale
dei tweet negativi rilevati
22.043
geolocalizzati

Periodo di rilevazione:
marzo 2019 - maggio 2019

LE PAROLE INTOLLERANTI

Nella scelta delle parole da mappare, abbiamo lavorato sui termini e le offese più ricorrenti sui social, evidenziati anche nelle ricerche scientifiche, che si sono occupate di studiare i meccanismi implicati nell'atteggiamento discriminatorio.

Terrone • zingaro • negro • muso giallo
rumeno • albanese • muso da scimmia
bangla • mangia rane • kebabbaro • crucco
rabbino • sbarcato • mangia banane

NAZIONALITÀ:
1.190.001 Romania
415.531 Marocco
290.681 Cina
237.047 Ucraina
167.859 Filippine

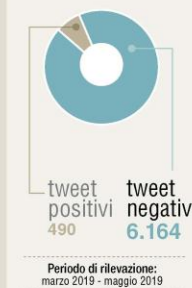
Casi di discriminazione in ITALIA
nel 2016 **69,4%**
nel 2017 **82,9%**

Nel 2018 in Italia
70%

EPISODI
di intolleranza
e razzismo verso gli
immigrati

23,4% pensa che gli
immigrati
in Italia siano
troppi
35,6% ha paura di
subire
REATI
50,9% ha problemi
economici ed è
insoddisfatto

IL FATTORE AGGRESSIVITÀ



Periodo di rilevazione:
marzo 2019 - maggio 2019

Il grafico mostra la quantità di messaggi aggressivi, rilevati secondo un'ulteriore metodica di analisi: si è proceduto a estrarre i gruppi di parole più ricorrenti nei messaggi intolleranti, al di là di insulti e parole offensive. Il software è stato infatti "istruito" per estrarre i tweet più aggressivi, evidenziandone il livello di virulenza: la valutazione è stata orientata dalle categorie utilizzate dalla scala MOAS (Modified Overt Aggression Scale). Tale metodica, in questa sua prima forma sperimentale, si è dimostrata utile per meglio comprendere non solo la negatività, gli atteggiamenti intolleranti e discriminatori, ma anche l'orientamento aggressivo dei messaggi.

xenofobia

VOX



UNIVERSITÀ
DEGLI STUDI
DI MILANO



SAPIENZA
UNIVERSITÀ DI ROMA



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

In collaborazione con
ISIME



Mappe dell'Intolleranza

XENOFOBIA

OMOFOBIA

