



PROGETTO PER IL CORSO DI SOCIAL MEDIA ANALYTICS

ITALIANI ALLE URNE: COME SI SCHIERA LA COMUNITÀ ITALIANA DI TWITTER RIGUARDO LE #ELEZIONIPOLITICHE22

ANDREA MAVER 828725, ELEONORA PALOMBA 876479, OSCAR ZANOTTI 873763





002

RESEARCH QUESTIONS

ROAD TO ELECTION DAY

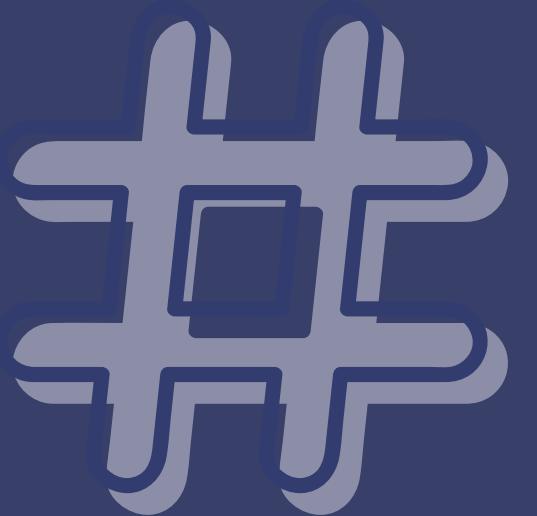


- La divisione politica della società si rispecchia anche nell'utenza di Twitter?
- È possibile riconoscere delle comunità nette in base alle preferenze politiche?
- Quali sono i sentimenti dei messaggi che vengono rivolti ai principali attori?
- C'è differenza tra le emozioni che suscitano i messaggi dei leader politici e quelli di chi invece ne parla?

STEP 1 - RACCOLTA DATI

@Tweepy

- Download dei tweet nei giorni 01/09 e 03/09
- 35.000 tweet raccolti in lingua italiana
- Un dataset per i tweet dell'utenza
- Un dataset per i tweet dei leader politici



#elezioniPolitiche22

#elezioniPolitiche2022

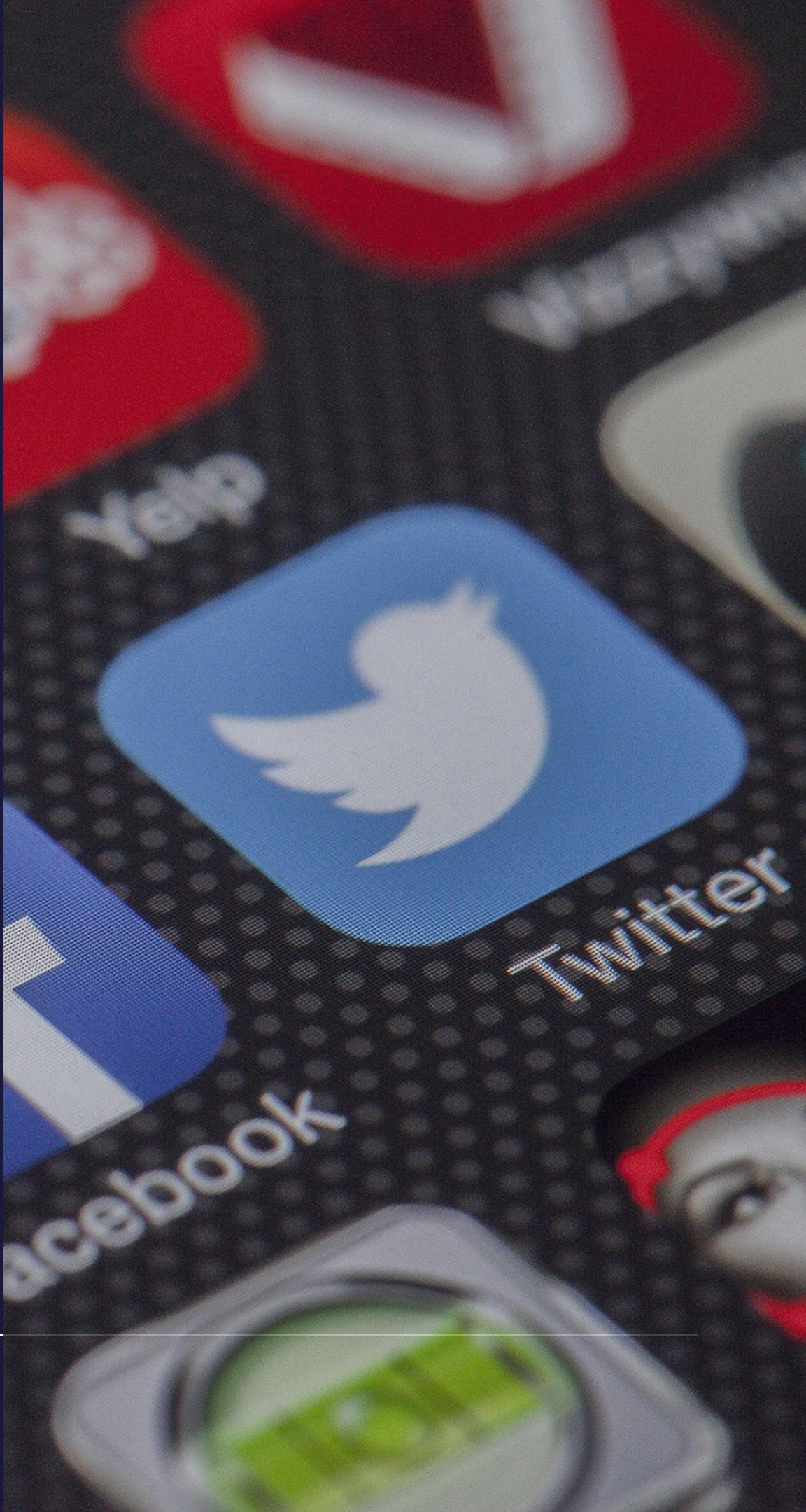
#elezioni2022

#elezioni22

#elezionipolitiche

≡ PREPROCESSING - I

1. Aggiunta delle colonne per hashtag, menzioni e autori
2. Rimozione delle emoji
3. Rimozione delle menzioni e riferimenti ad altri utenti
4. Rimozione degli hashtag
5. Rimozione degli url
6. Rimozione di eventuali spazi extra
7. Trasformazione del testo in lowercase
8. Rimozione di simboli contenuti nel testo
9. Normalizzazione dei caratteri accentati



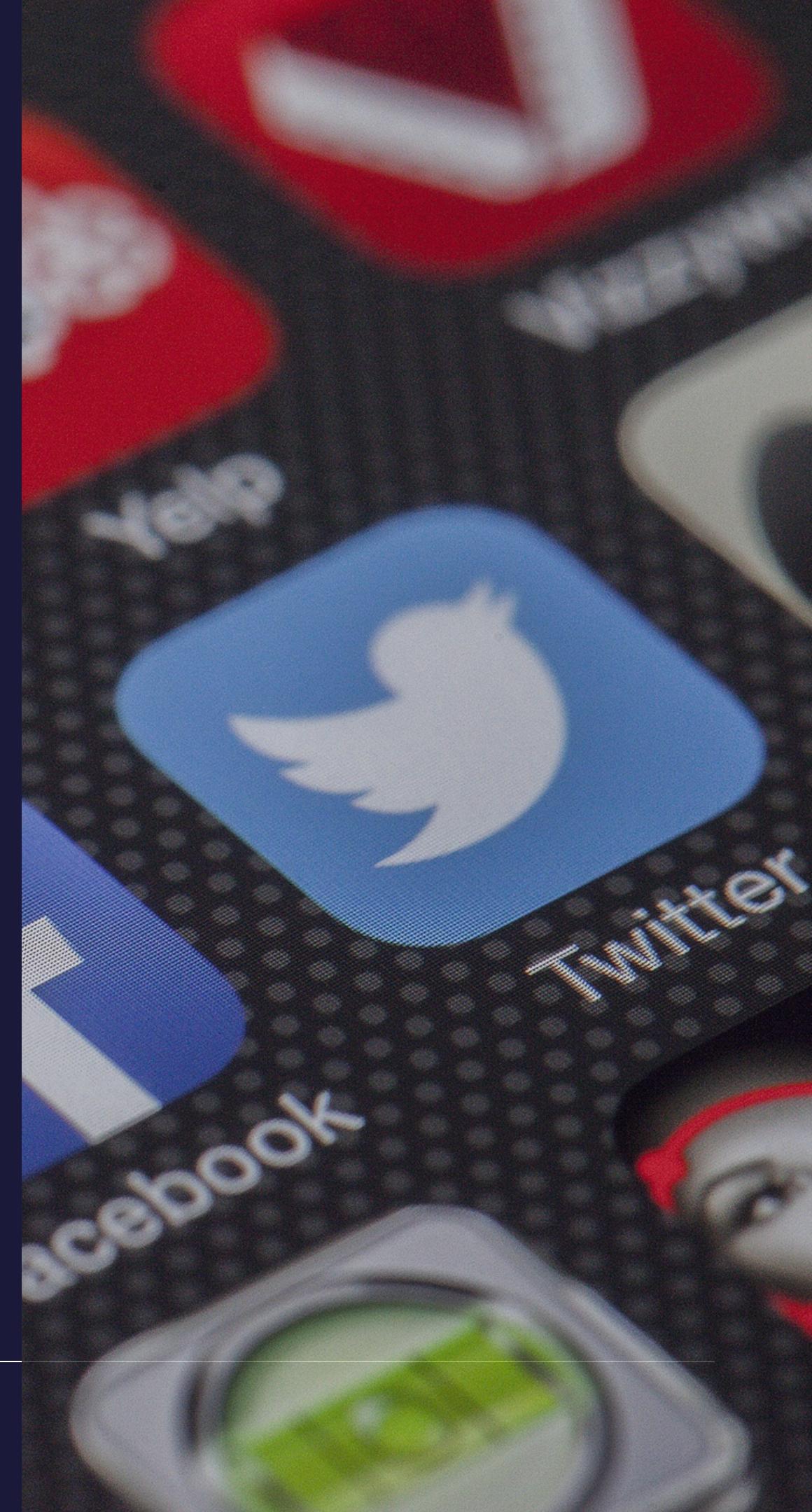
≡ PREPROCESSING - II

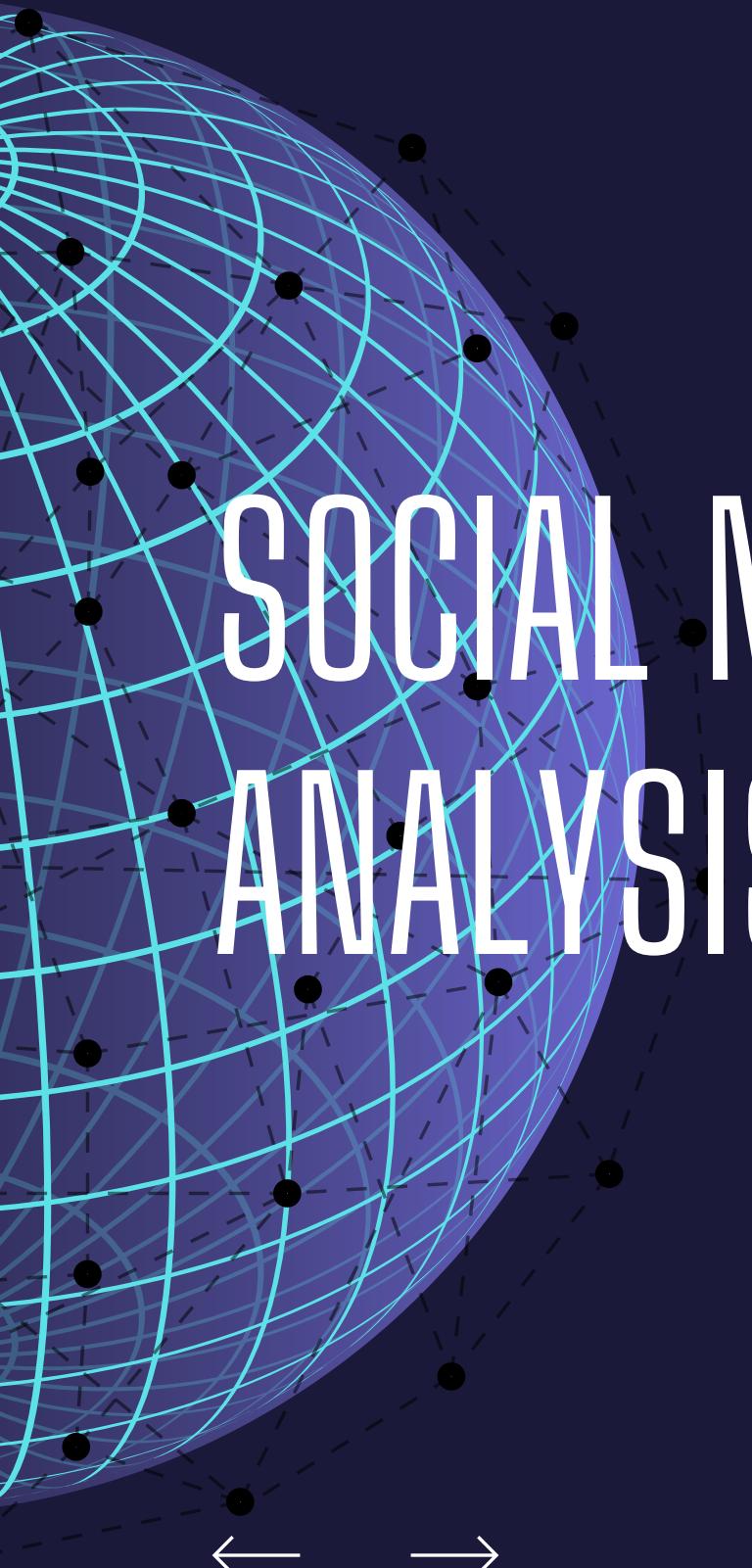
Per la sentiment analysis si utilizzano solo i tweet che contengono nel testo il nome dei principali leader



1. FRATELLI D'ITALIA: GIORGIAMELONI, MELONI
2. PARTITO DEMOCRATICO: ENRICOLETTA, LETTA
3. LEGA: MATTEOSALVINIMI, SALVINI
4. MOVIMENTO 5 STELLE: GIUSEPPECONTEIT, CONTE
5. AZIONE - ITALIA VIVA: CARLOCALENDÀ, CALENDA

- Creazione di 5 dataset, uniti per creare il dataset finale, composto inizialmente da 19174 righe (ridotte poi a 14829 perché rimossi tweet postati da account falsi o 'bot') e 19 colonne.
- Tokenizzazione e lemmatizzazione del testo con la libreria Tree Tagger
- Rimozione delle stopwords





SOCIAL NETWORK ANALYSIS



006

Viene creato il grafo di interazione tra utenti utilizzando i retweet: se un utente retweetta un post, viene creato un arco tra i due nodi.

Viene applicato un peso all'arco in base al numero di volte in cui un utente ha retweettato l'altro.

**IL GRAFO RISULTANTE È PESATO,
DIREZIONATO E COMPOSTO DA 12947 NODI
E 27097 ARCHI**



METRICHE E ANALISI ESPLORATIVA

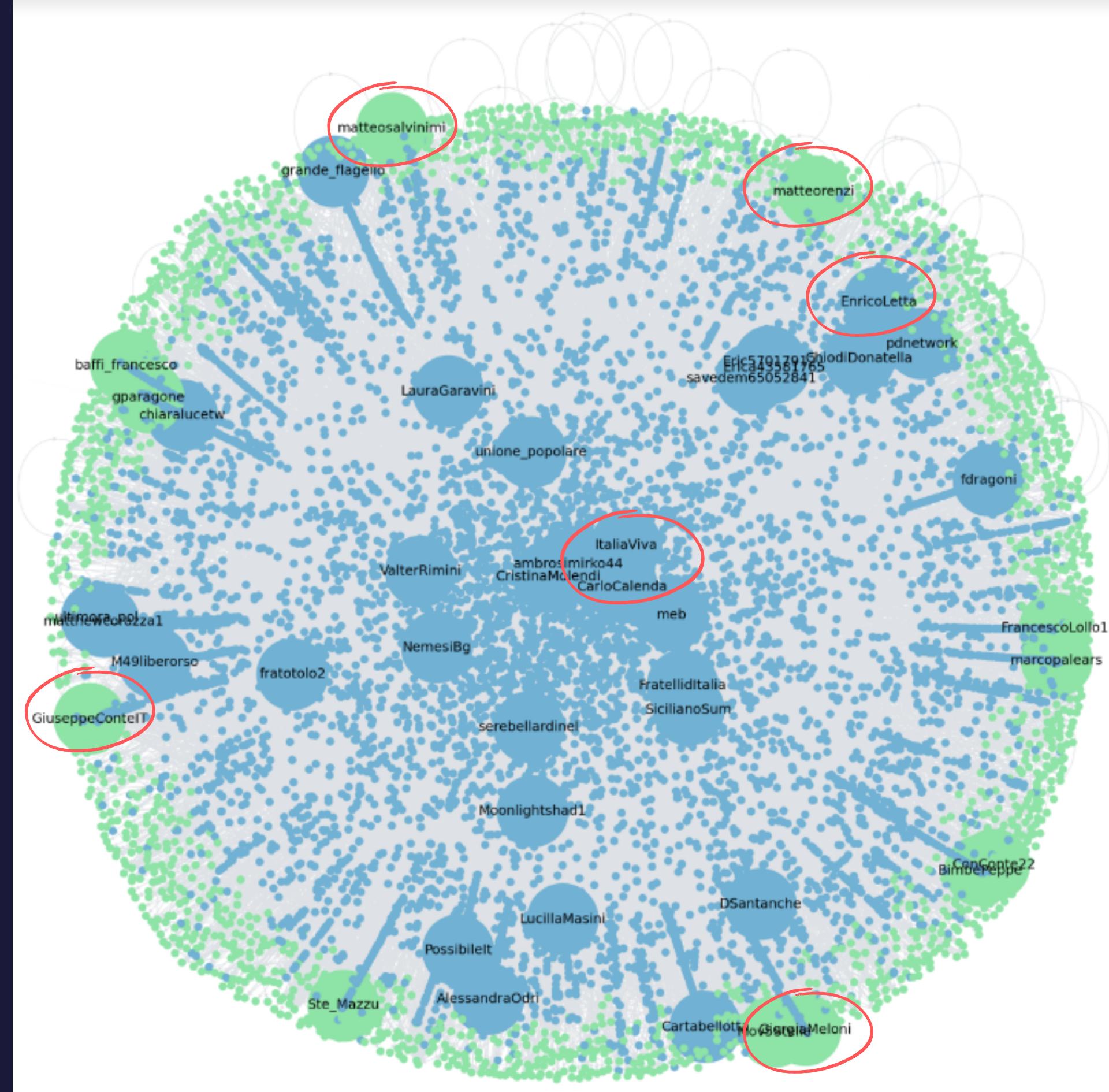
metrica	grado medio	densità	connettività nodi	connettività archi	assortatività	reciprocità generale	componenti fortemente connesse
valore	2.8	0.000162	0	0	-0.0616	0.01705	12678

- Network è tendenzialmente neutro e leggermente disassortativo
- Grafo molto sparso
- Gli utenti non interagiscono particolarmente tra loro
- Grafo sia disconnesso
- Ogni utente ha effettuato in media quasi tre retweet nell'arco dei giorni presi in considerazione



≡ VISUALIZING THE NETWORK

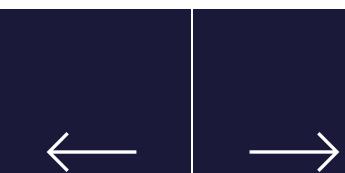
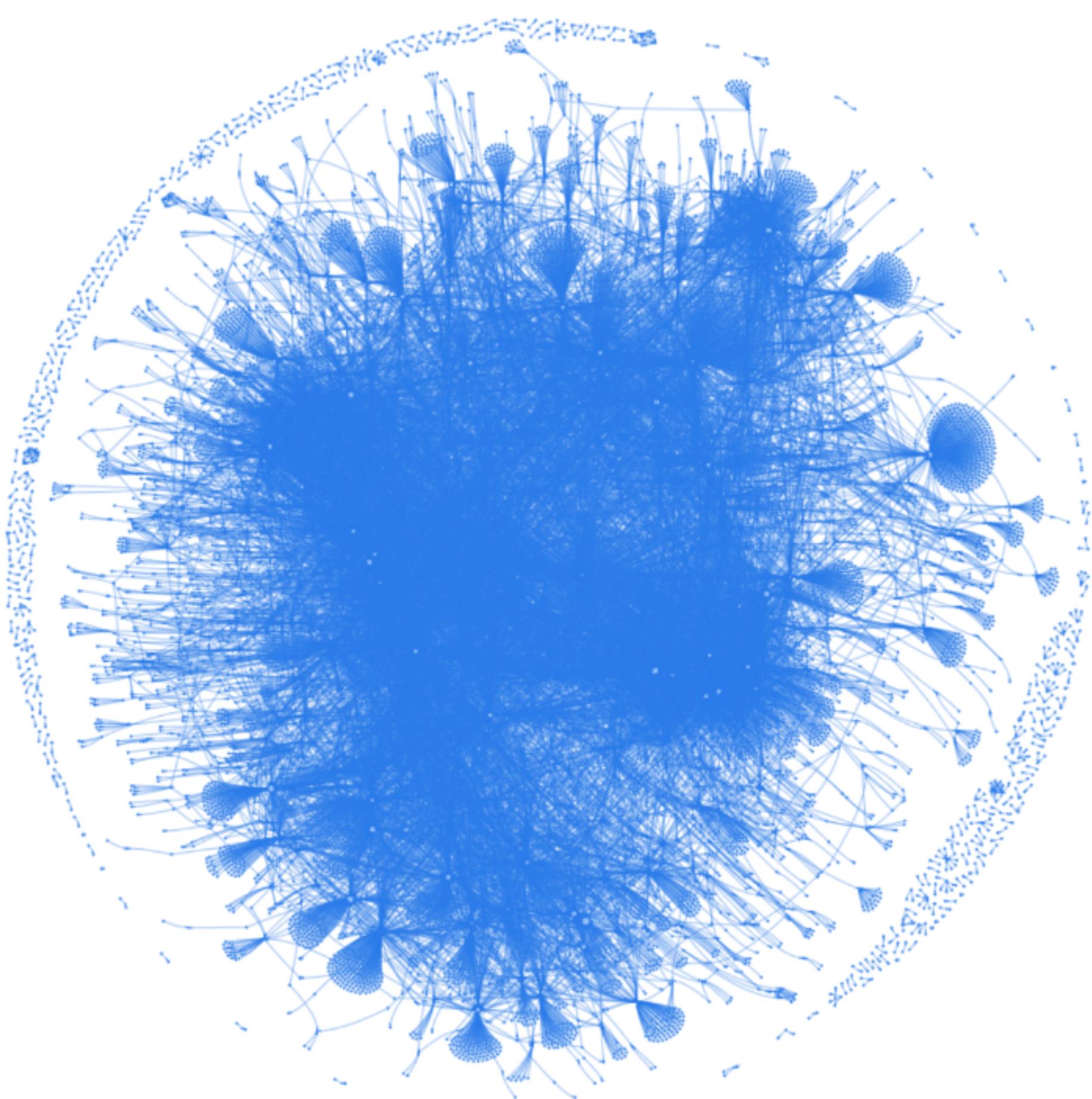
@NetworkX



008

≡ VISUALIZING THE NETWORK

@Pyvis



Si conferma che il grafo
è disconnesso.

Le strutture a
mongolfiera sembrano
delle indicazioni di
comunità create intorno
agli hub.

009

author_name	author_follower
SkyTG24	3'618'782
TwitterGov	2'371'841
fattoquotidiano	2'179'848
sole24ore	1'741'955
marcotravaglio	1'652'386
MediasetTgcom24	1'407'494
TgLa7	770'899
EnricoLetta	716'015
ilpost	713'621
meb	654'442
Adnkronos	593'177
La7tv	589'914
ilgiornale	567'333
Roma	506'153
civati	408'480

Tabella 2: Top 15 account più seguiti

author_name	author_follower
EnricoLetta	716'015
meb	654'442
pdnetwork	400'377
CarloCalenda	396597
DSantanche	191'623
FratellidItalia	182'065
gparagone	172'004
Cartabellotta	76'223
PossibileIt	58'214
fdragoni	54'974
ItaliaViva	50'143
grande_flagello	25'756
M49liberorso	25'744
Moonlightshad1	21'249
ChiodiDonatella	20'074

Tabella 3: Top 15 hub più seguiti

Nella tabella a sinistra si possono osservare gli account con più follower, tra quelli che hanno postato almeno un tweet.

Nella tabella a destra si considerano solo i nodi etichettati, in particolare i quindici con più follower.

Osservando la closeness risultano presenti i leader politici dei due principali schieramenti, con il loro rispettivo partito.

Considerando la betweenness si notano valori molto bassi. Si osserva un pattern interessante: molti degli account sono evidenti sostenitori del Movimento 5 Stelle, partito centrista che sembra posizionarsi tra le due fazioni di destra e sinistra.

closeness centrality	betweenness centrality
ultimora_pol	0.127
GiorgiaMeloni	0.0838
pdnetwork	0.0747
peppeprovenzano	0.0600
EnricoLetta	0.0586
FratellidItalia	0.0575
ilgiornale	0.0511
direpuntoit	0.0511
robertomenia	0.0504
DSantanche	0.0493
ambrosimirko44	0.00316
Yoda15271485	0.00176
Rapace781	0.00169
Finiguerra	0.00159
romano_tano	0.00156
BuselliGiulia	0.00154
UPrieste	0.00152
demagistris	0.00119
Marzo531	0.00110
radionowher	0.00109

Tabella 4: Top 10 closeness e betweenness centrality

COMMUNITY DETECTION

Viene utilizzata la massimizzazione della modularità per creare cluster distanti da raggruppamenti casuali

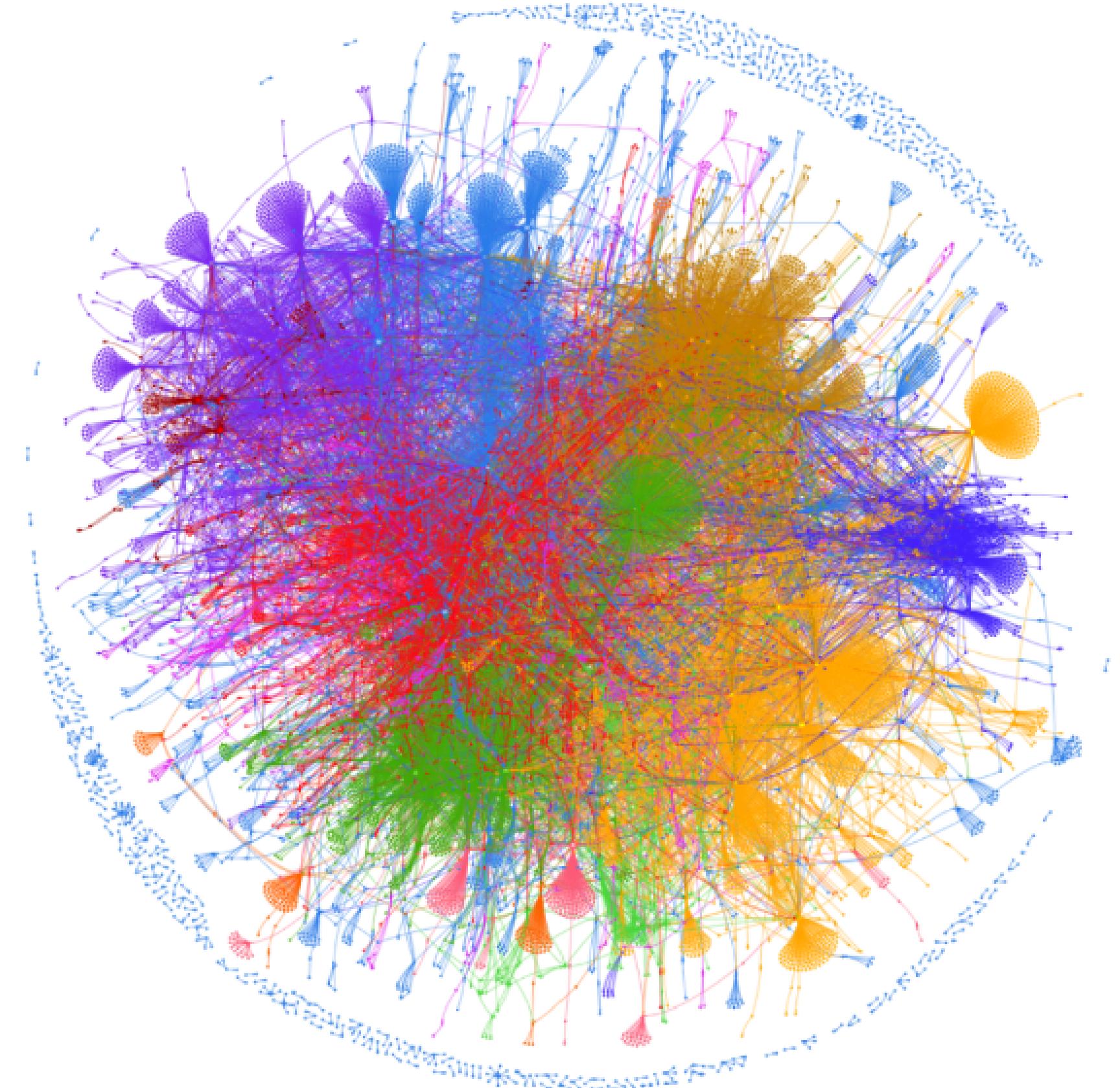


Figura 6: Visualizzazione delle comunità trovate

Vengono identificate 351 comunità con un valore di modularità medio di 0.712.

Si nota subito che in ognuna di esse sia presente uno degli hub trovati precedentemente, come Giorgia Meloni e Enrico Letta.

COMMUNITY EVALUATION

Viene effettuata una valutazione semantica delle comunità, non avendo ground truth, per cercare pattern di termini utilizzati all'interno delle diverse comunità.



Figura 7: Wordcloud delle 10 comunità più numerose

≡ SENTIMENT ANALYSIS

È stata eseguita utilizzando un approccio di tipo lexicon-based, dopo aver ricondotto le parole contenute nei tweet ai loro lemmi originari, per poi associarvi un punteggio di sentiment tramite un dizionario (lexicon).

Tale metodologia utilizza i testi tokenizzati e lemmatizzati, e ha seguito alcuni passaggi:

1

Gestione del lessico italiano tramite **Sentix**:
Lessico che assegna polarità positiva o negativa a lemmi in lingua italiana.

2

Calcolo del sentiment score tramite **VADER**:
VADER permette di analizzare la polarità (positivo/neutrale/negativo) e l'intensità delle emozioni.

3

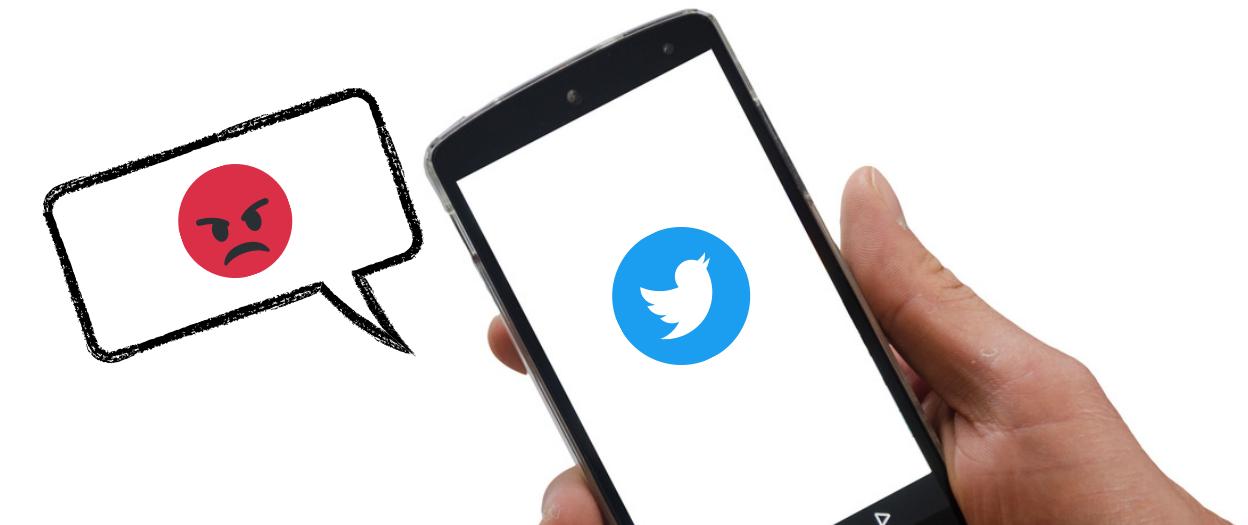
VADER viene aggiornato con il dizionario ottenuto da **Sentix** e fornisce un compound score normalizzato, ovvero un valore di sentiment complessivo compreso tra -1 ed 1.

SENTIMENT ANALYSIS



Figura 8: Sentimenti dell'utenza

Figura 9: Sentimenti dei leader



EMOTION ANALYSIS

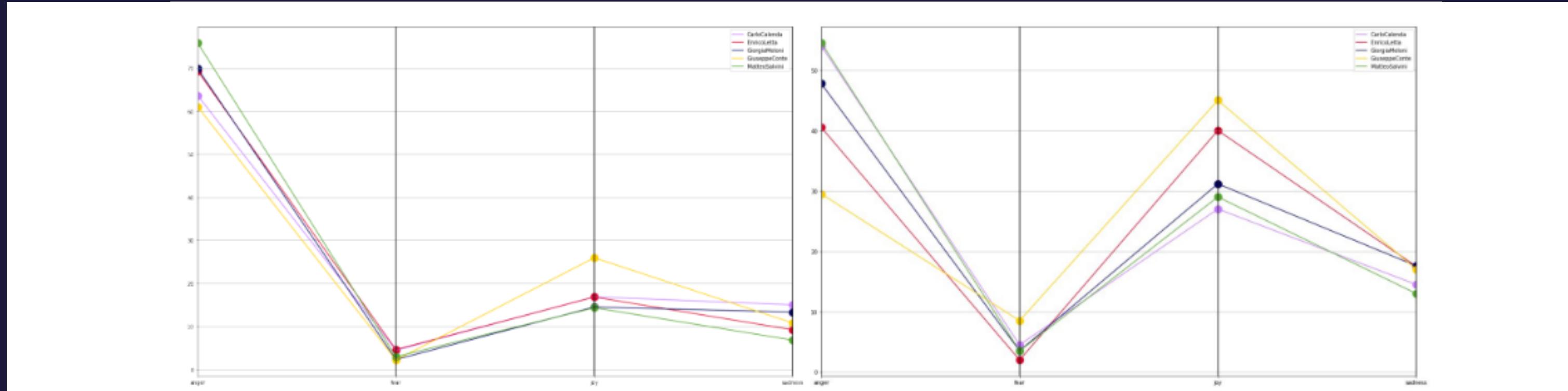
Viene utilizzata la libreria FEEL-IT, modello nato appositamente per assegnare a tweet in lingua italiana quattro emozioni elementari, cioè rabbia, gioia, paura e tristezza.

Il modello è stato creato facendo fine-tuning di UmBERTO, una versione italiana di BERT, famoso modello per svolgere task di NLP.

Come per la sentiment, l'analisi confronta i tweet dei leader di partito e quelli dell'utenza che ne parla per poter cercare se esistano differenze.



EMOTION ANALYSIS



Tra i tweet dell'utenza spicca fortemente la rabbia, che stacca di molto le altre emozioni a dimostrare come i commenti sulla piattaforma siano per la maggior parte negativi.

Nei tweet dei leader si nota una fortissima crescita per quanto riguarda la gioia: questo è un risultato interessante che mostra come ci sia una forte differenza tra quello che l'utenza percepisce e comunica e quello che i vari segretari tentano di veicolare.



Rabbia

"#Elezioni2022 Giorgia #Meloni:
"Non ci interessano le
provocazioni, io non vi racconto
balle sull'attuale situazione
italiana... Non siamo messi
bene, no... Non è un momento
facile,
ma penso che con un po' di
coraggio e buona volontà, si può
invertire tendenza"

48 120 350

@ElisaVerdi

"L'unica persona che mi
rappresenta è #Salvini. Mi spiega
che molta gente preferisce
sudici come #Letta e #Conte
miserabili che ci porteranno alla
distruzione. Ieri il grullino ha
ribadito che le sanzioni alla
Russia sono necessarie. Finirà
molto male. #ElezioniPolitiche22
#credo"

1 37 52
4

Tristezza



Vengono visualizzate delle wordcloud per poter confrontare gli argomenti di cui parlano i diversi leader.



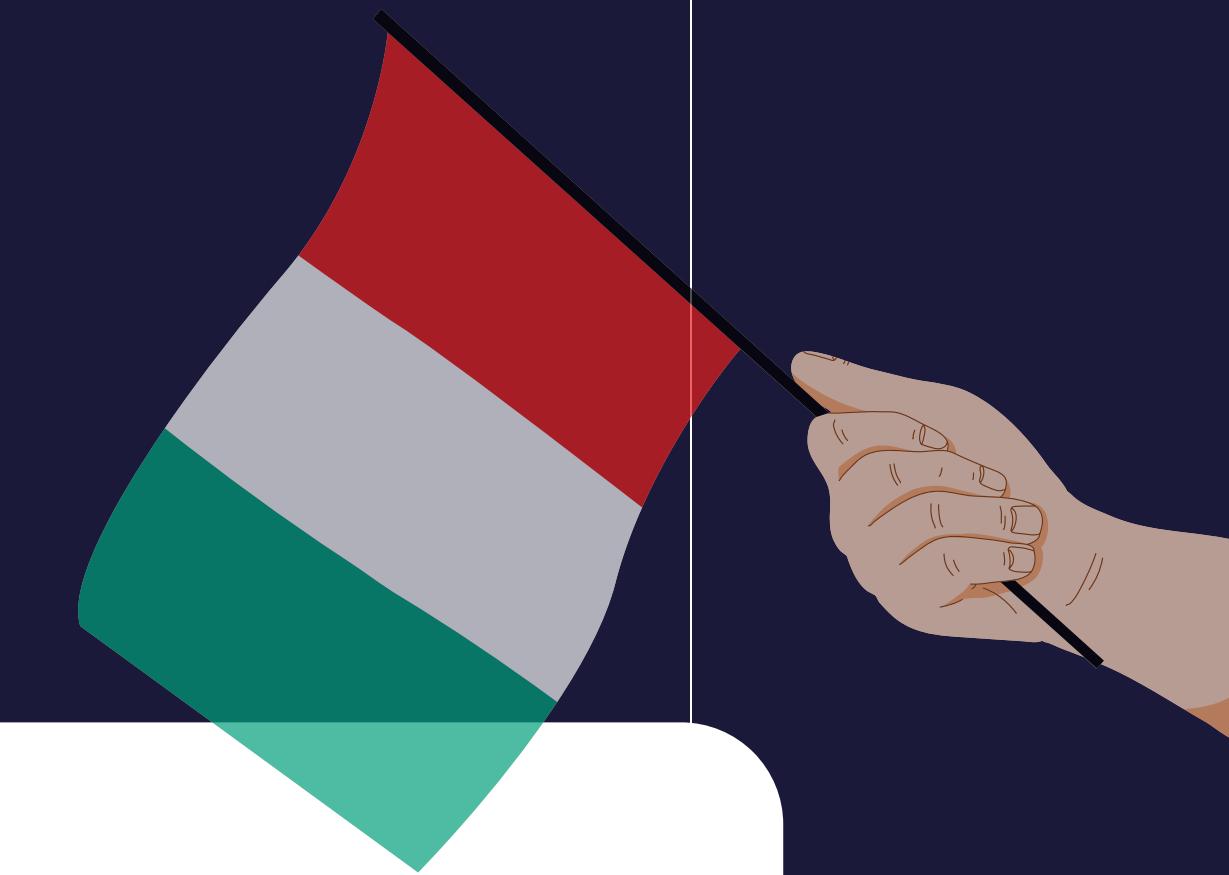
Figura 13: Wordcloud dei tweet dei leader



@matteosalvinimi

"I libri di scuola sono un salasso per le famiglie. Come aiutare gli italiani? Ecco la proposte della Lega. Fatemi sapere cosa ne pensate. Ti aspetto su TikTok per rimanere aggiornato"

598 1200 350



@GiorgiaMeloni

"La sinistra dei salotti ha dimostrato in questi anni di essere lontana dai temi del lavoro, della sicurezza e dello sviluppo. Gli italiani sanno bene che solo un governo capace di difendere gli interessi nazionali potrà combattere delocalizzazioni e concorrenza sleale. #VotaFdI"

48 120 350



@GiuseppeConteIT

"Chiediamo che le urgenze dei cittadini siano anche le urgenze del governo. Nè più, né meno. E su questo il Mov5Stelle e il Paese aspettano le risposte di Palazzo Chigi."

58 3740 64
0



GRAZIE PER
L'ATTENZIONE



ANDREA MAVER 828725, ELEONORA PALOMBA 876479, OSCAR ZANOTTI 873763