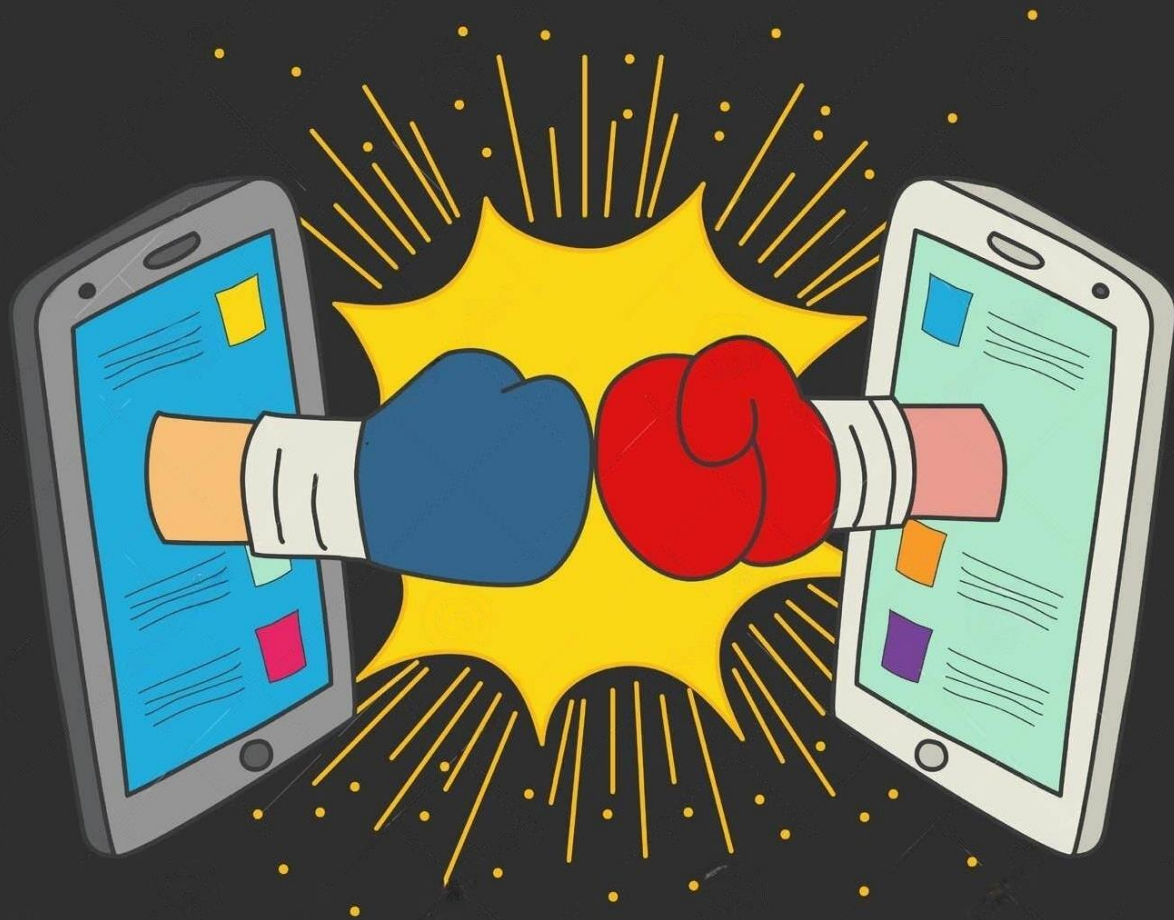


# DISCUSSIONI ITALIANE SU REDDIT: ANALISI DI DINAMICHE E SENTIMENT

ZARANTONELLO MASSIMO E FRANCESCA PULERÀ

ANNO ACCADEMICO 2023/2024



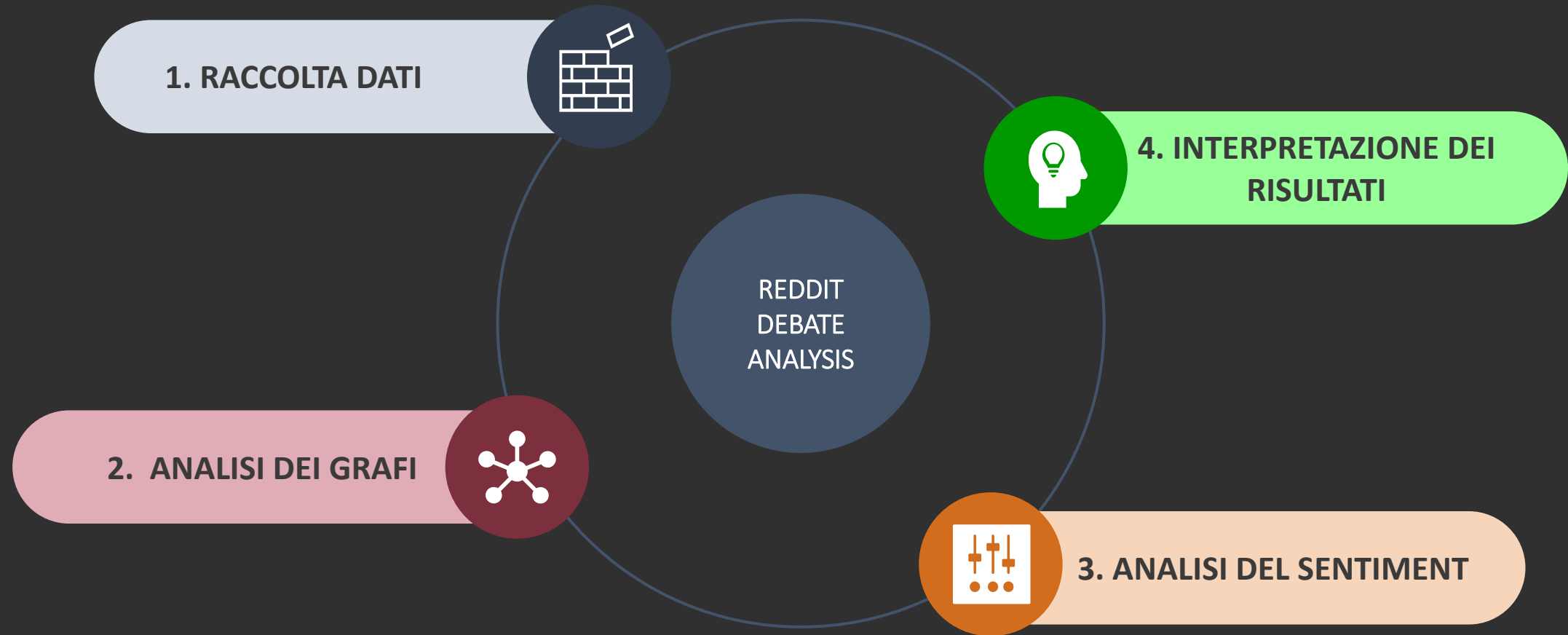


QUANTO TENDONO A  
DISCUTERE GLI  
ITALIANI SUI SOCIAL?

CHE CARATTERE  
HANNO QUESTE  
DISCUSSIONI?

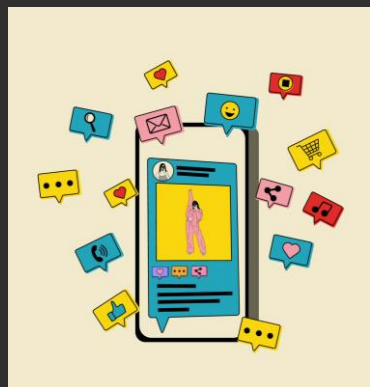
QUALI SONO GLI  
ARGOMENTI  
MAGGIORMENTE  
DIBATTUTI?

**OBIETTIVO:** Analizzare la propensione degli italiani a discutere sotto post riguardanti temi di attualità e caratterizzare la natura di queste discussioni.



# RACCOLTA DATI

Realizzazione di un dataset a partire dalla raccolta dei commenti di interesse su Reddit. Verranno effettuate analisi preliminari per definire le proprietà fondamentali del dataset, essenziali per i successivi task.

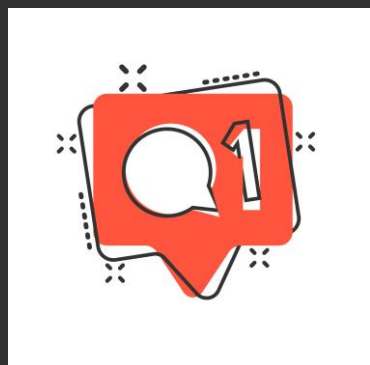


### 1. SELEZIONE DEI POST

Diversi **temi di attualità** con qualche post anche su argomenti più leggeri, al fine di confrontarli e comprendere che cosa suscita maggiore discussione.

### 2. SCELTA DELLE FEATURE

Tramite PRAW, una libreria Python che permette di interagire con l'API di Reddit, sono state selezionate le **feature di interesse**.

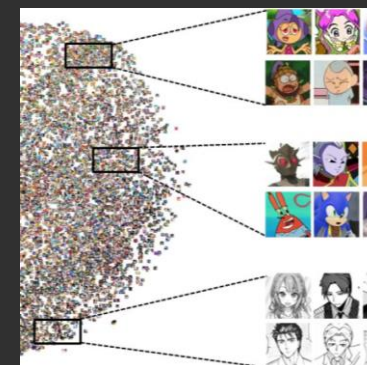


### 3. ESTRAZIONE DEI COMMENTI

Sono stati scaricati sia i **commenti principali** sotto i post di interesse, sia i **commenti innestati**, ovvero le risposte ai commenti.

### 4. CREAZIONE DEL DATASET

È stato dunque creato un dataset **ad hoc** composto da **4207 istanze** relative ad **9 post** di interesse.





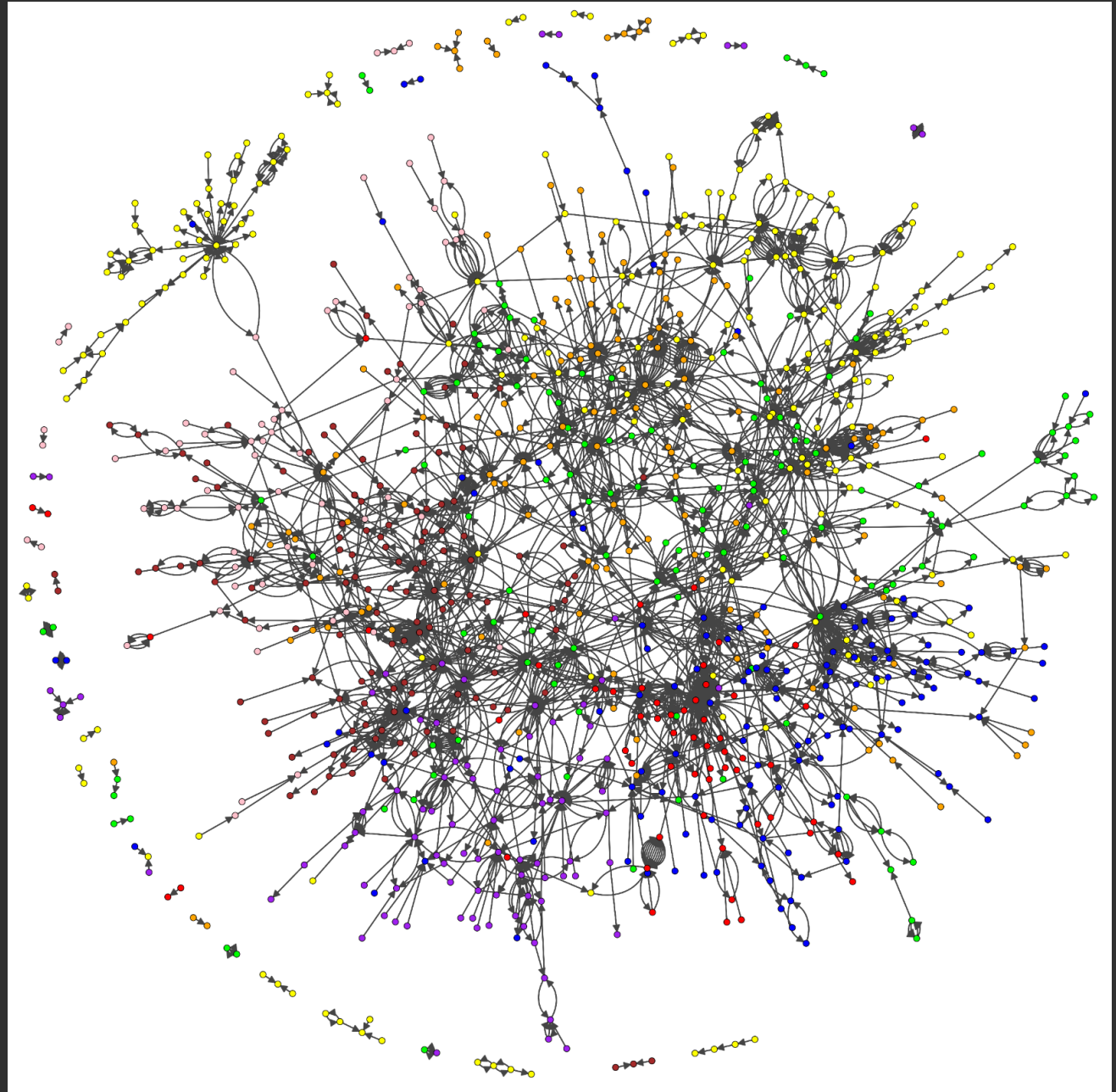
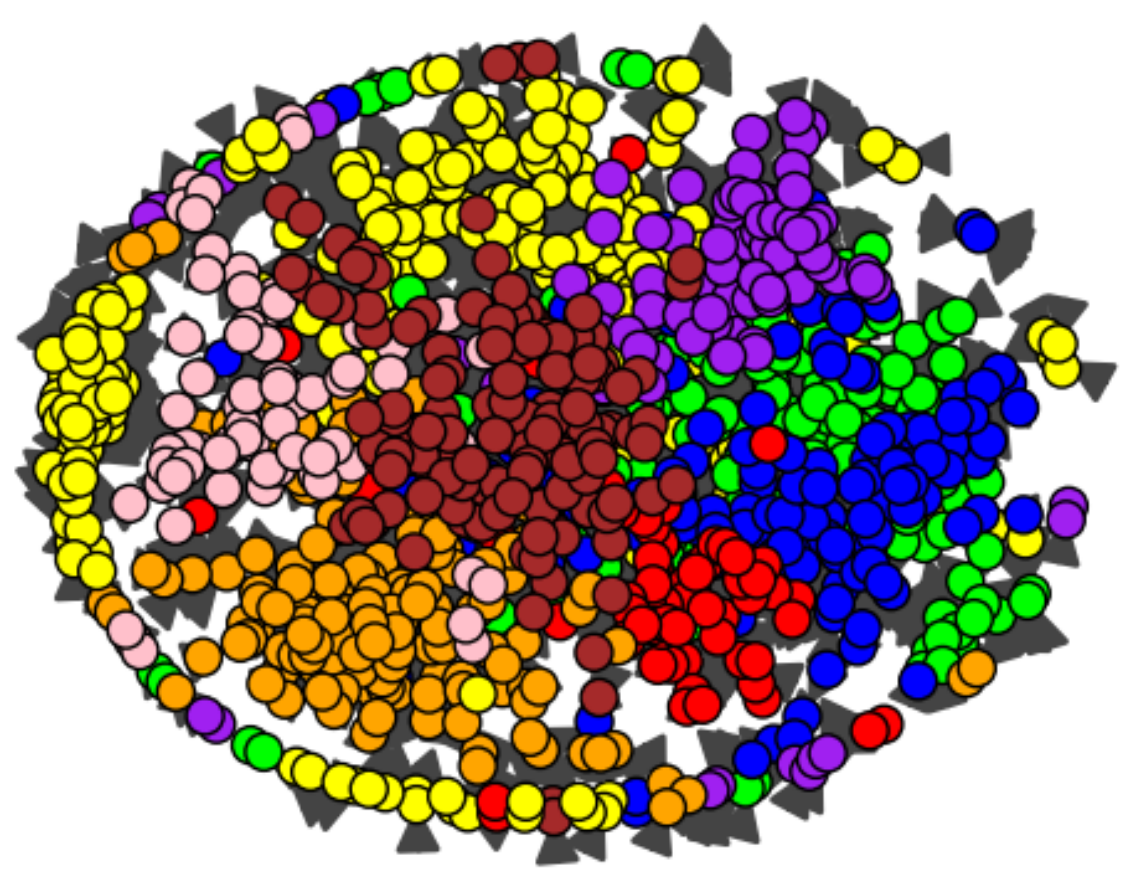
id	author_name	parent_name	score	comment_body	post_id	post_title
j7geyvi	dnxpb64	OP	336	Secondo me ha a che fare con il fatto che la maggior parte della gente è scettica in questo ambito, e quindi non vuole avere a che fare con cibo a base di insetti.	10v8sey	Natura sì: niente cibo a base di insetti/sintetici. Non rientrano nella nostra definizione di salute
ktawo2v	00ishmael00	OP	327	il cliente da lei chiamato non è al momento raggiungibile...	1b6cg4q	Possibile guerra NATO - RUSSIA. Se davvero dovessero succedere e vi chiamassero al fronte per combattere ...
l6ld98z	OnlyRugrats	OP	316	No way. Fino a qualche anno fa i ragazzi per non emigrare venivano a Milano. Ora che Milano è una cloaca, un misto tra un suq e Gotham city, andarsene via è l'unica scelta possibile. Non c'è volontà politica di sistemare la questione. Chi ci governa ha già nuova mano d'opera a basso costo da impiegare nelle loro realtà.	1d5h5h6	Boom di giovani in fuga dall'Italia, sono oltre un milione
k9xjbzd	deleted	OP	310	Non è un idiota l'ex ragazzo, è un assassino, punto.	17z2hci	Ho visto troppi post sulla Giulia Cecchettin
kikcanu	emblemadelp	OP	269	Quello bianco è il pulsante che avvia la Guerra Termonucleare Globale?	19aeo2k	Eccovi il mio gatto
kxtj91m	Tonieland	OP	254	Fatto tra di noi questo resta un discorso tra "poveracci", nel senso che a noi gente normale tra un po' ci tagliano le palle se facciamo un peto più del solito mentre ai ricchi e megaricchi (tipo ai 500 che appena finito l'ultimo superbowl sono ripartiti coi loro jet privati super ecologisti), alle grandi multinazionali e ai paesi in via di sviluppo (e non solo) è tutto concesso.	1bulhj9	Il problema del caldo anomalo(causato dal riscaldamento globale) che a nessuno sembra fregare qualcosa...
k7dqj9h	TheNomadologist	OP	203	Ci sono proteste, marce e manifestazioni continuamente in molte parti del mondo. Ce ne sono anche in Israele. Colombia e Cile hanno richiamato i loro ambasciatori da Israele citando i crimini di guerra mentre la Bolivia è andata oltre, interrompendo le reazioni diplomatiche. Ogni organizzazione umanitaria possibile chiama al cessate il fuoco, la mozione dell'assemblea generale delle Nazioni Unite, è passata nonostante l'astensione e il voto negativo in blocco del cosiddetto occidente che nel mondo resta comunque la minoranza. Crescono gli appelli al boicottaggio e alla discesa in campo della corte penale internazionale. Tutto ciò non è neanche lontanamente abbastanza ma quello che manca è la copertura mediatica mainstream.	17lese9	4000 bambini morti in Palestina. Possibile che nessuno faccia niente a livello politico?



# ANALISI DEI GRAFI

# SITUAZIONE INIZIALE

Inizialmente si ha un unico grafo i cui **nodi** rappresentano tutti gli **utenti** che hanno commentato sotto uno dei nove post e gli **archi orientati** le **interazioni** fra di essi.





# COSA STIAMO CERCANDO?

## Le discussioni

## COSA CI SERVE?

1

Definizione per induzione di «discussione»:

### Caso base

- $\text{INdegree}(n)=0 \wedge \text{OUTdegree}(n) = 0 \Rightarrow n$  non è una discussione
- $\text{INdegree}(n)=0 \wedge \text{OUTdegree}(n) \geq 1 \Rightarrow n$  non è una discussione
- $\exists m \in V, n \neq m, (n,m) \in E, (m,n) \in E \Rightarrow \{n,m\}$  è una discussione

### Passo induttivo

Supponiamo che  $D \subseteq V$  sia una discussione.

Se  $u \in V \setminus D, (u, v) \in E, v \in D, \Rightarrow D' = D \cup \{u\}$  è una discussione

2

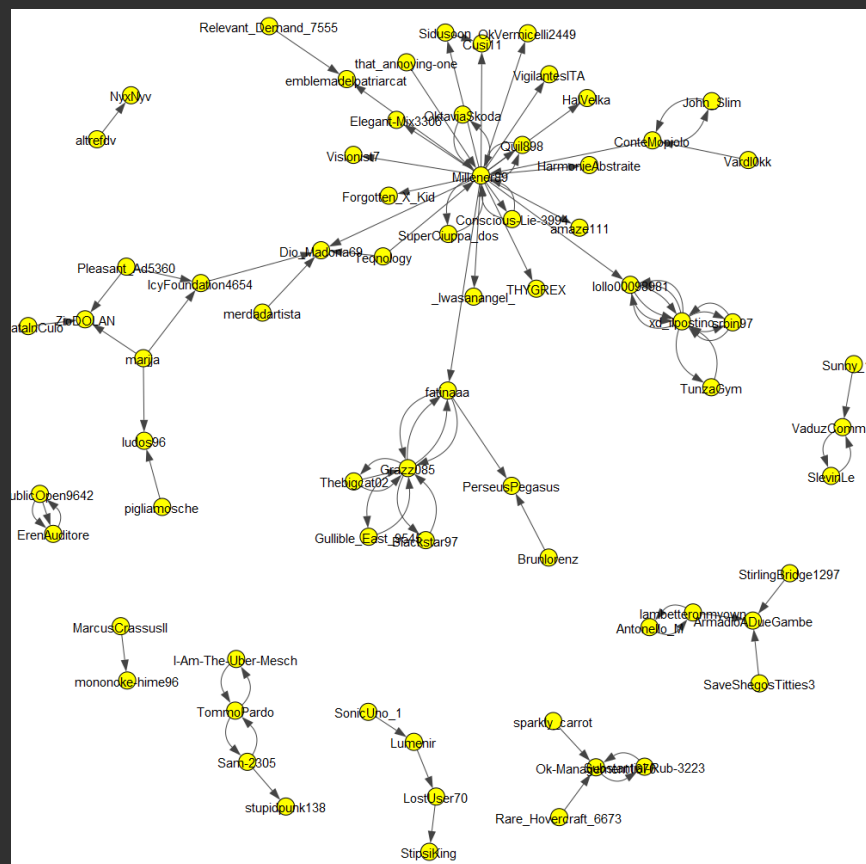
Estrazione dei dataset relativi alle discussioni ed associazione id discussione:

I dibattiti relativi ai vari post sono stati isolati e consolidati in un **nuovo dataset** composto esclusivamente dai nodi pertinenti.

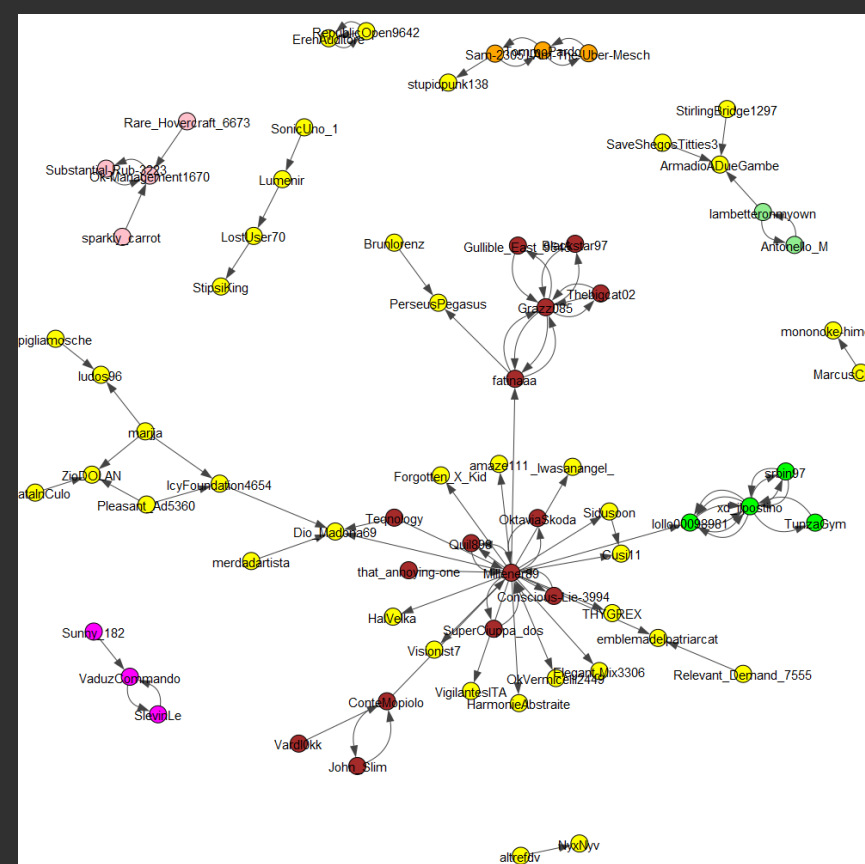
Aggiunto inoltre ***debate\_group***: un id univoco che identifica all'interno di un post la discussione alla quale il commento fa parte

# RISULTATI

# PRIMA



# DOPO



# RISULTATI

## PRIMA

comment_body	post_title
Secondo me ha a che fare con il fatto che la maggior parte della gente è scettica in questo ambito, e quindi non vuole avere a che fare con cibo a base di insetti.	Natura sì: niente cibo a base di insetti/sintetici. Non rientrano nella nostra definizione di salute
il cliente da lei chiamato non è al momento raggiungibile...	Possibile guerra NATO - RUSSIA. Se davvero dovessero succedere e vi chiamassero al fronte per combattere ...
No way. Fino a qualche anno fa i ragazzi per non emigrare venivano a Milano. Ora che Milano è una cloaca, un misto tra un suq e Gotham city, andarsene via e l'unica scelta possibile. Non c'è volontà politica di sistemare la questione. Chi ci governa ha già nuova mano d'opera a basso costo da impiegare nelle loro realtà.	Boom di giovani in fuga dall'Italia, sono oltre un milione

## DOPO

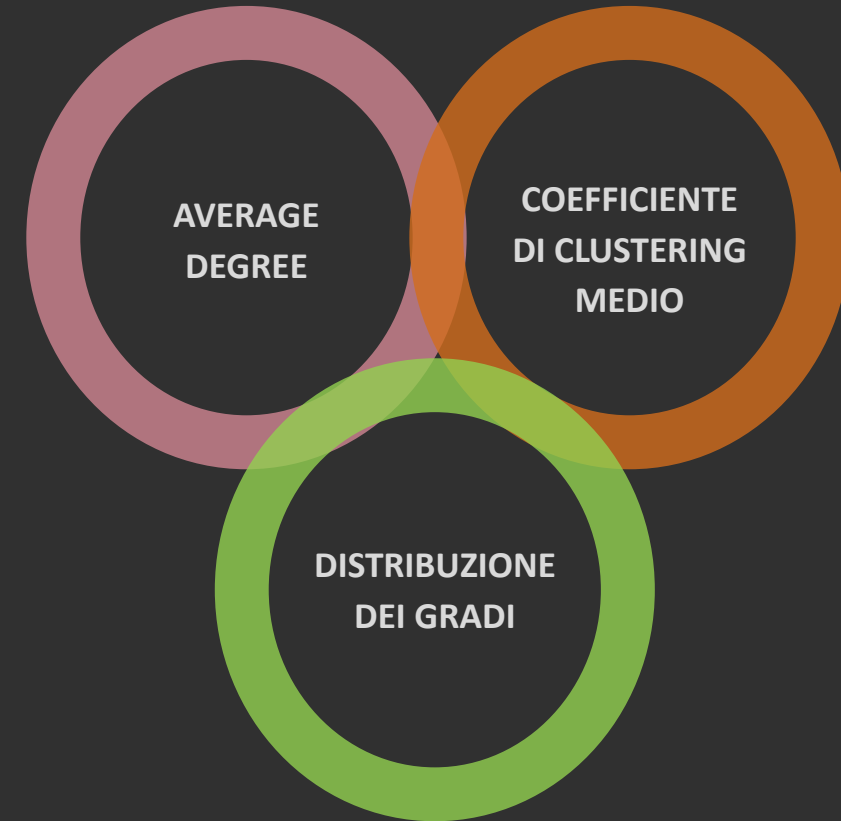
comment_body	post_title	debate_group
Secondo me ha a che fare con il fatto che la maggior parte della gente è scettica in questo ambito, e quindi non vuole avere a che fare con cibo a base di insetti.	Natura sì: niente cibo a base di insetti/sintetici. Non rientrano nella nostra definizione di salute	1
il cliente da lei chiamato non è al momento raggiungibile...	Possibile guerra NATO - RUSSIA. Se davvero dovessero succedere e vi chiamassero al fronte per combattere ...	2
No way. Fino a qualche anno fa i ragazzi per non emigrare venivano a Milano. Ora che Milano è una cloaca, un misto tra un suq e Gotham city, andarsene via e l'unica scelta possibile. Non c'è volontà politica di sistemare la questione. Chi ci governa ha già nuova mano d'opera a basso costo da impiegare nelle loro realtà.	Boom di giovani in fuga dall'Italia, sono oltre un milione	6

# ANALISI CONDOTTA

## METRICHE SUL DIBATTITO

Quanto gli utenti tendono  
ad interagire tra di loro?

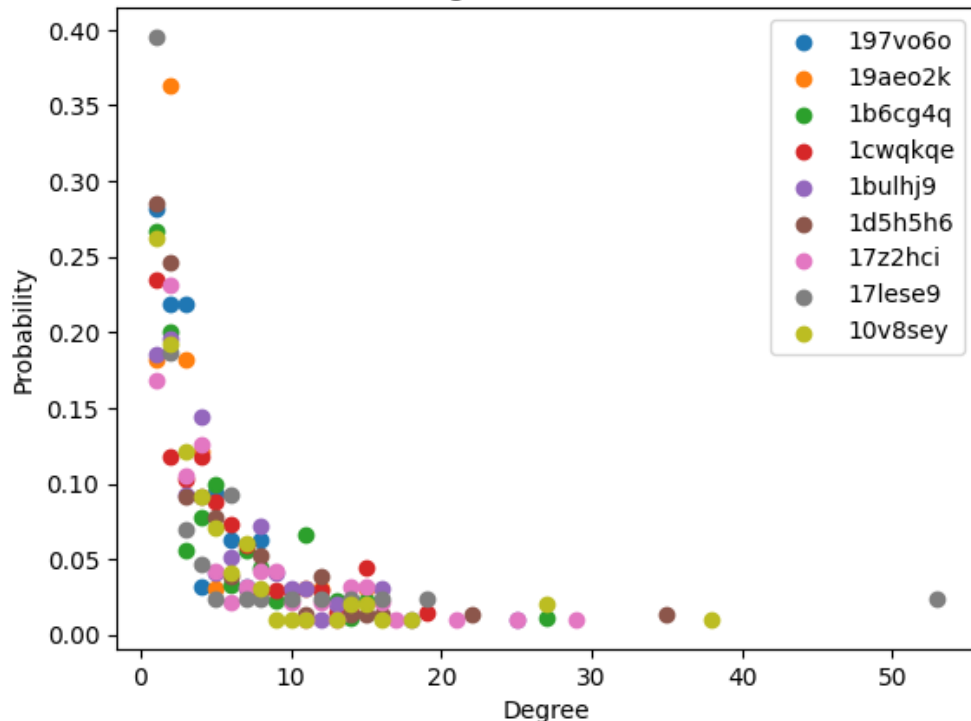
Quanto è denso il dibattito?



## METRICA SUL POST

Qual è la percentuale di utenti  
che hanno interagito  
tanto/poco con il post?

Degree distribution





# ANALISI DEL SENTIMENT

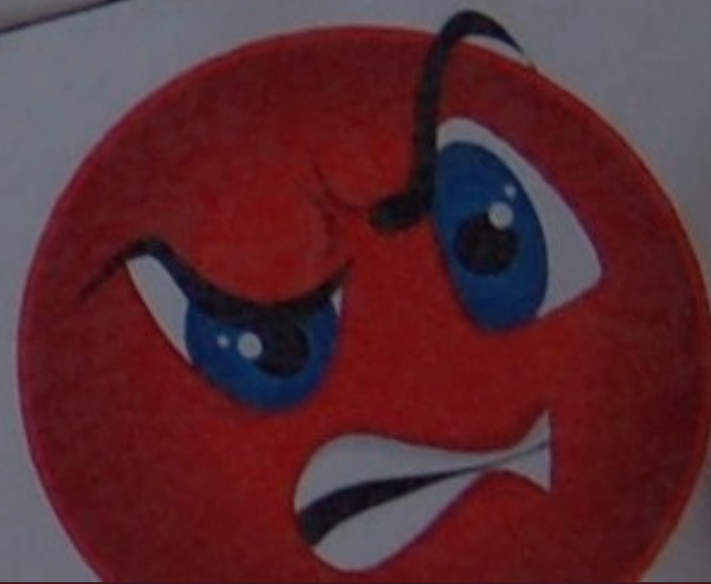


Tristezza



Paura

obia



# PREPROCESSING

Si converte il testo in minuscolo per garantire una rappresentazione uniforme del testo

Esegue una serie di operazioni di pre-processing, come la rimozione di link URL e tag HTML

La punteggiatura non aggiunge necessariamente significato al testo e può essere considerata "rumore"

Conversione in lowercase

Funzione che processa il testo

Rimozione della punteggiatura

Tokenizzazione

Rimozione di stopwords

Lemmatizzazione

Suddivisione del testo in unità più piccole, chiamate token (TweetTokenizer)

Rimozione di parole comuni come articoli e congiunzioni, prive di informazioni specifiche

Riduce ad una radice comune parole che hanno approssimativamente lo stesso significato



## PRIMA

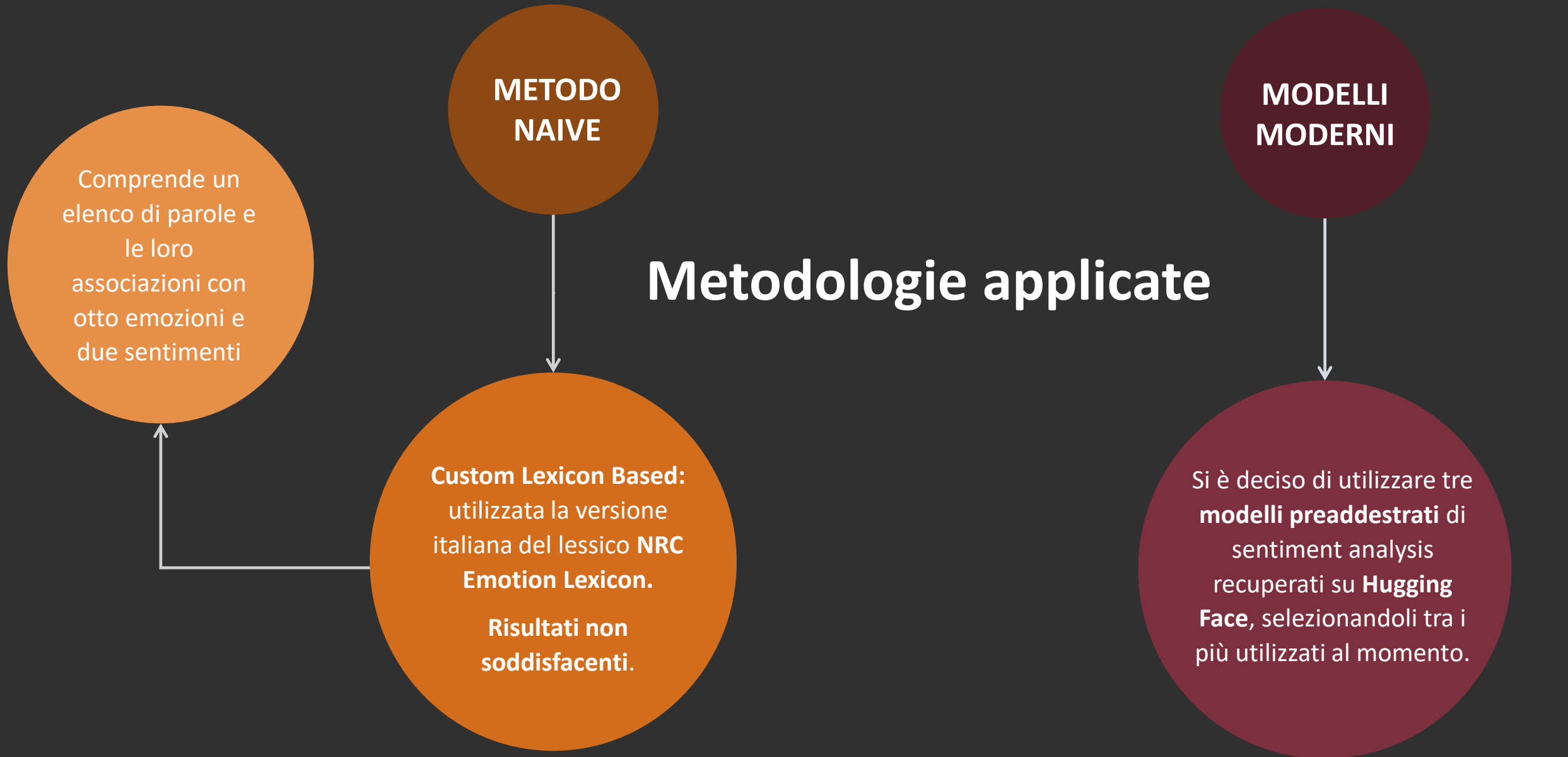
comment\_body

Voterei un partito più a sinistra in Italia se ci fosse un...  
Sì io ti parlo di quello che vedo di come mi sembra...  
Verissimo. Da sempre più anni i termini sinistra / destra...  
Vengono paragonati ai gamberi per il tema degli allergeni...  
L'Unione europea ha emesso una normativa che consentE...

## DOPO

Parola 1	Parola 2	Parola 3	Parola 4	Parola 5	...
voterei	partire	più	sinistra	italia	...
si	parlo	vedere	sembrare	funzionino	...
verissimo	da	sempre	più	anno	...
vengono	paragonare	gambero	temo	allergeno	...
il	unione	europea	emettere	normativa	...

# QUAL È IL CARATTERE DELLE DISCUSSIONI?





# MODELLI UTILIZZATI

Dotato di un corpus di riferimento di **Twitter italiani** annotati con quattro emozioni fondamentali: **rabbia, paura, gioia, tristezza**. Comprimidoli, possiamo anche fare l'analisi del sentiment.

## Italian BERT Sentiment model

Esegue l'analisi del sentiment sulle frasi **italiane**. È stato addestrato a partire da un'istanza di **bert-base-italian-cased**, e messo a punto su un dataset italiano di **tweets**, raggiungendo su quest'ultimo l'82% di precisione.

## FEEL-IT: Emotion and Sentiment Classification for the Italian Language

Modello **bert-base-multilingual-uncased** ottimizzato per l'analisi del sentiment sulle **recensioni dei prodotti in sei lingue**, tra cui l'italiano. Prevede il sentiment della recensione sotto forma di **numero di stelle (tra 1 e 5)**.

## MULTILINGUAL: bert-base- multilingual- uncased- sentiment

# ETICHETTATURA

Si procede etichettando il dataset risultante per maggioranza: si comparano i risultati dei tre modelli più utilizzati e, il sentimento che prevale, sarà quello associato al dato commento. Quando ho tre sentimenti differenti, imposto "neutral".

comment_body	post_title	debate_group	NRC	BERT	FEEL_IT	MULTILINGUAL	MAJORITY_VOTE
amo i miei gatti	eccovi il mio gatto	1	[positive', 'joy']	neutral	positive	positive	<b>positive</b>
il cliente da lei chiamato non è al momento raggiungibile...	Possibile guerra NATO - RUSSIA. Se davvero dovessero succedere e vi chiamassero al fronte per combattere ...	2	['positive', 'trust']	neutral	negative	positive	<b>neutral</b>
Fino a qualche anno fa i ragazzi per non emigrare venivano a Milano. Ora che Milano è una cloaca, un misto tra un suq e Gotham city, andarsene via è l'unica scelta possibile. Non c'è volontà politica di sistemare la questione. Chi ci governa ha già nuova mano d'opera a basso costo da impiegare nelle loro realtà.	Boom di giovani in fuga dall'Italia, sono oltre un milione	6	['positive', 'fear']	negative	negative	negative	<b>negative</b>

# QUAL È IL SENTIMENT DEI DIBATTITI?

Il sentiment delle discussioni è calcolato sommando il sentiment dei commenti, pesati in base alla loro rilevanza e consenso espresso dai lettori.

# QUAL È IL SENTIMENT DEL SINGOLO POST?

Il sentiment dei post è determinato combinando il sentiment dei dibattiti con altre metriche (numero di nodi e archi, grado medio, coefficiente di clustering) per valutare l'importanza delle discussioni.

# INTERPRETAZIONE DEI RISULTATI

The background of the slide is a dark blue gradient. It features a bokeh effect with out-of-focus light circles in shades of teal, yellow, and white. On the right side, there is a vertical strip showing a close-up of a circuit board with various electronic components and solder points.



# COMMENTI SUI RISULTATI

Che carattere hanno queste discussioni?

La maggior parte dei dibattiti ha sentiment negativo, ad esclusione del post dei gatti, inserito per verificare la possibilità di score positivi.

post_title	post_sentiment
Farina di grillo, e voi la mangereste?	-184.24194841517496
4000 bambini morti in Palestina. Possibile che nessuno faccia niente a livello politico?	-83.36907841837238
Ho visto troppi post sulla Giulia Cecchettin	-192.99597794495756
Altre nubi sui Ferragnez settimana dopo settimana	-52.085561497326196
Eccovi il mio gatto	35.666666666666664
Possibile guerra NATO - RUSSIA. Se davvero dovessero succedere e vi chiamassero al fronte per combattere cosa fareste?	-188.61665824915823
Il problema del caldo anomalo(causato dal riscaldamento globale) che a nessuno sembra fregare qualcosa...	-171.99146645021645
Chi voterete alle europee?	-115.23436471362739
Boom di giovani in fuga dall'Italia, sono oltre un milione	-112.51599326599329

Questo risultato è in linea con le aspettative, poiché il post sui gatti è stato intenzionalmente inserito per verificare la possibilità di ottenere score positivi.

**MA...**  
**ABBIAMO RISPOSTO**  
**ALLE ALTRE DOMANDE**  
**INIZIALI?**

# Quanto tendono a discutere gli italiani sui social?

TITOLO DI POST	POPOLOSITA' DELLE DISCUSSIONI (numero di utenti che partecipano)
Farina di grillo, e voi la mangereste?	99
4000 bambini morti in Palestina. Possibile che nessuno faccia niente a livello politico?	43
Ho visto troppi post sulla Giulia Cecchettin	97
Altre nubi sui Ferragnez settimana dopo settimana	36
Eccovi il mio gatto	33
Possibile guerra NATO - RUSSIA. Se davvero dovessero succedere e vi chiamassero al fronte per combattere cosa fareste?	98
Il problema del caldo anomalo(causato dal riscaldamento globale) che a nessuno sembra fregare qualcosa...	97
Chi voterete alle europee?	68
Boom di giovani in fuga dall'Italia, sono oltre un milione	83

# Quali sono gli argomenti maggiormente dibattuti?

TITOLO DI POST	NUMERO DI DIBATTITI
Farina di grillo, e voi la mangereste?	6
4000 bambini morti in Palestina. Possibile che nessuno faccia niente a livello politico?	2
Ho visto troppi post sulla Giulia Cecchettin	3
Altre nubi sui Ferragnez settimana dopo settimana	4
Eccovi il mio gatto	7
Possibile guerra NATO - RUSSIA. Se davvero dovessero succedere e vi chiamassero al fronte per combattere cosa fareste?	6
Il problema del caldo anomalo(causato dal riscaldamento globale) che a nessuno sembra fregare qualcosa...	8
Chi voterete alle europee?	3
Boom di giovani in fuga dall'Italia, sono oltre un milione	6



# SVILUPPI FUTURI

1

**Espansione del numero di post:** analizzare più post da diversi subreddit italiani per ottenere un quadro più completo del sentiment.

2

**Estrazione dei topic:** implementare l'analisi di estrazione dei topic per identificare i temi principali che stimolano le discussioni.

3

**Nuovi modelli e metriche:** esplorare nuovi modelli di sentiment analysis e introdurre nuove metriche per un'analisi più dettagliata.



GRAZIE DELL'ATTENZIONE