

16 Maggio 2019

Cittadini di età minore e welfare locale

Enrico Aiello



Cineca - Chi siamo

CINECA è un Consorzio Interuniversitario, costituito nel 1969 dal MIUR per il calcolo a supporto della ricerca scientifica.
Organizzazione senza scopo di lucro.

Consoziati:

- MIUR
- 70 Università
- 8 Enti di Ricerca

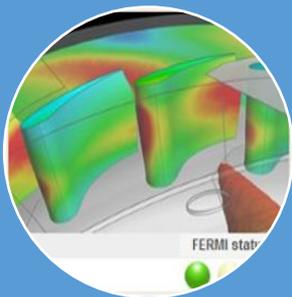
In house provider:

Vigilato dal MIUR, i consorziati esercitano un controllo analogo congiunto
Nessuna partecipazione di capitali privati
80% delle attività è effettuata verso i consorziati

Dipendenti: circa 830



Le aree di attività del consorzio Cineca



Supporto alla ricerca e all'innovazione

- High Performance Computing e Data Analytics
- Tier 0 Nazionale e Europeo
- Visualizzazione Scientifica e Ambienti interattivi Virtuali
- Ricerca e sviluppo, formazione e training



Servizi alle Università e al MIUR

- Soluzioni e Servizi per l'Amministrazione Universitaria
- Servizi per il Ministero Istruzione Università e Ricerca
- Innovazione dei work flow, dematerializzazione, sicurezza



Trasferimento Tecnologico

- Sperimentazione numerica, prototipizzazione virtuale
- Problem solving mediante modellistica numerica
- Analisi predittiva e machine learning
- Soluzioni e Servizi per la Sanità e la Ricerca Clinica

Indice

- Bigdata (definizione, ciclo di vita, etica)
- Giovani e social media
- Applicazione Analytics (Text mining - sentiment analysis - topic detection)

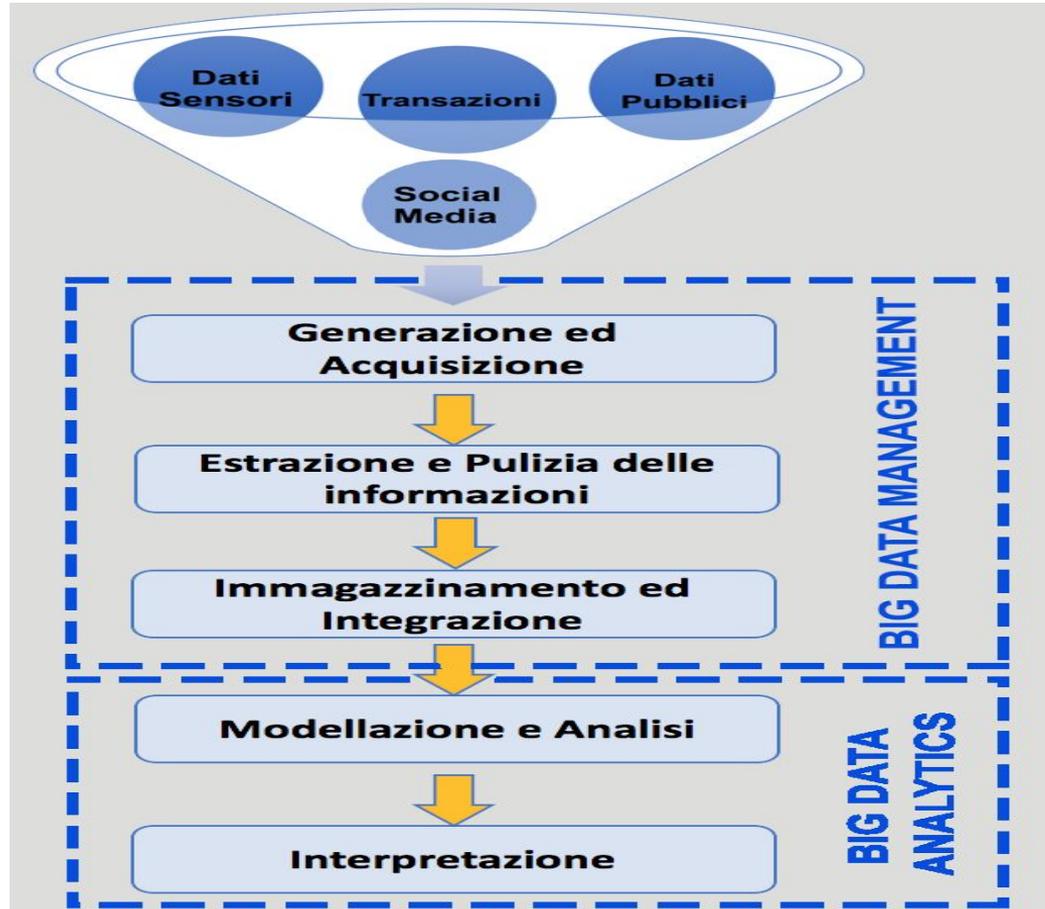
Big Data - definizione e caratteristiche

- Si parla di Big Data quando si ha un insieme talmente grande e complesso di dati che richiede la definizione di nuovi strumenti e metodologie per estrapolare, gestire e processare informazioni entro un tempo ragionevole.
- Non esiste una dimensione di riferimento, ma questa cambia sempre, poiché le macchine sono sempre più veloci e i dataset sono sempre più grandi.

Big Data - Caratterizzazione



Big Data - Ciclo di vita



Big Data - Etica

- Lo sviluppo di metodologie di acquisizione e di elaborazione di grandi masse di dati tramite algoritmi coinvolge tutti i settori economici ha sollevato dubbi sulla loro applicazione in mancanza di adeguate normative e controlli.
- Nel 2008, in seguito della crisi economica mondiale, due ingegneri finanziari, Emanuel Derman e Paul Wilmott, hanno elaborato un manifesto etico per gli scienziati dei dati, sul modello del Giuramento di Ippocrate per i medici
- Nel 2017 la data scientist Cathy O'Neil ha esposto in un libro le varie problematiche sorte con l'uso dei modelli basati su Big Data. Molti dei quali, lungi dall'essere equi ed obiettivi, si sono dimostrati codificazioni di pregiudizi umani che hanno portato ad errori sistemici senza possibilità di appello nei software che controllano le nostre vite in diversi ambiti, da quello legale a quello lavorativo e politico

Internet as data source



European
Commission

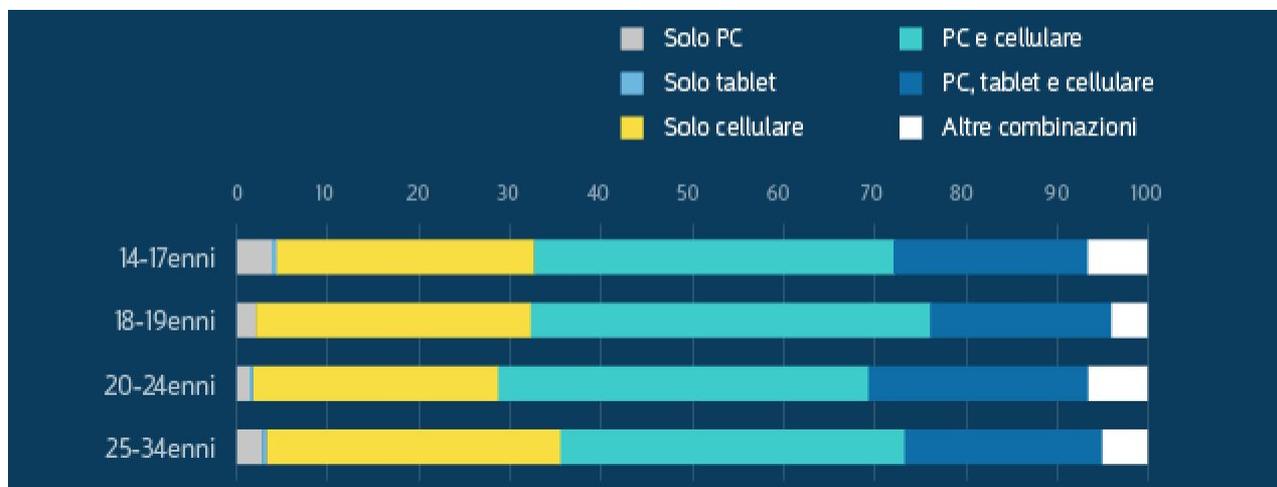
<http://www.dialogic.nl/documents/2010.080>

Internet as data source

Feasibility Study on Statistical Methods on Internet as a
Source of Data Gathering

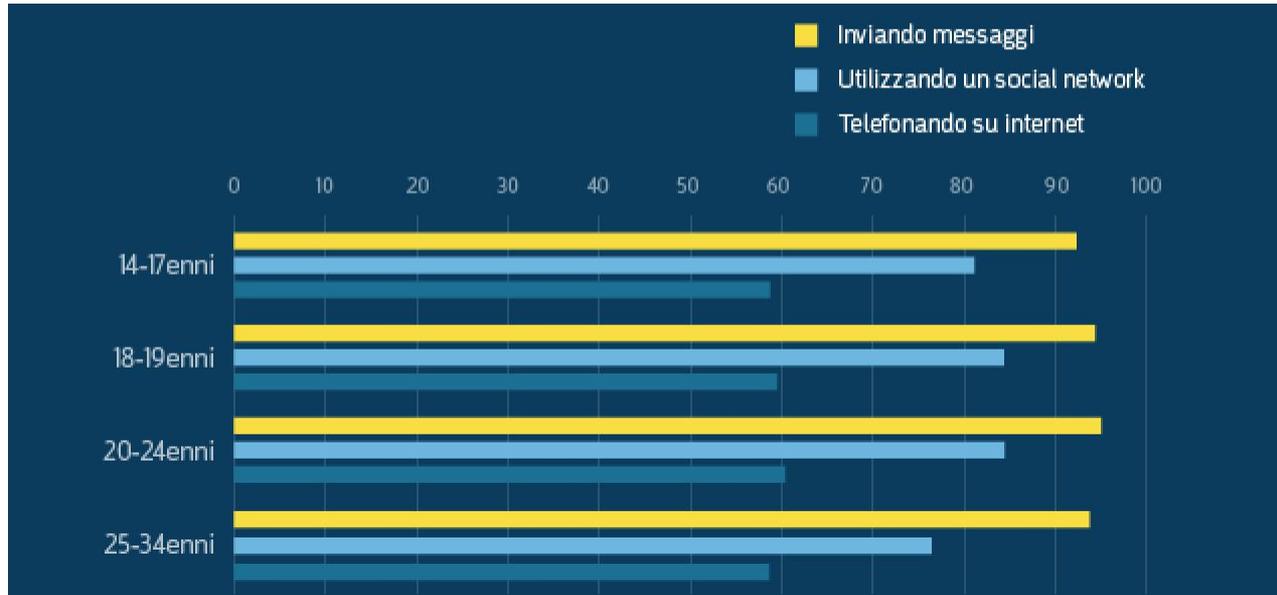
Giovani e social media

Come si connettono?



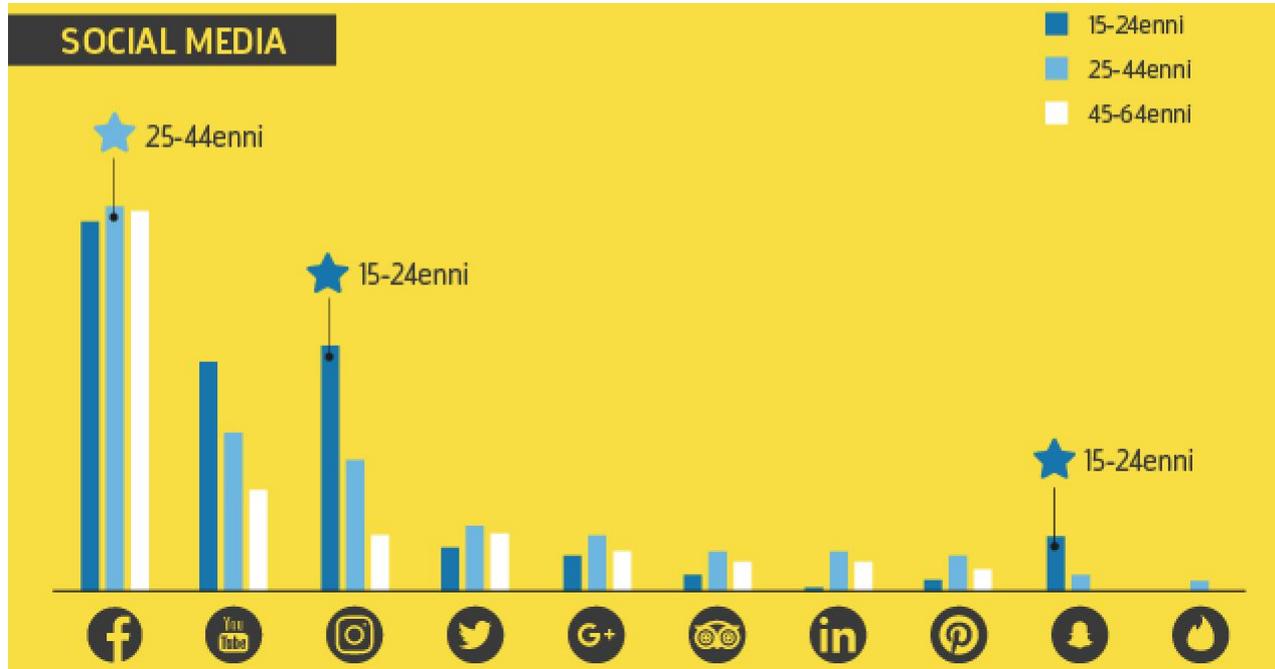
Giovani e social media

Come comunicano?



Giovani e social media

Utilizzo social media in base all'età



Perché i giovani si stanno allontanando da facebook e utilizzano instagram?



Applicazione: Twitter Analysis

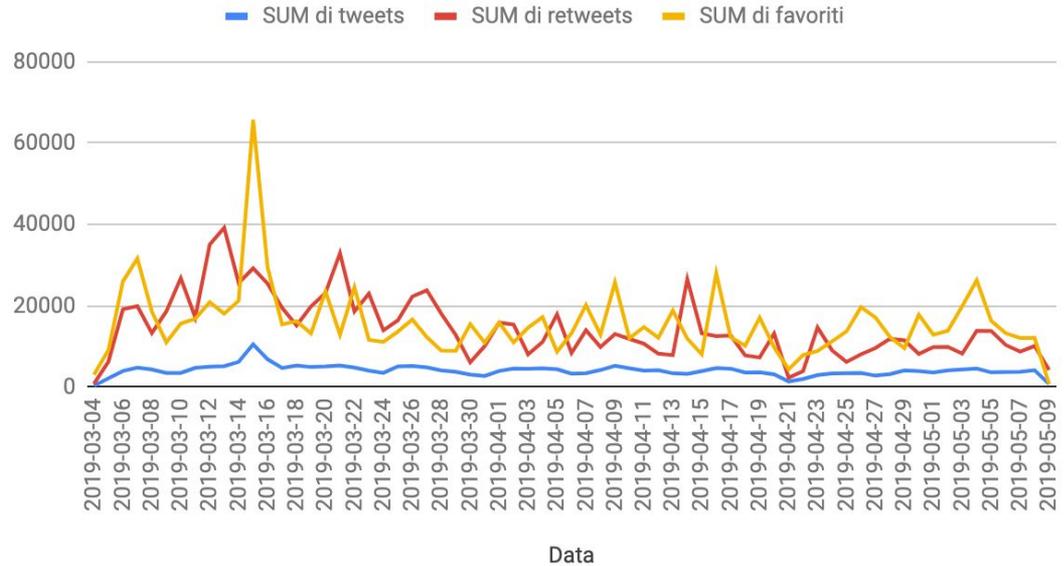
Implementato un algoritmo che estrae ed analizza i tweets riguardanti le seguenti parole chiave:

- Scuola
- Bullismo
- Giovani
- Gioventù
- Adolescenti

Twitter analysis - Volumi totali

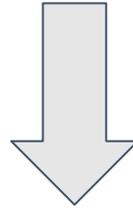
Nr. tweets	274.519
Nr. retweets	974.032
Nr. favoriti	10.194.116
AVG tweets per giorno	circa 4.100
AVG retweet per giorno	circa 28.700
AVG favoriti per giorno	circa 34.000

Tweets e interazioni



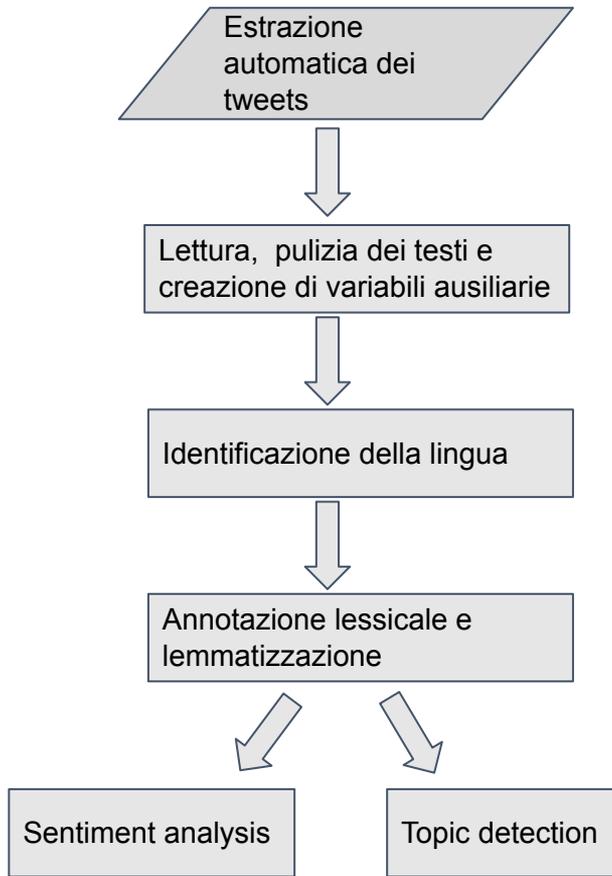
Text Mining - Analisi automatica dei testi

La grande mole di tweets estratti non rende agevole una valutazione “a vista” delle informazioni.



E' pertanto necessaria l'analisi automatica per poter acquisire conoscenza da questo tipo di basi dati, soprattutto in ottica di valutazione in tempo reale.

Come funziona l'algoritmo



Sentiment Analysis

L'obiettivo principale della sentiment analysis è determinare la polarità generale di un documento (sia che si tratti di una recensione, di un commento ad un post e via dicendo), ossia classificare un documento o frase in positiva, negativa o neutrale.

Sentiment Analysis

- Consiste nell'estrazione e analisi delle opinioni che gli utenti esprimono nel web rispetto a diversi temi per misurarne la percezione.
- Si riferisce all'elaborazione del linguaggio e all'analisi del testo per identificare informazioni soggettive nelle fonti.

Identificazione automatica dei argomenti più discussi

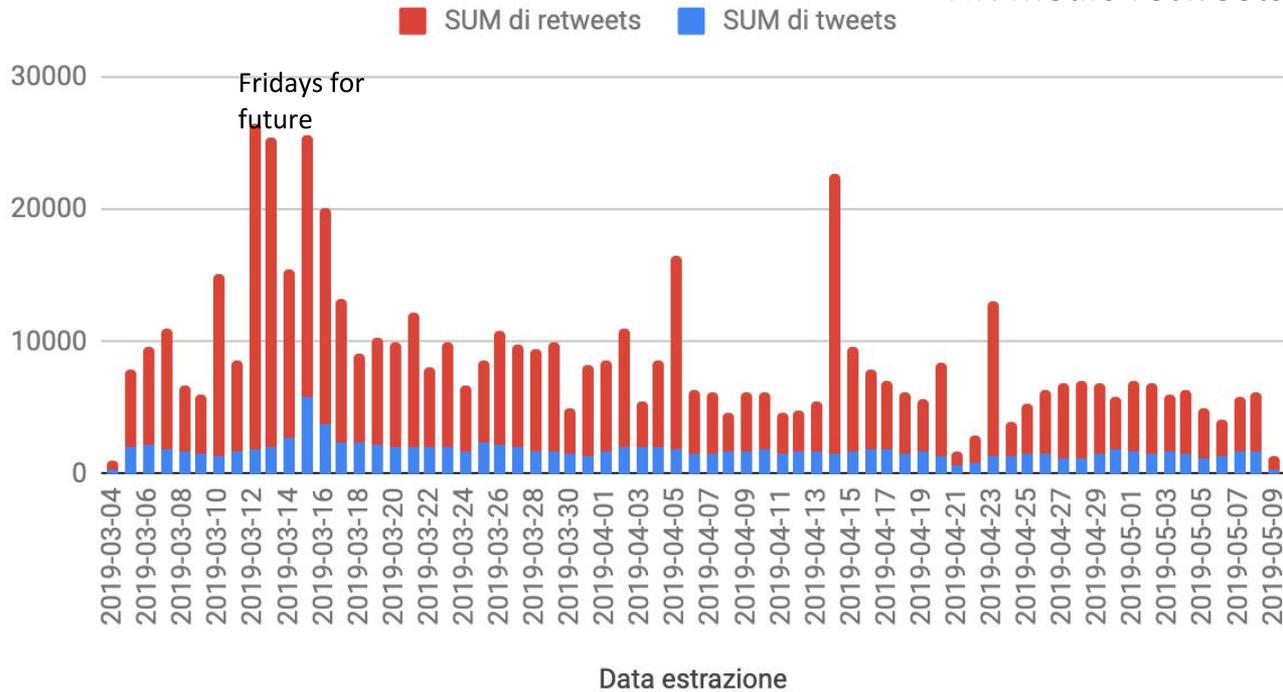
- Attraverso algoritmi di machine learning opportunamente addestrati è possibile estrapolare automaticamente gli argomenti più discussi nelle fonti dati considerate.
- Maggiore è il grado di addestramento degli algoritmi maggiore è il livello di accuratezza dei risultati.

Parole cercate: giovani - gioventù - adolescenti

Nr. tweets e interazioni per giorno

Nr. medio tweets per giorno: 1780

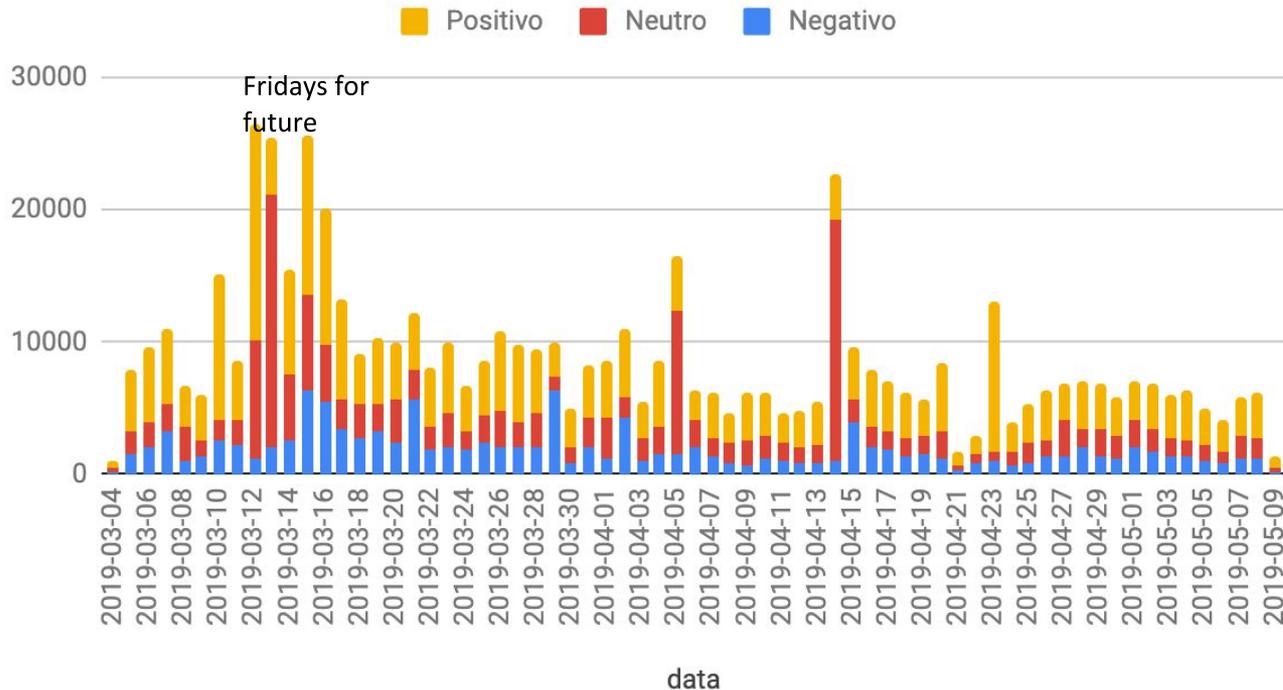
Nr. medio retweets per giorno: 7000



Sentiment analysis

Parole cercate: giovani -
gioventù - adolescenti

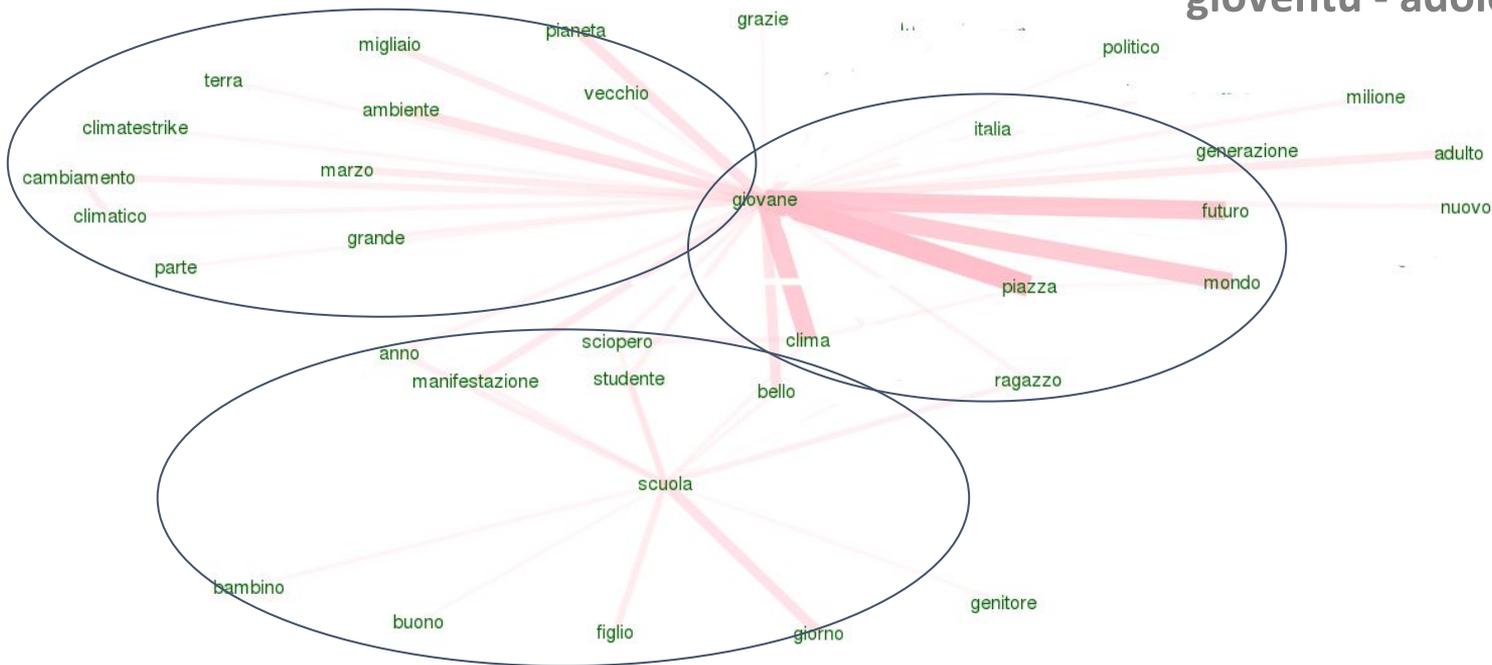
Andamento giornaliero della polarità



Argomenti discussi - Es. Analisi giornaliera

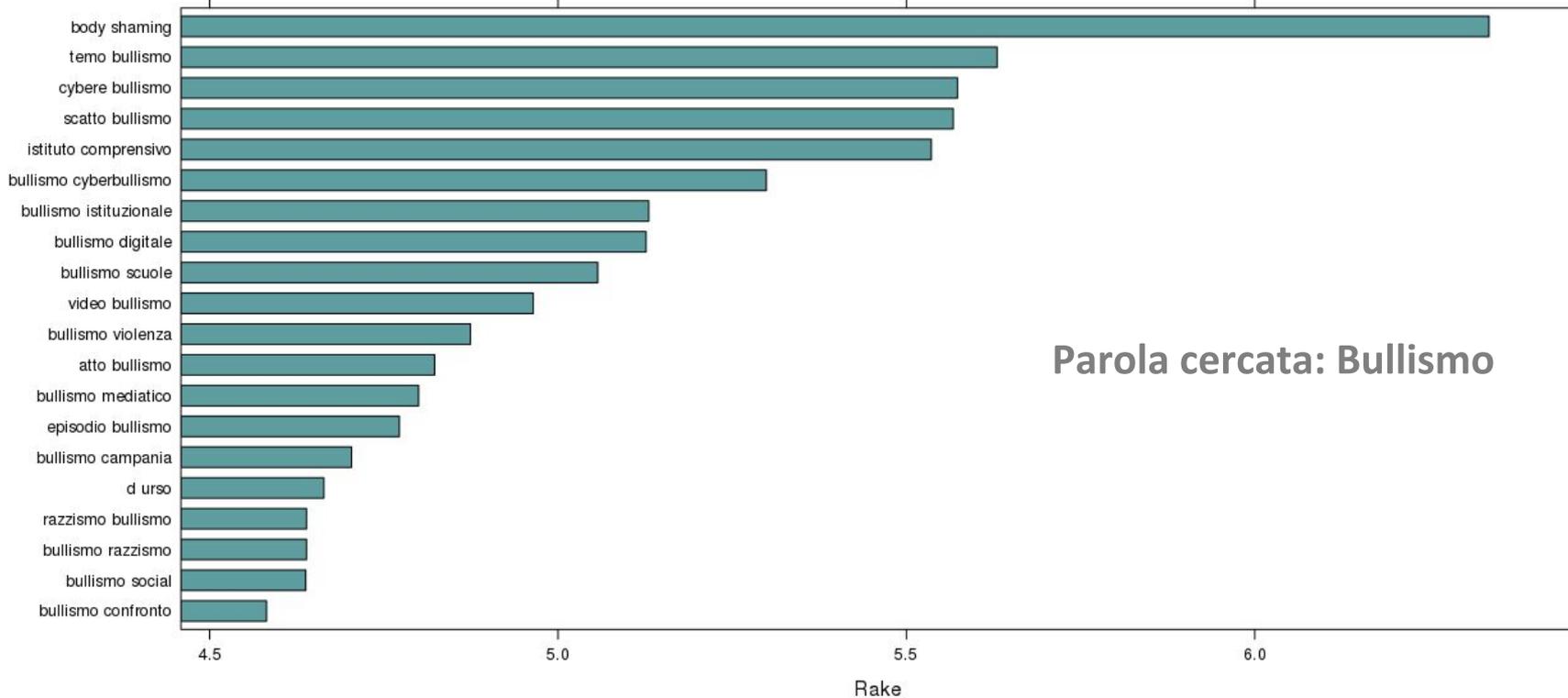
Cooccurrences within sentence

Nouns & Adjective



Identificazione automatica di parole chiave

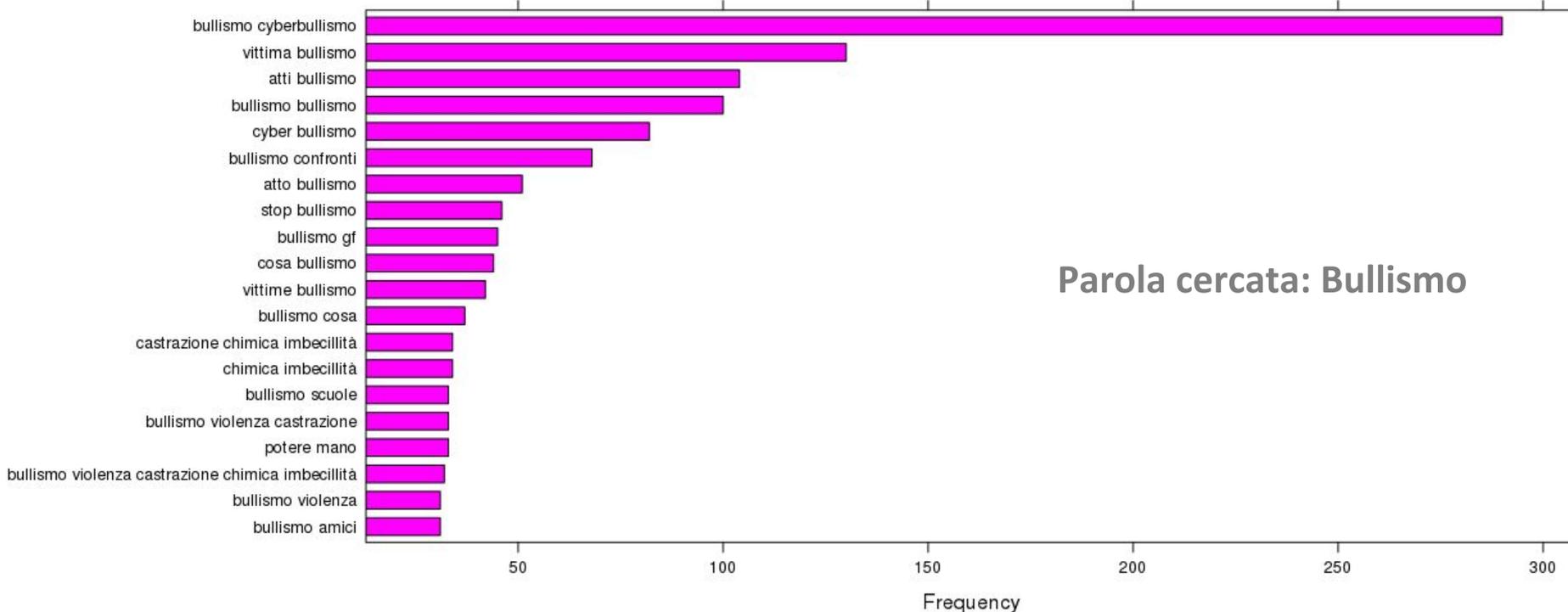
Keywords identified



Parola cercata: Bullismo

Identificazione automatica di parole chiave

Keywords - simple noun phrases



Grazie per l'attenzione

e.aiello@cineca.it