

Movimento QAnon Analisi del dibattito online

Raccolta dei tweets

La ricerca è filtrata in base a espressioni
e hashtag tipici della narrazione QAnon:

#QSees, #QKnows, etc.



Vengono estratti 49732 tweet non duplicati
in lingua inglese in data 5 giugno 2020

I testi sono stati pubblicati da **26770** utenti diversi e contengono menzioni/retweet/risposte/citazioni ad altri **2152**.



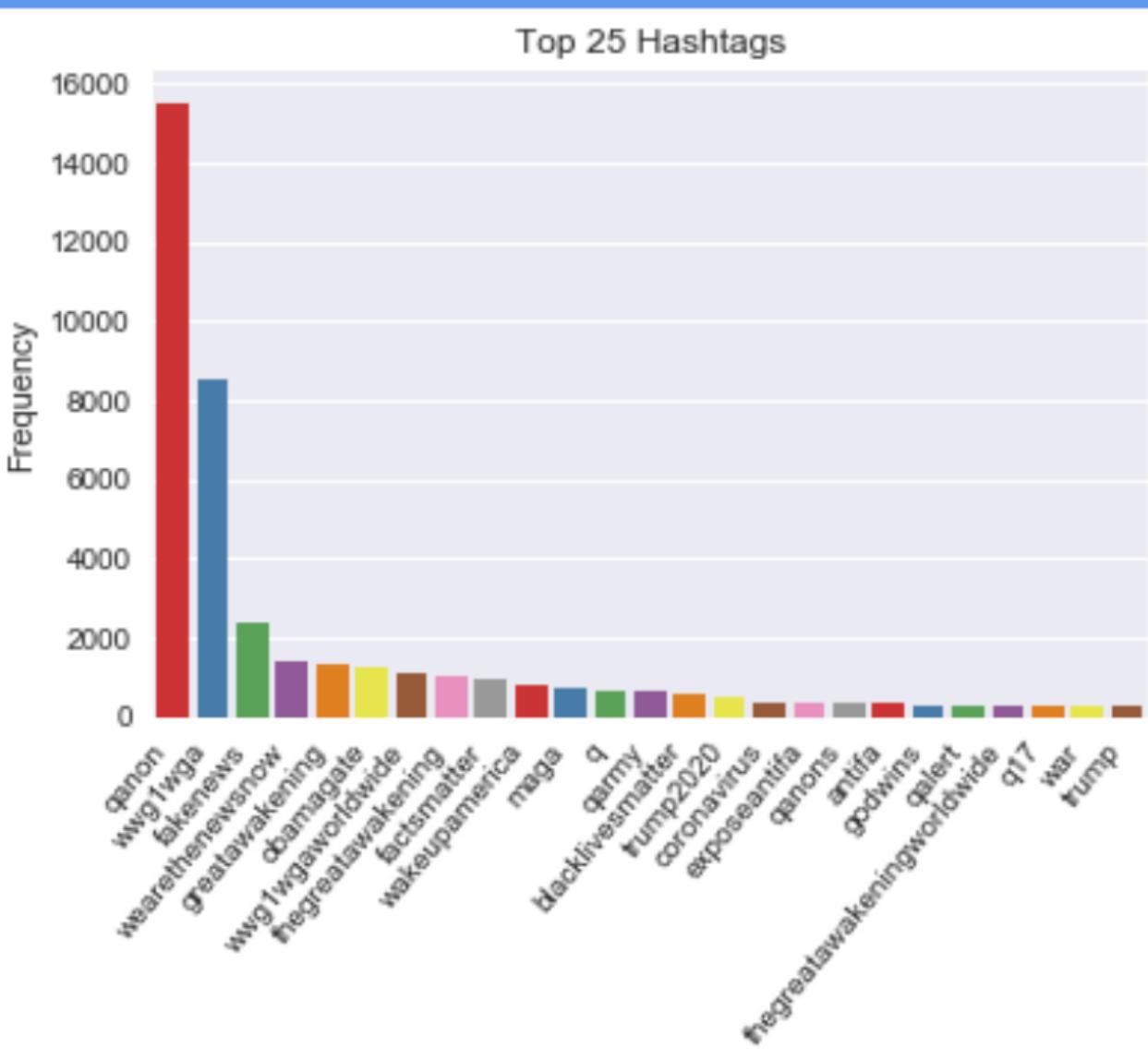
Informazioni interessanti

I sottocampi archiviati nel db finale sono
relativi a

- ✓ Utente: nome, follower e following
- ✓ Post: id, testo e hashtags
- ✓ Interazioni: mention, retweet, quote...

Analisi preliminari

Text Preprocessing



Hashtags

Espressioni ricorrenti

Riferimenti religiosi

Lessico militare

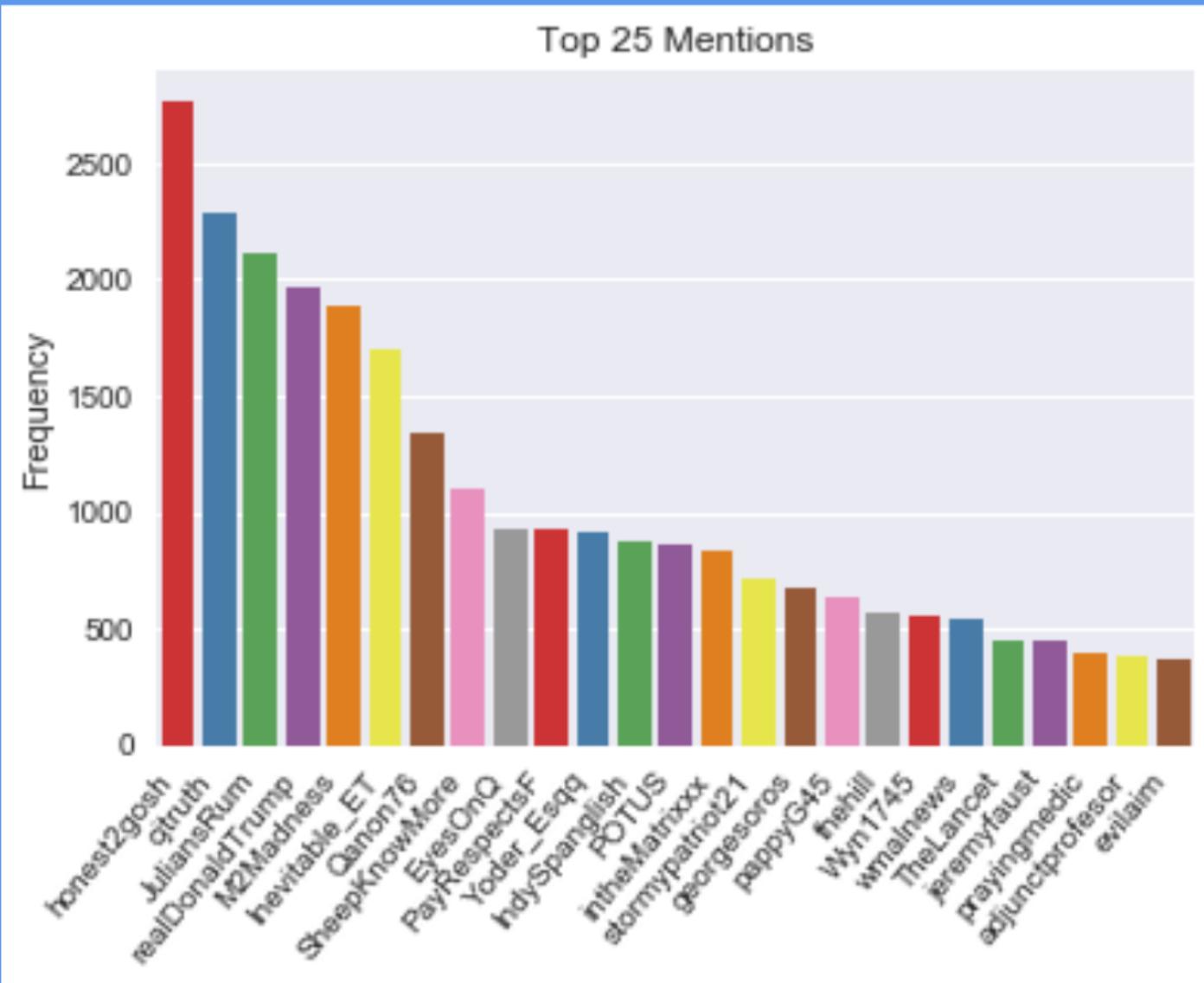
Sostegno a Trump

Coronavirus, BLM e AntiFa

Interazioni virtuali

- ✓ Retweet e menzioni > 76%
- ✓ Citazioni 15.4%
- ✓ Risposte 8.6%

Top 25 mentions



Text Preprocessing

-  **Rimozione parti inutili:** URLs, menzioni, punteggiatura, numeri, caratteri speciali
-  **Conversione hashtags in parole**
-  **Conversione emoji/ emoticons in parole**   "happy face smiley"
-  **Normalizzazione:** conversione contrazioni e espressioni colloquiali, testo in minuscolo
-  **POS Tagging + Lemmatizzazione**  **Rimozione stopwords**



L'individuazione dei bot avviene col modello
fornito dalla libreria `tweetbotornot`

Impostando **90%** come soglia , il modello
trova **387** profili automatizzati (840 tweet)

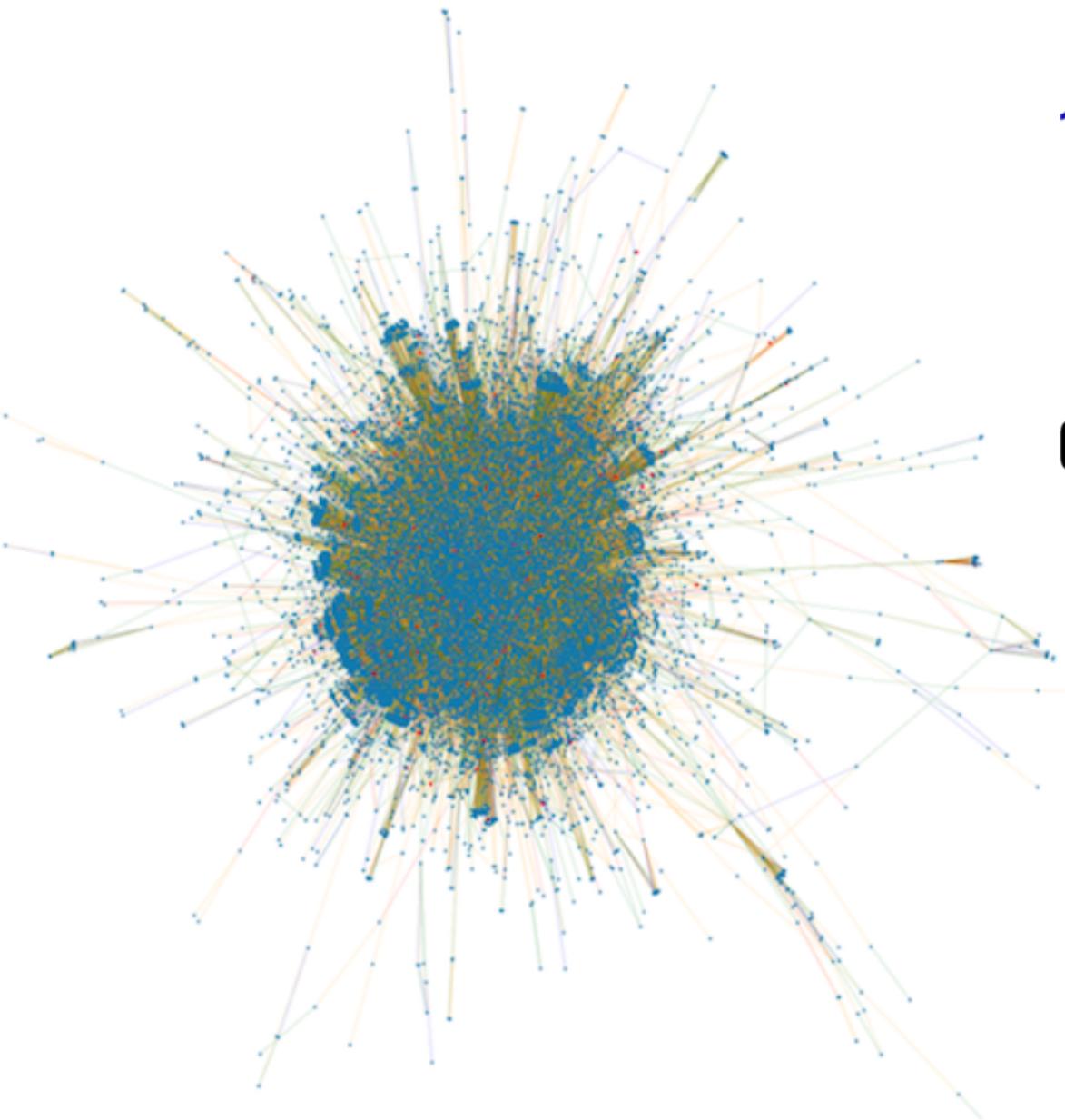
**La media del numero di follower per bot è molto elevata (4199),
facendo supporre una grande potenziale influenza sull'utenza reale.**



Si definisce un grafo dal complesso di interazioni, composto da 30914 nodi e 65001 archi.

Poichè molto sparso, si è estratta la componente connessa più estesa e filtrati i nodi con grado superiore al primo.





16061 nodi e 50627 archi

Grado compreso fra 2 e 1975 con una media di 6.3 e una moda di 2



Metriche di connessione

- ✓ Coeff di clustering medio ~0.284(moderato)
- ✓ Transitività 0.0047 (basso)
- ✓ Densità ~0.0004 (molto bassa)

Nonostante sottografo, la struttura rimane **molto sparsa**.



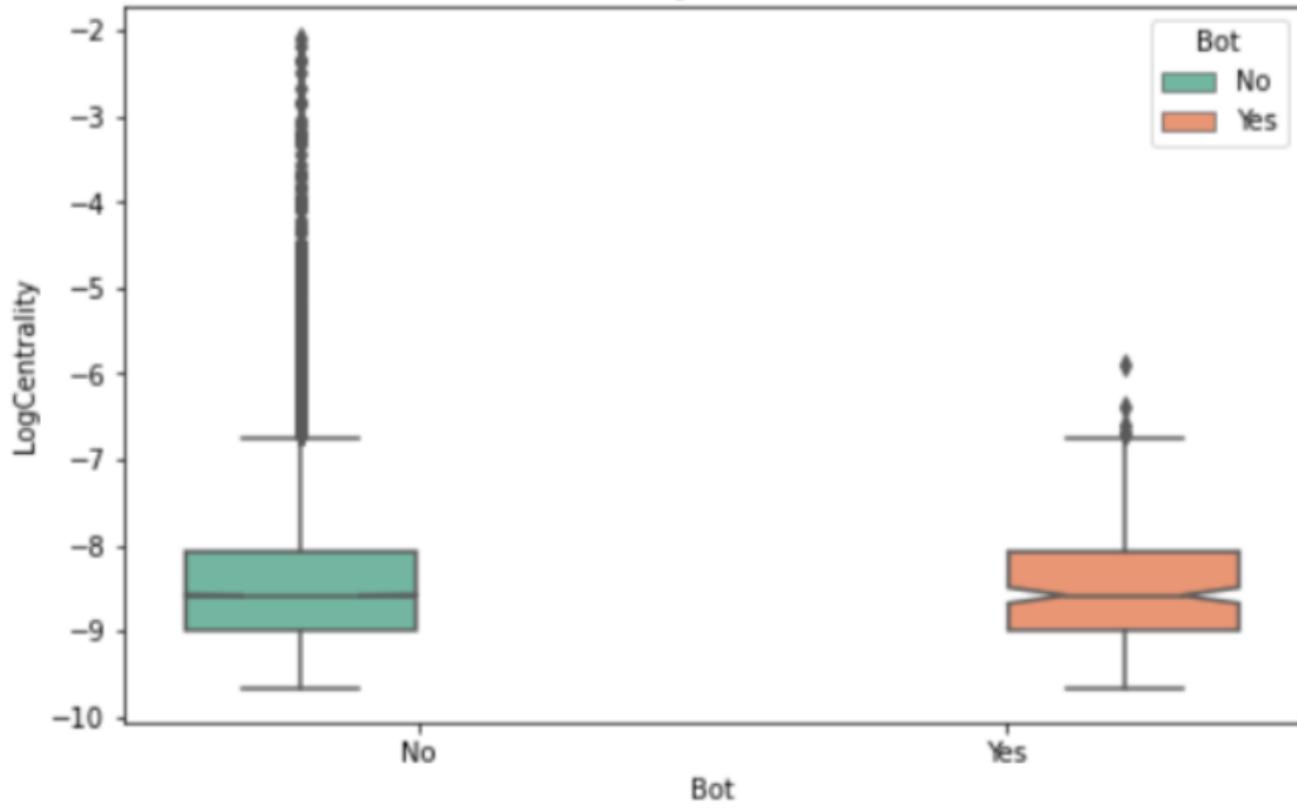
Metriche di distribuzione

La **degree-centrality** media dei bot (**0.0003**) è leggermente minore rispetto a quella dei vertici relativi agli utenti umani (**0.0004**)

Tuttavia, questa differenza non è significativa



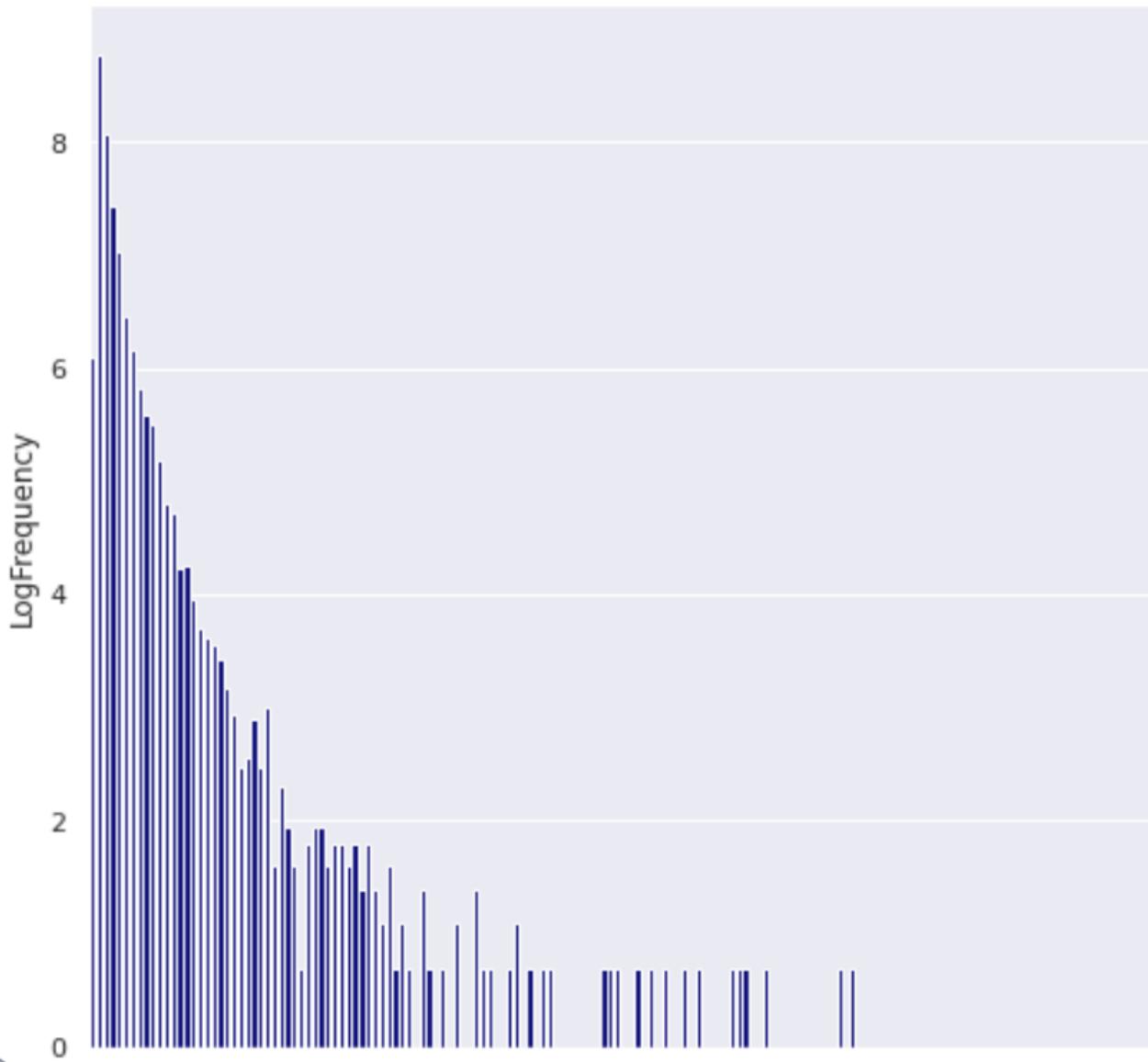
Centrality distribution



Metriche di distribuzione II



Degree distribution



Relazione esponenziale negativa
fra il num. di archi incidenti e la
frequenza di tale grado

Presenza di hub all'interno della
rete, che alimenta l'effetto
small world e preserva
l'invarianza di scala

Correlazione di grado

È possibile cogliere i pattern di connessione grazie al calcolo di un **coefficiente di assortatività** (correlazione di Pearson fra le freq. dei gradi)

Tale coefficiente è pari a **-0.23** e indica che la rete è **disassortativa**: gli hub si evitano sistematicamente

Il modello d'interazione nei social network è infatti definito "**the richer get richer**"



Strutture principali della rete

Sono stati individuati **38435 cricche massimali** e **645 ponti**, in entrambi i casi con una concentrazione verso la zona centrale

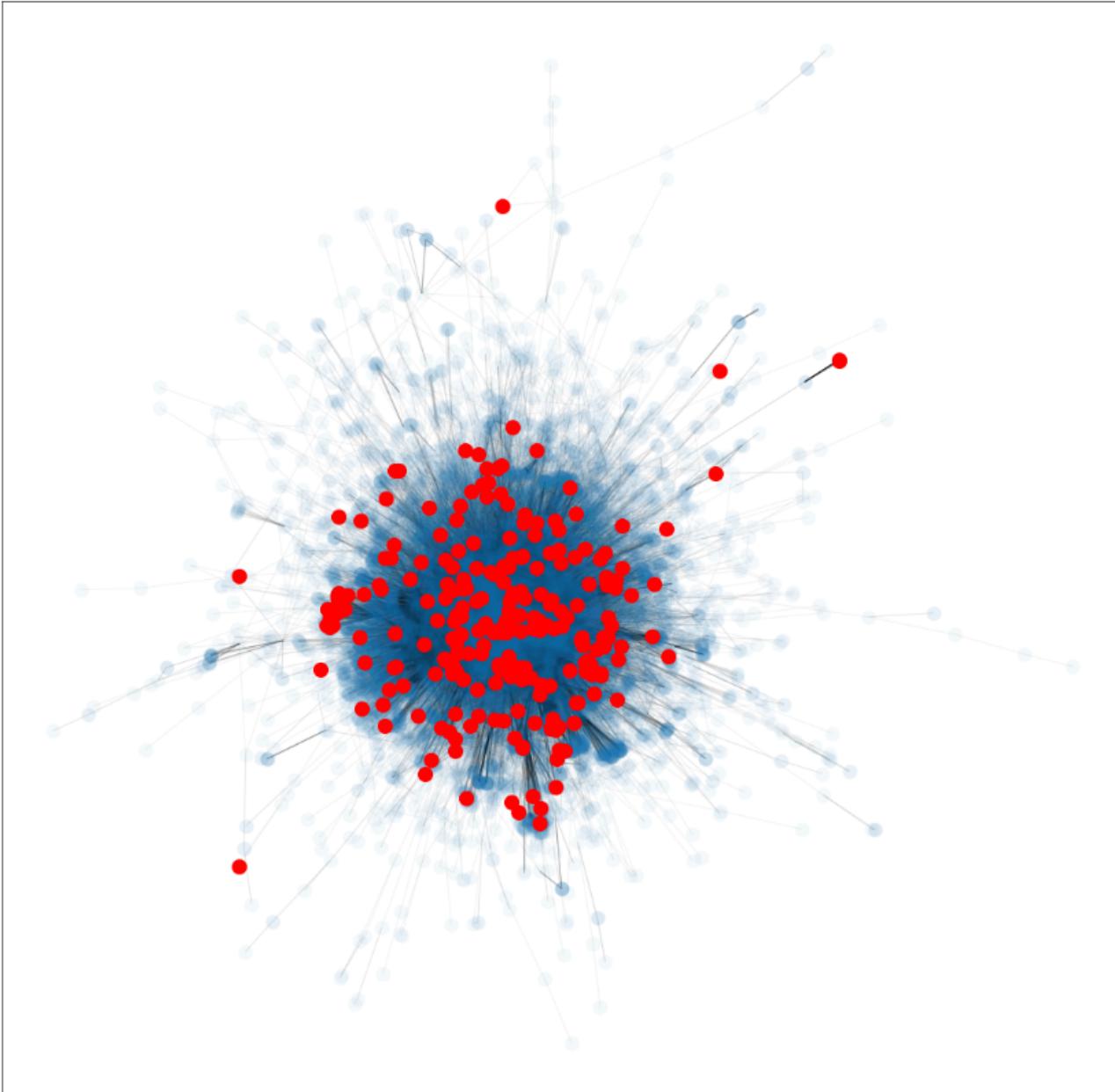


Strutture principali della rete: **bot**

Il sottografo dei bot è composto da **248 vertici** e **9 archi**, concentrati nel centro, a conferma della potenziale influenza

In questo subset, le connessioni sono rarissime e l'assenza di coesione viene confermata dal fatto che coefficiente di clustering e transitività abbiano valore **nullo**





Strutture principali della rete: set dominanti

**L'importanza dei bot viene però ridimensionata
verificandone la presenza all'interno dei set dominanti.**



Strutture principali della rete: **set dominanti**

L'importanza dei bot viene però ridimensionata verificandone la presenza all'interno dei set dominanti.

Sono esaminati due tipi di architetture reggenti:

✓ Node-dominating set

✓ Edge-dominating set



Node-dominating set

È il subset di vertici D tale per cui ogni nodo esterno a D è
adiacente ad almeno un suo membro

Mentre la percentuale di profili reali inclusi in D sul totale
degli stessi è del 20%, quella dei bot è inferiore al 10%



Edge-dominating set

È il subset di archi F tale per cui ogni legame in esso non compreso
è incidente all'estremo di almeno un arco di F

In questo caso, il set dominante non contiene alcun bot



**Sono state testate 4
metodologie diverse per
estrarre comunità dal grafo**

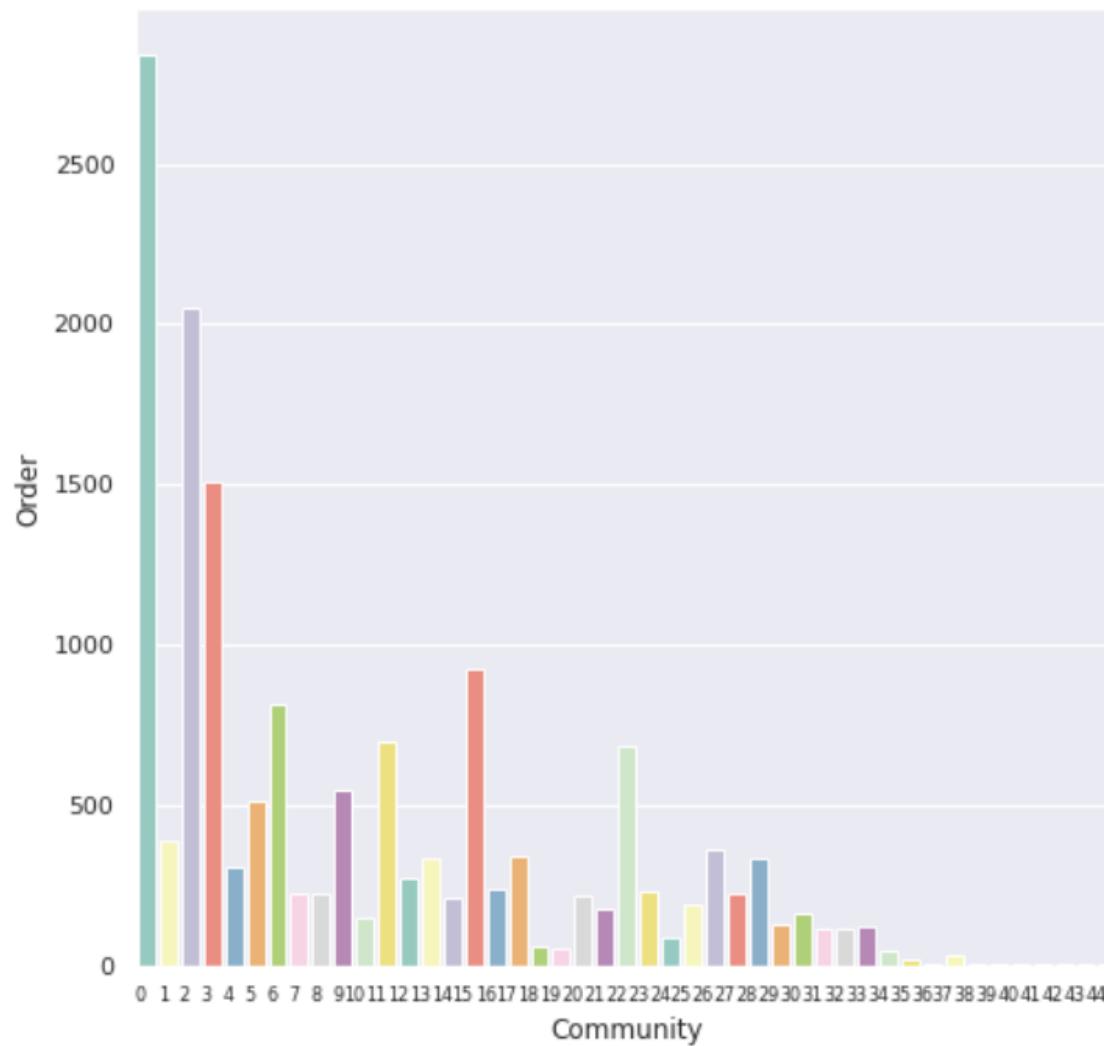


Algoritmo di Louvain

Algoritmo greedy per massimizzare la modularità, intesa come densità relativa dei legami all'interno delle comunità rispetto a quella degli archi all'esterno.

**Estraе 42 comunità molto sbilanciate, per un
valore finale di modularità pari a 0.58**



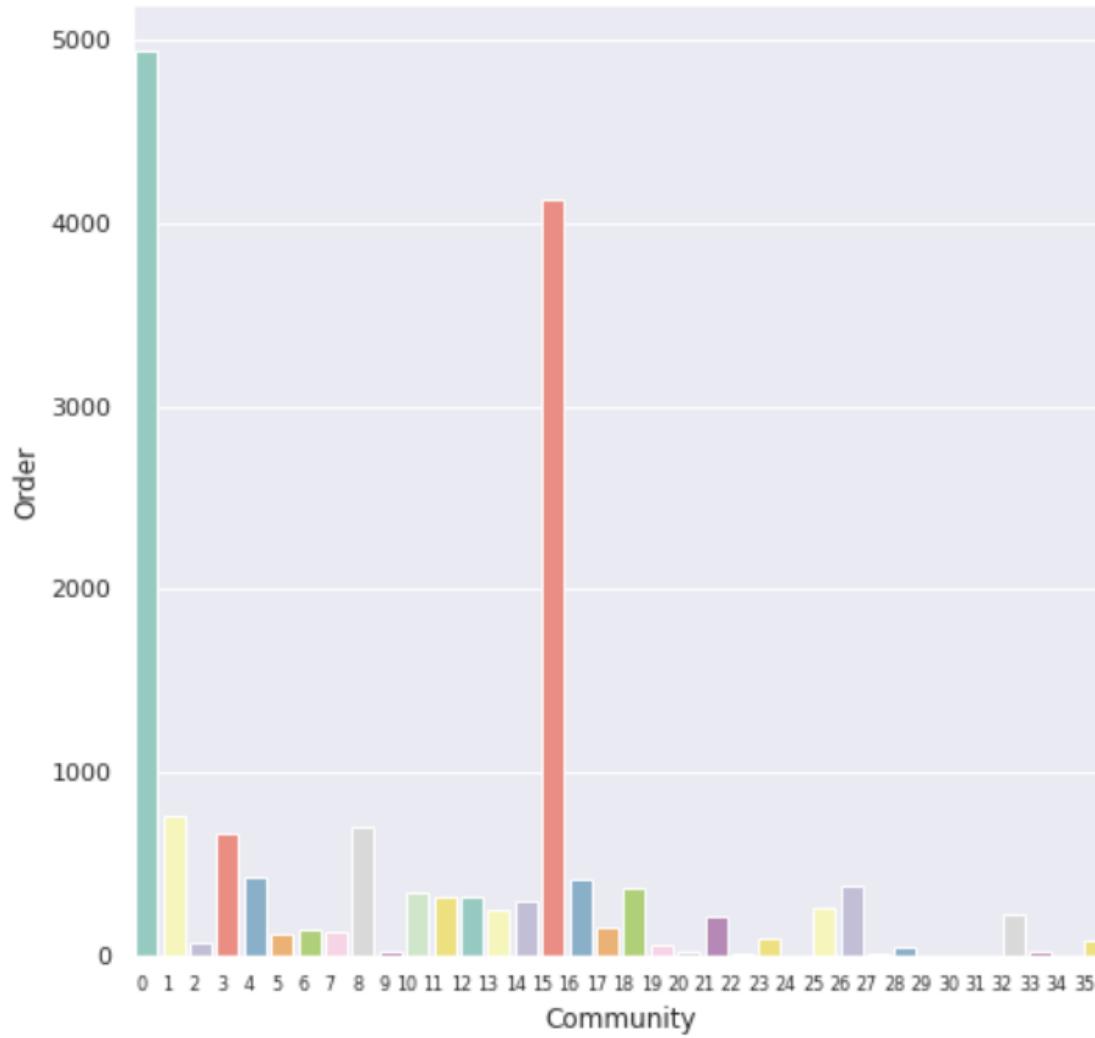


Metodo di Newman

Algoritmo ricorsivo di clustering divisivo basato sugli autovettori principali della matrice di modularità, per cui la massimizzazione della modularità viene risolta in termini di autospettro di tale matrice

Individua 36 comunità, ancor più sbilanciate.





Propagazione asincrona dell'etichetta

Inizializza ciascun vertice con un'etichetta unica



Assegna label più frequente nel vicinato

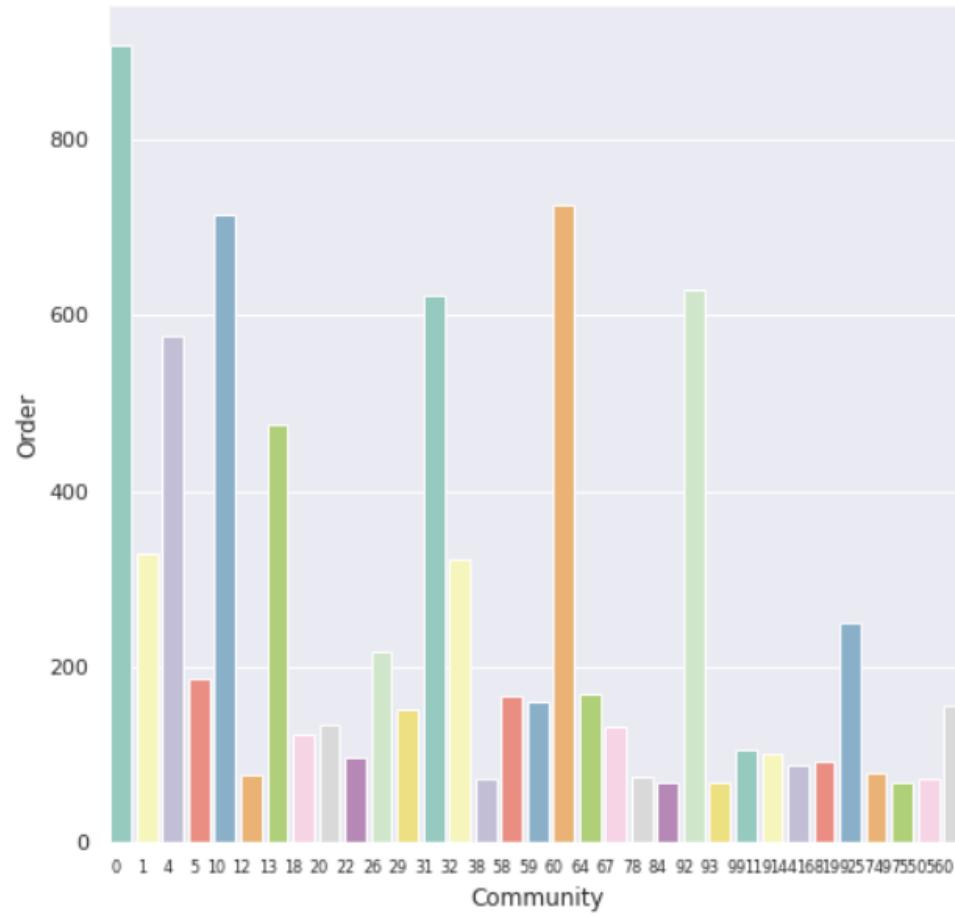


Stop quando la condizione è verificata in tutti i nodi



Pr

a



Rileva 1551 comunità di dimensione ridotta.

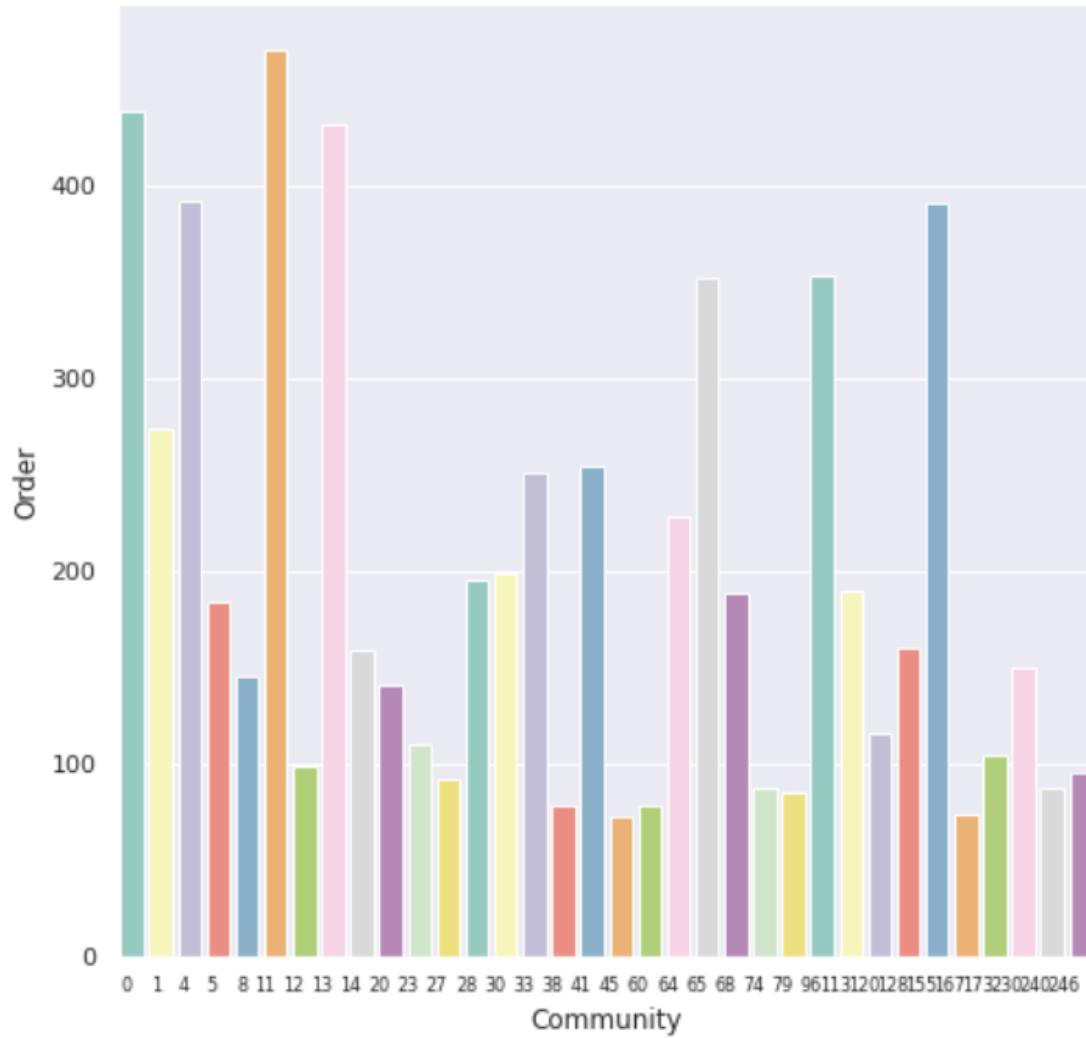


Metodo Information Map

**Sfrutta l'algoritmo di mappatura dell'informazione elaborato da
Rosvall e Bergstrom**

Trova 853 comunità, meno sbilanciate rispetto a quanto visto finora





La selezione della miglior partizione avviene con l'indice di copertura: ratio fra numero di legami interni a ciascuna comunità e totale di archi nel grafo

	Comunità	Copertura
Louvain's	42	0.66
Newman's	36	0.54
InformationMap	853	0.52
ALP	1551	0.48



L'analisi dei risultati è focalizzata sui 12 gruppi con numero di utenti superiore all'ordine medio





Si è proceduto al calcolo dei pesi **TF-IDF** per ciascun termine, considerando come documenti unici i set di testi per ogni comunità

Questa operazione non è però stata sufficiente ad individuare i topic tipici di ogni preciso gruppo

È stato necessario esaminare più in profondità i pattern semantici, stimando un modello del linguaggio con il metodo Latent Dirichlet Allocation

Grazie a questo modello, sono stati individuati i termini che identificano i 12 topic più rappresentativi di ciascun cluster

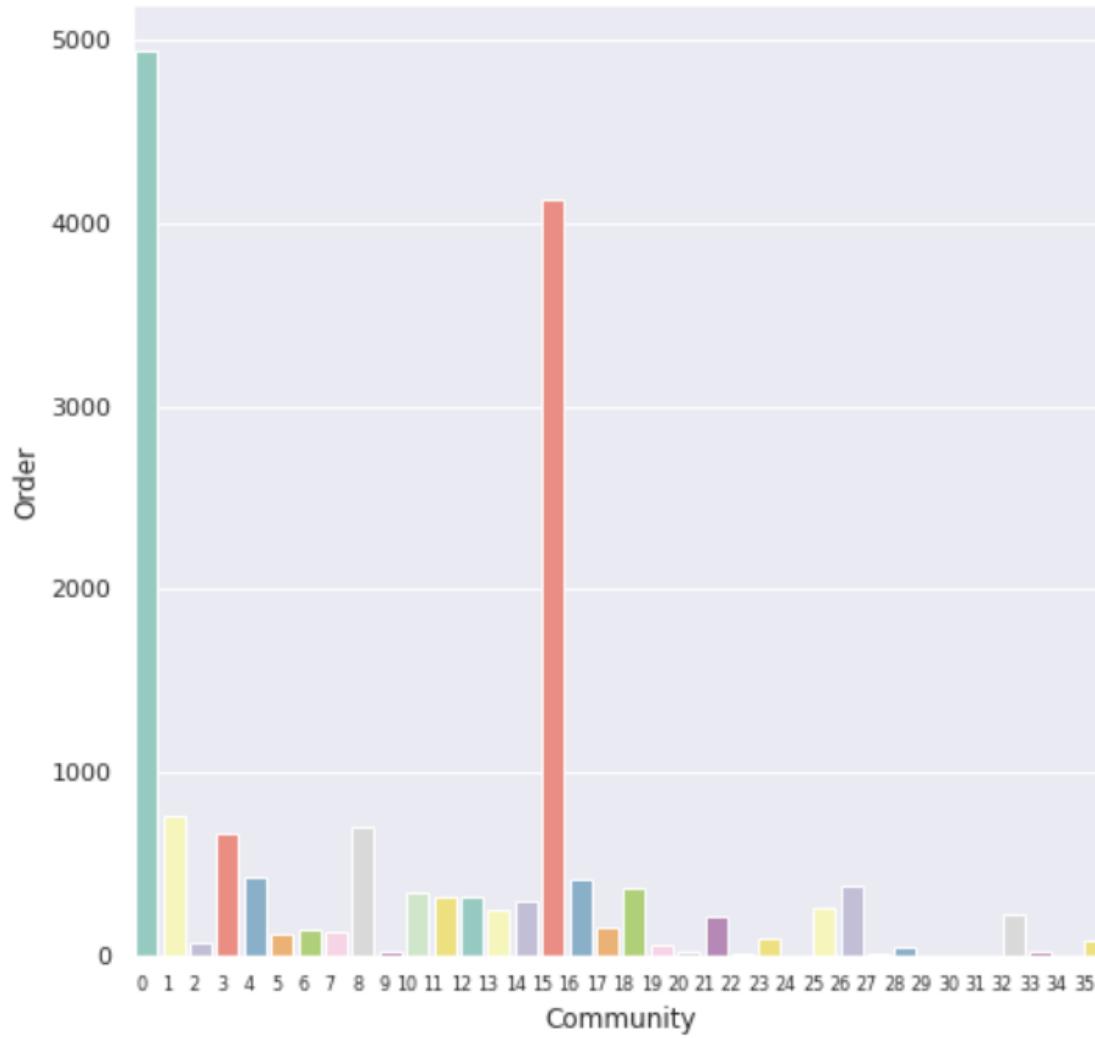


Edge-dominating set

È il subset di archi F tale per cui ogni legame in esso non compreso
è incidente all'estremo di almeno un arco di F

In questo caso, il set dominante non contiene alcun bot





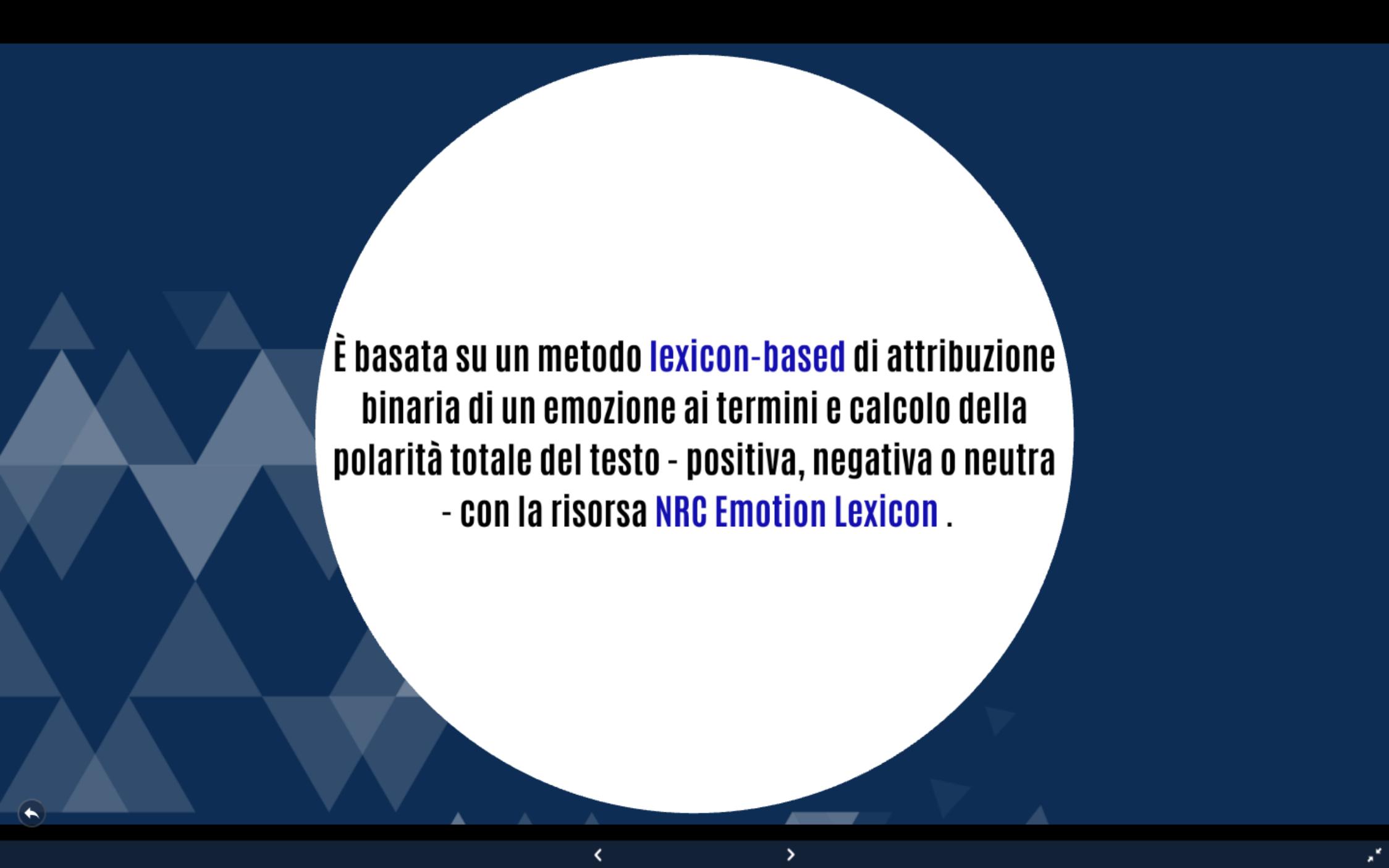
La selezione della miglior partizione avviene con l'indice di copertura: ratio fra numero di legami interni a ciascuna comunità e totale di archi nel grafo

	Comunità	Copertura
Louvain's	42	0.66
Newman's	36	0.54
InformationMap	853	0.52
ALP	1551	0.48



Si è proceduto al calcolo dei pesi **TF-IDF** per ciascun termine, considerando come documenti unici i set di testi per ogni comunità

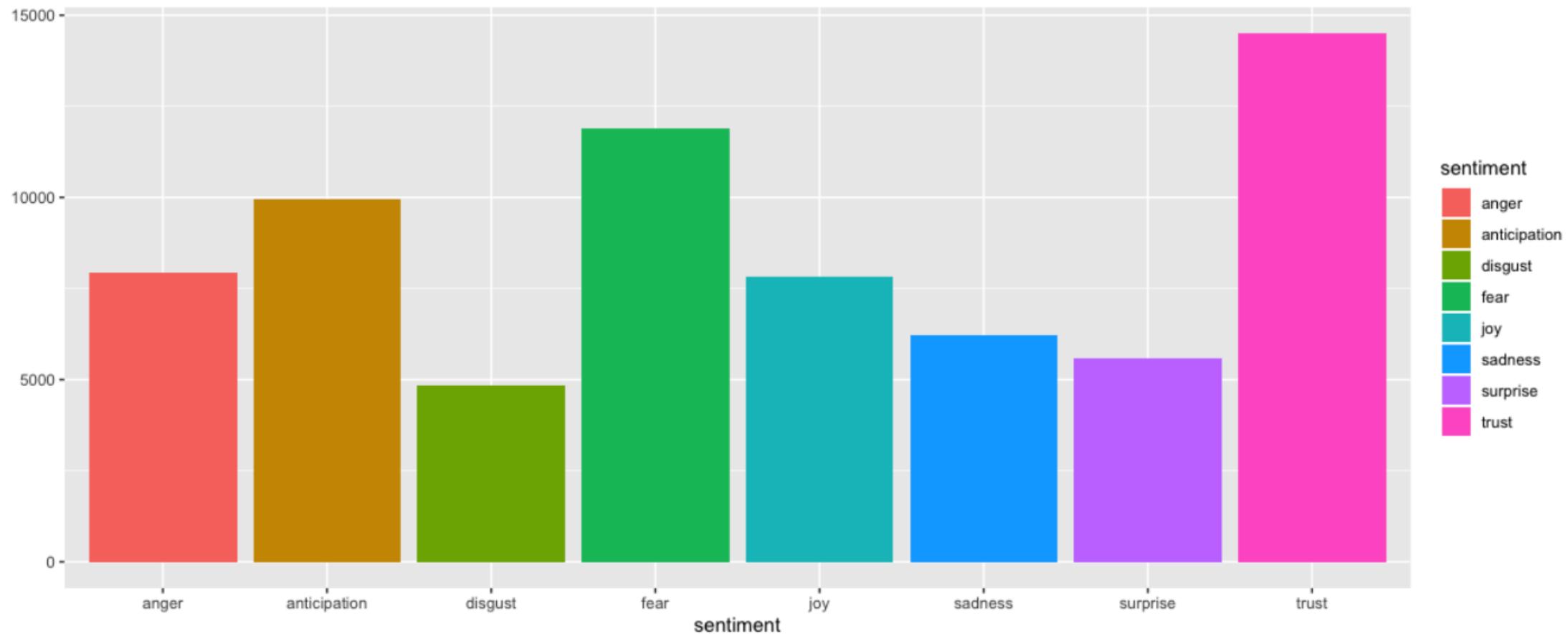
Questa operazione non è però stata sufficiente ad individuare i topic tipici di ogni preciso gruppo



È basata su un metodo **lexicon-based** di attribuzione binaria di un'emozione ai termini e calcolo della polarità totale del testo - positiva, negativa o neutra - con la risorsa **NRC Emotion Lexicon**.

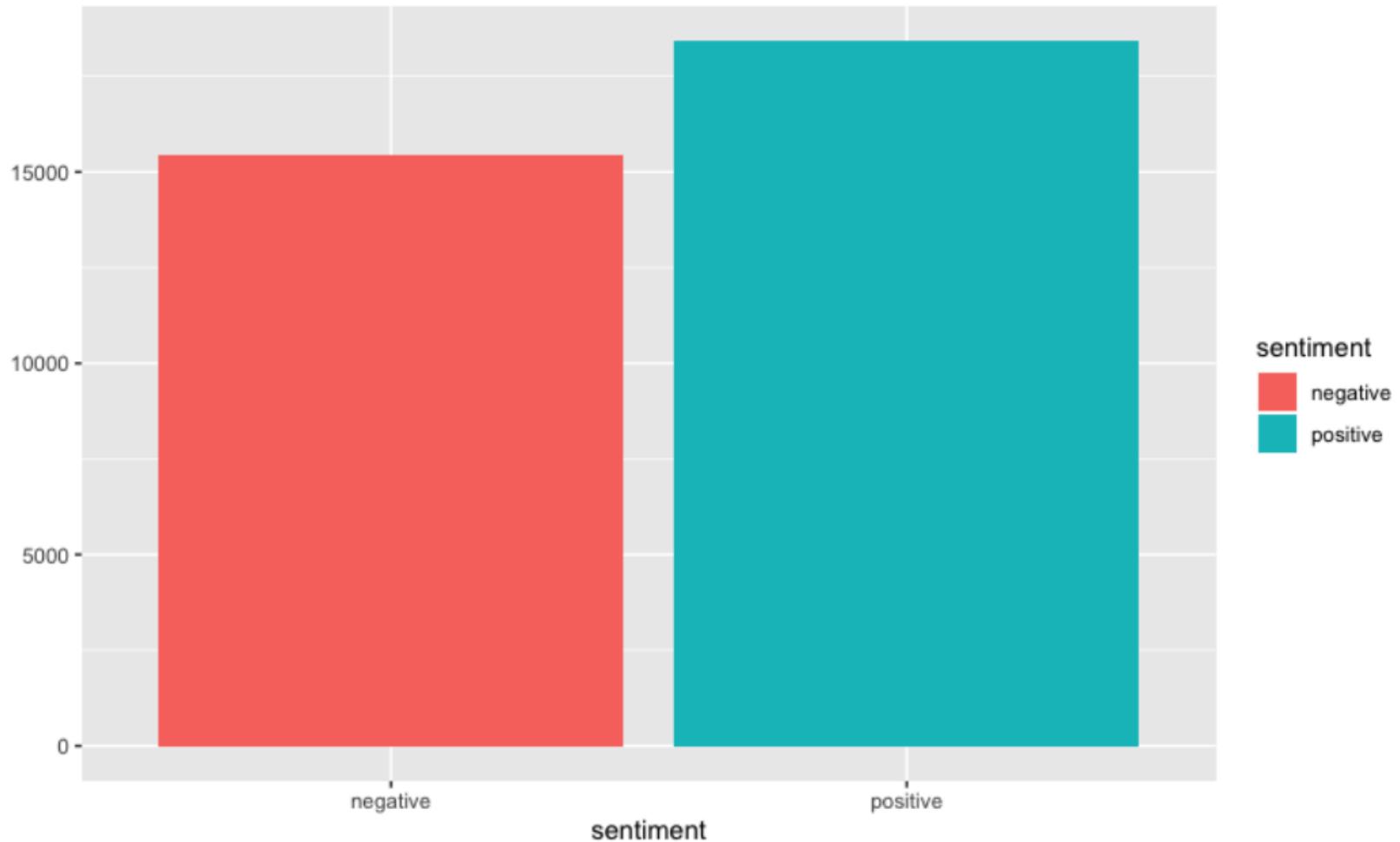
Frequenza totale delle emozioni

QAnon tweets - Emotions



Polarità a confronto

QAnon tweets - Polarity



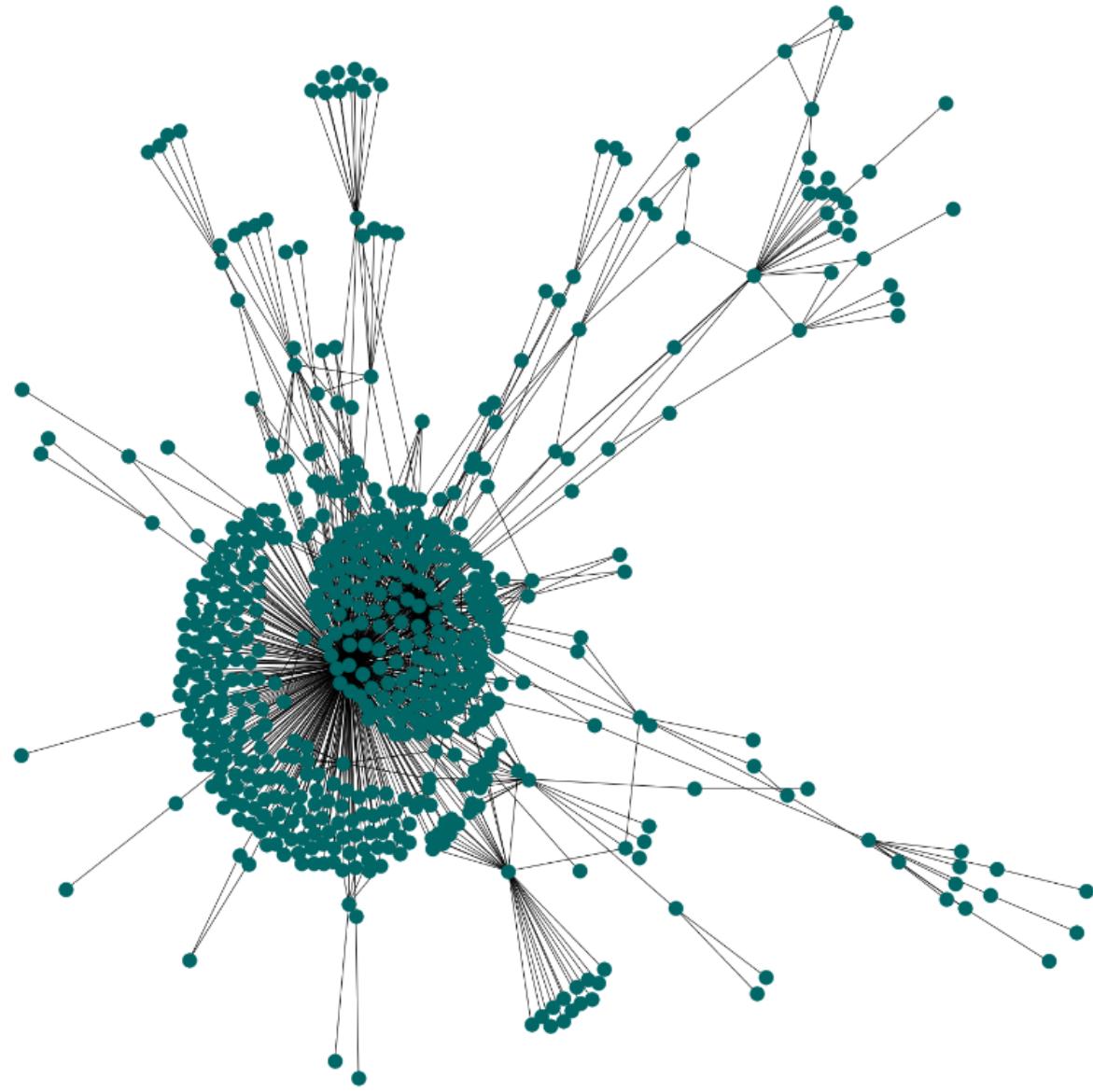
**Le visualizzazioni sono state
prodotte mediante libreria [NetworkX](#)**



Come layout si è optato per una tipologia spring, che posiziona i vertici in modo che tutti gli archi abbiano lunghezza costante e minimizzando gli incroci.

L'algoritmo itera la costruzione fino a una condizione di equilibrio.





Tuttavia, questo layout fatica nel proporre un disegno finale comprensibile al crescere del numero di vertici.

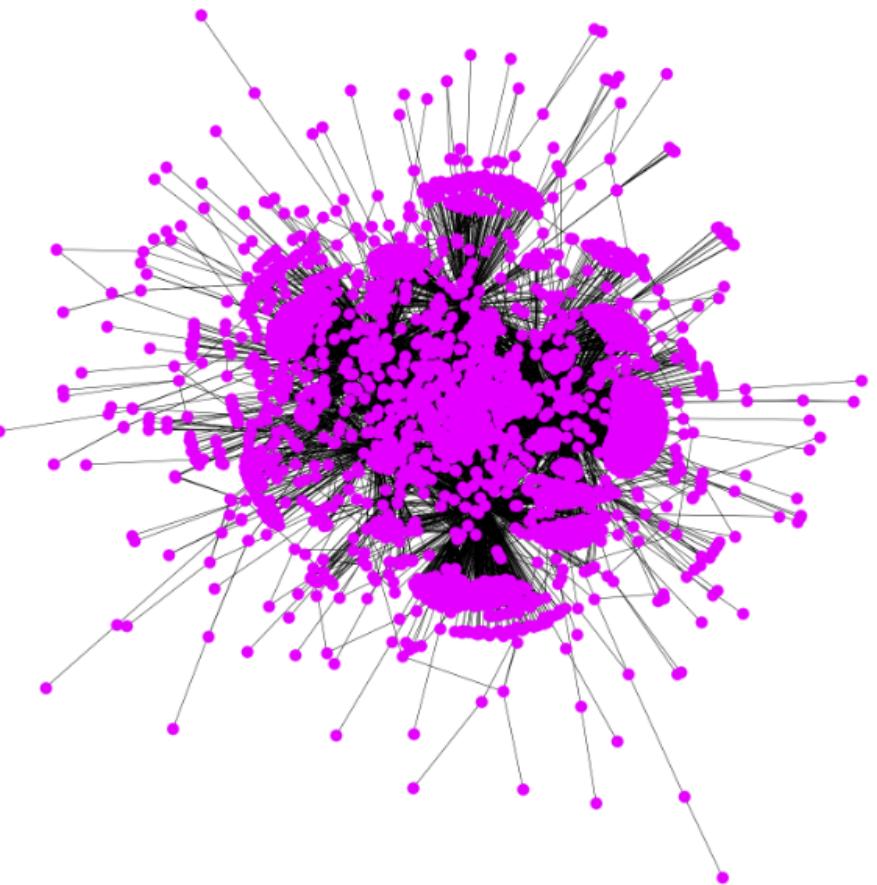
Per disegnare il grafo di partenza è stato usato il software GrapViz, che sfrutta un modello spring multi-scala.



I risultati mostrano la presenza di tre poli di attrazione nel dibattito: le comunità 0, 2 e 3.

Attorno ad esse si spargono nove comunità satellite.





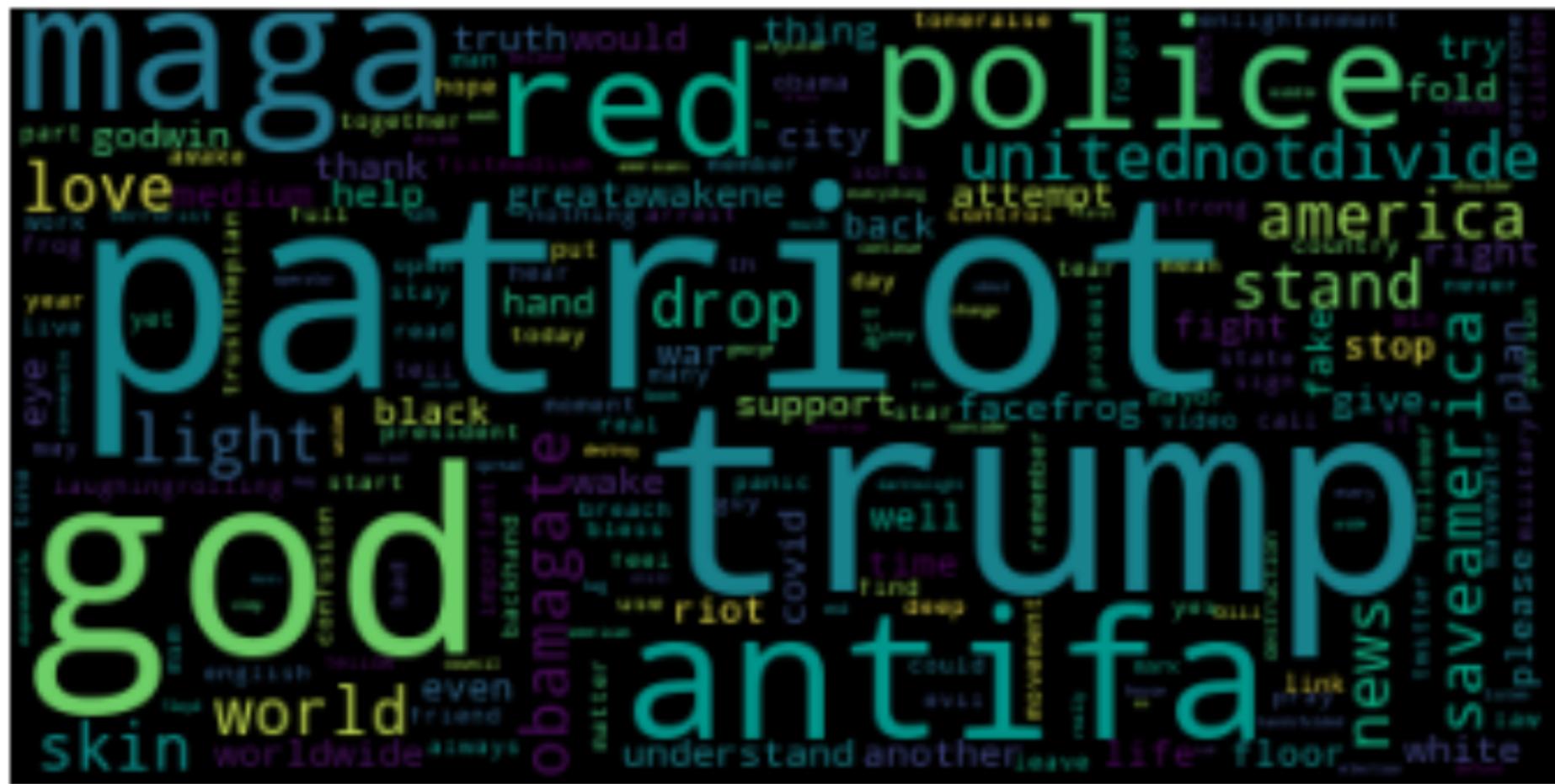
Comunità 0

- ✓ È la più popolosa (2862 nodi)
- ✓ Coeff di clustering massimo
- ✓ Densità minima
- ✓ È la rete più disassortativa (-0.57)
- ✓ Tuttavia, l'efficienza è moderata (0.37)

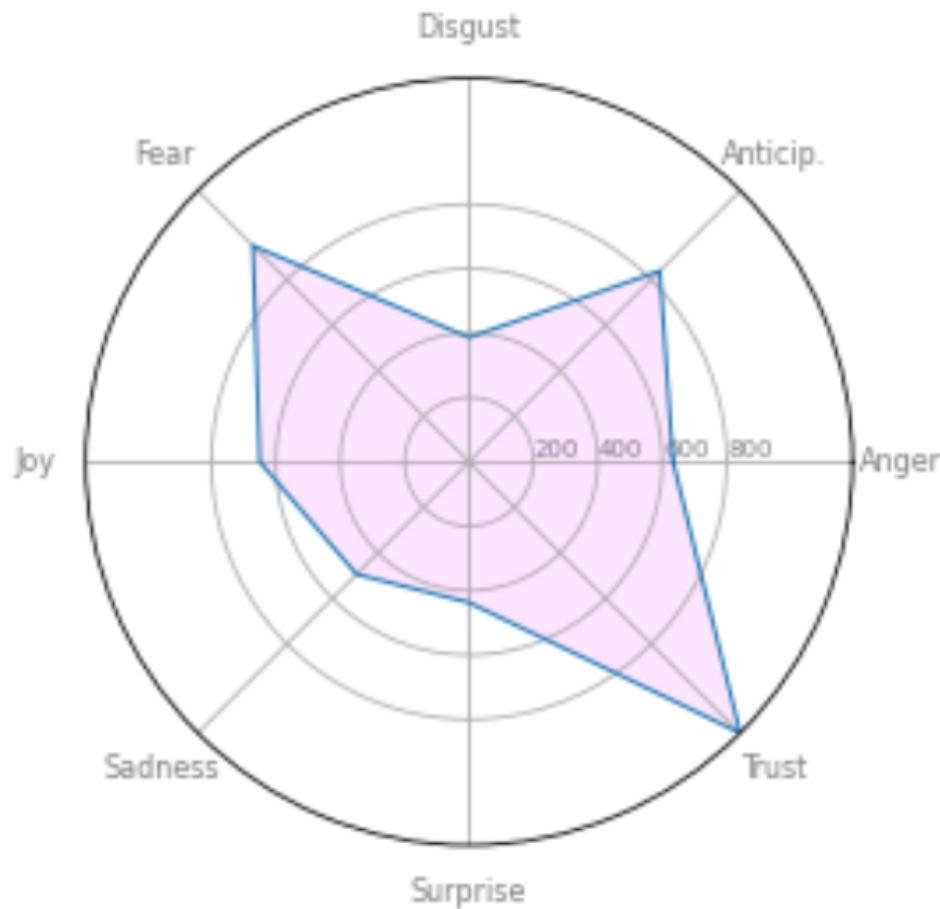
È composta dai sostenitori più attivi della teoria cospirativa

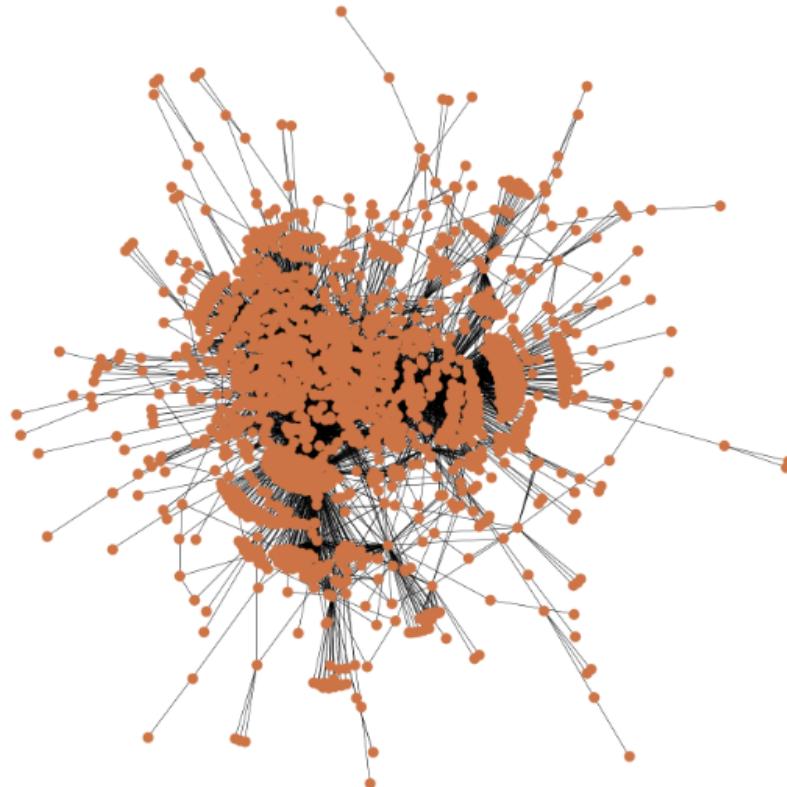


Topic comunità 0



Sentiment Analysis sul gruppo O

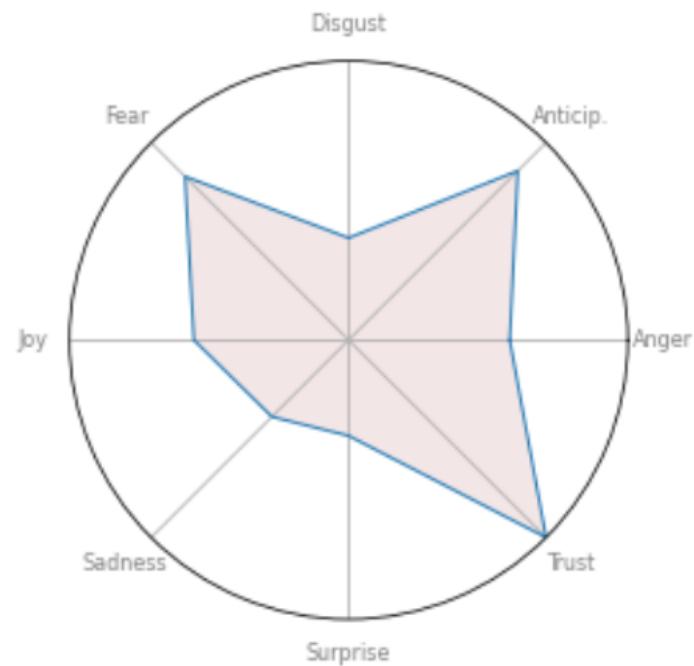




Comunità 2

- ✓ È la seconda più popolosa (2036 nodi)
- ✓ Ha connettività moderata, a livello globale e locale
- ✓ Disassortatività e efficienza minori rispetto al gruppo 0
- ✓ I temi appaiono simili a quelli della c. 0, con maggior attenzione al tema politico





Lo spettro emotivo non mostra differenze rilevanti nella composizione, a parte una crescita del livello di aspettativa.



Comunità 3

- ✓ È la terza più popolosa (1015 nodi)
- ✓ A liv. locale la connettività è minore rispetto a 0 e 2, ma a liv. globale è nettamente superiore
- ✓ L'efficienza è paragonabile a quella di 0 e 2
- ✓ I legami sono meno sistematici: coeff di assortatività minimo (-0.18)

Nonostante i legami non siano influenzati dalla popolarità, i nodi centrali sono gli account ufficiali di Donald Trump e la Casa Bianca



Sebbene la comunità sia legata a figure formali della politica americana, il dibattito continua ad assumere toni bizzarri: i topic sono quasi identici a quelli del primo cluster

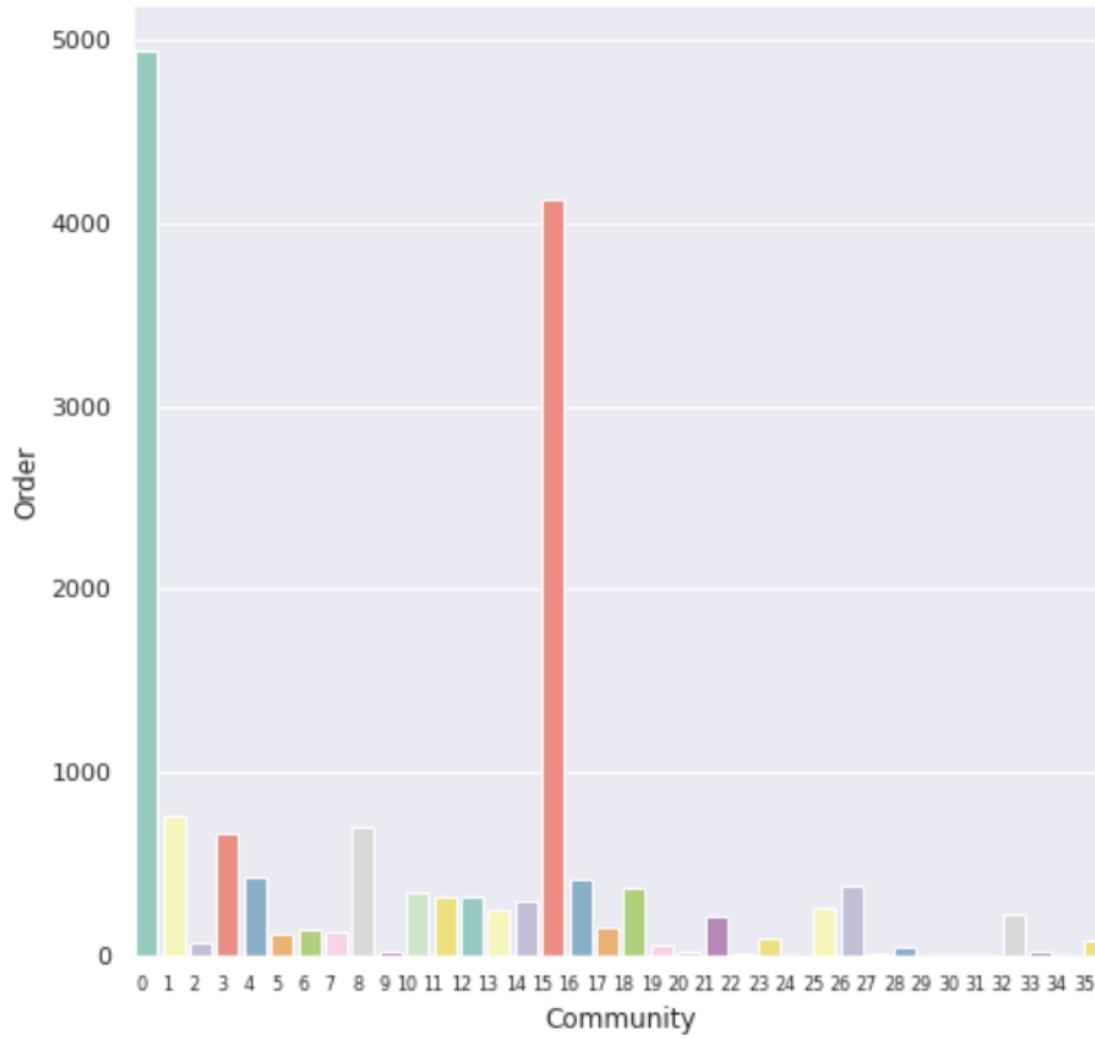


Edge-dominating set

È il subset di archi F tale per cui ogni legame in esso non compreso
è incidente all'estremo di almeno un arco di F

In questo caso, il set dominante non contiene alcun bot



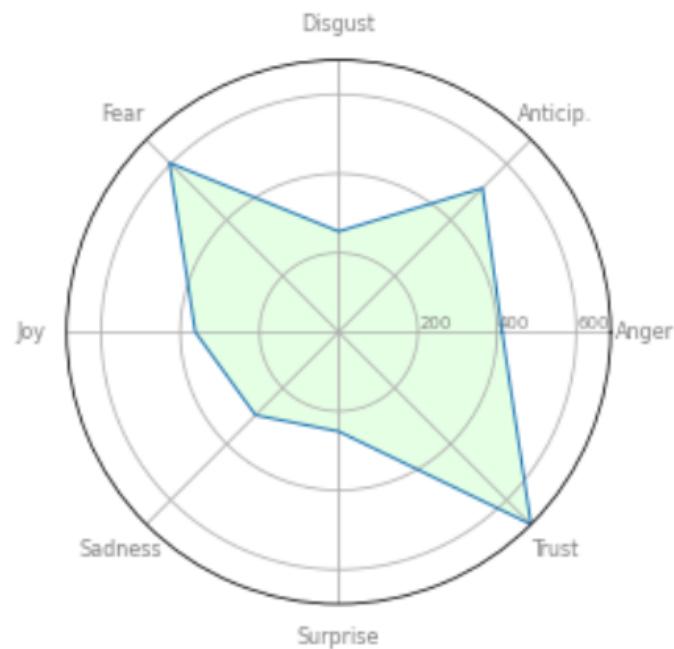


La selezione della miglior partizione avviene con l'indice di copertura: ratio fra numero di legami interni a ciascuna comunità e totale di archi nel grafo

	Comunità	Copertura
Louvain's	42	0.66
Newman's	36	0.54
InformationMap	853	0.52
ALP	1551	0.48



Comunità 3



Aumenta leggermente il livello di rabbia e discretamente la presenza di uno stato d'animo di paura.



Comunità satellite

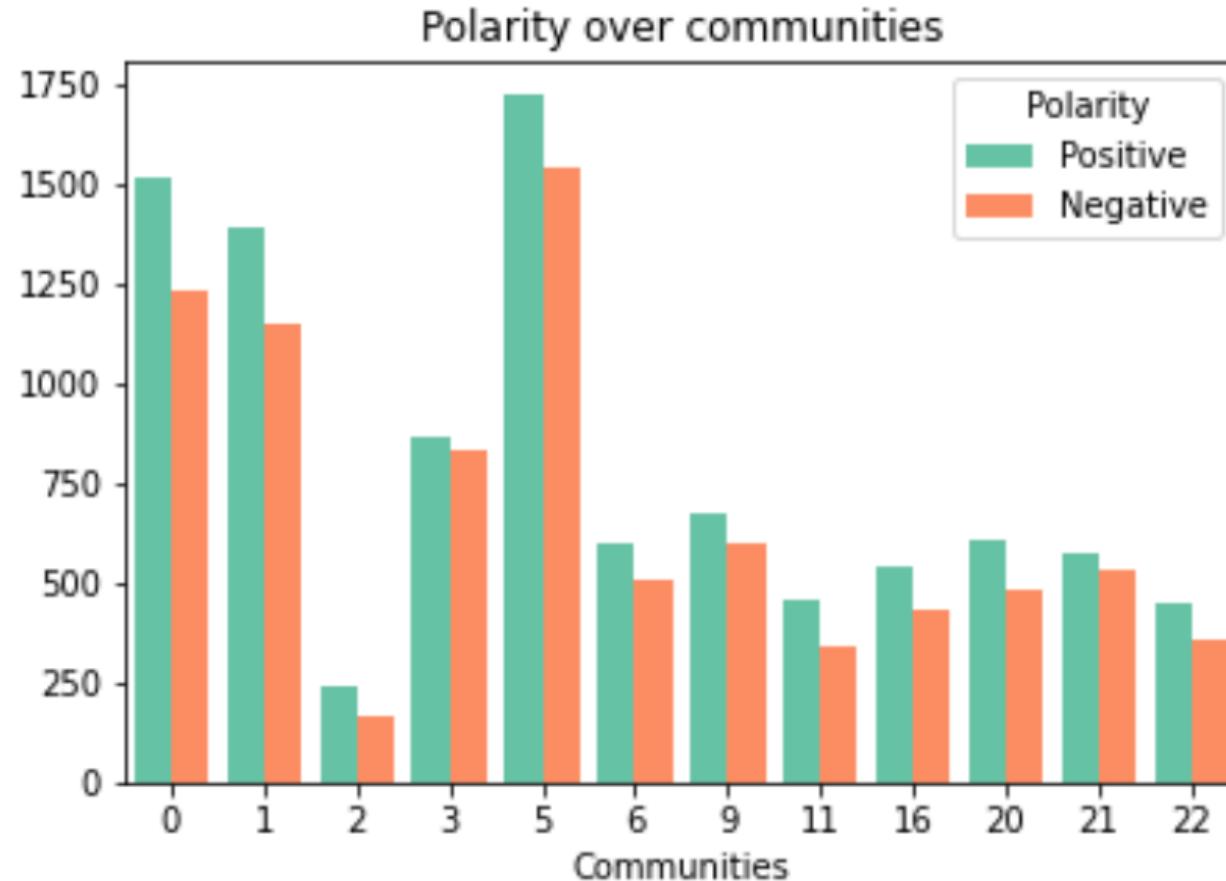
Le nove comunità rimanenti sono coerenti per quanto riguarda i topic principali, ma ne affrontano sfaccettature particolari, riferendosi ad eventi o precisi fatti di cronaca

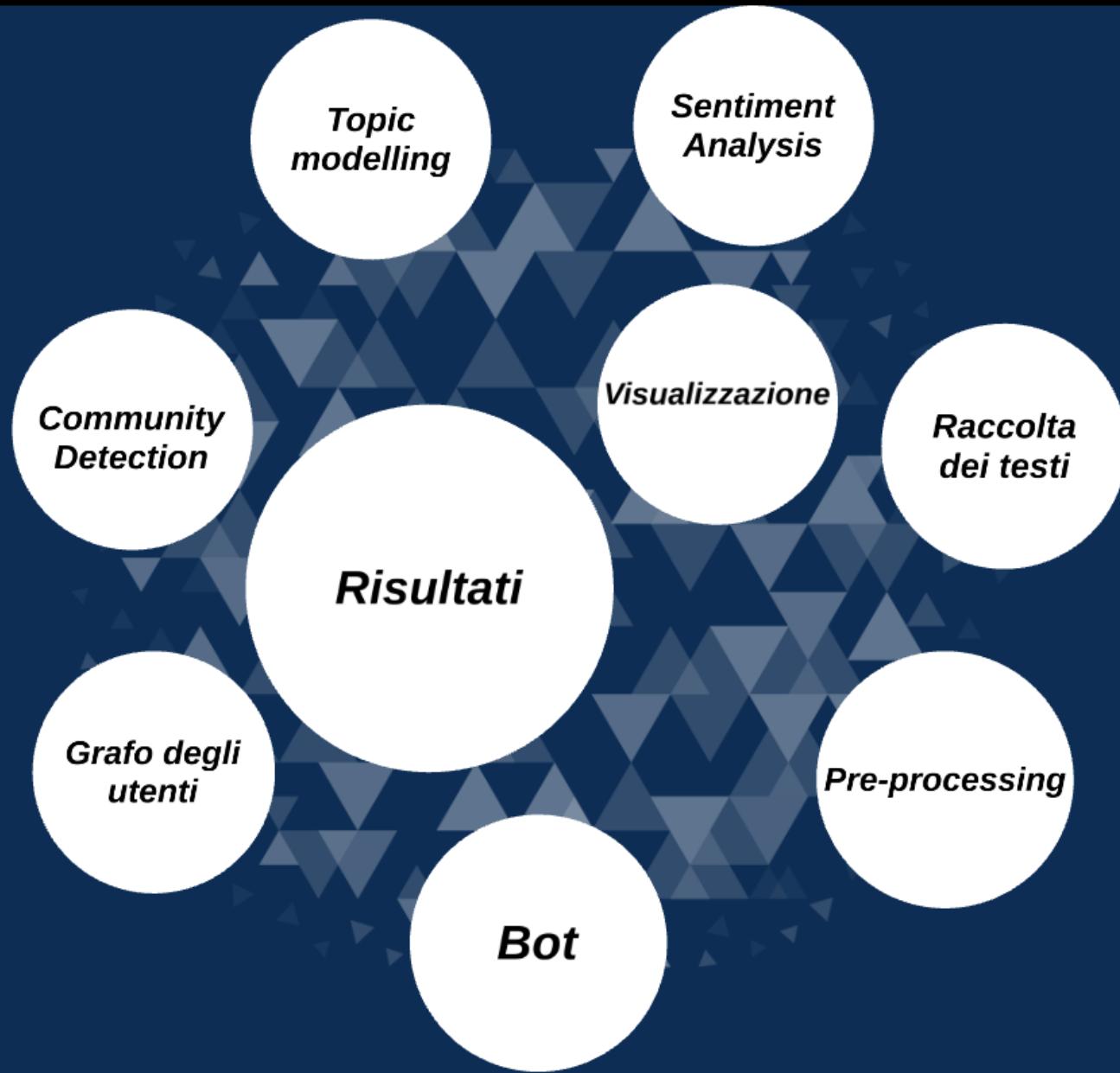
Il numero di membri rimane sotto le 800 unità, e le varie strutture sono eterogenee.

Si nota una correlazione assoluta del 50% fra i liv. di assortatività e l'efficienza, e superiore al 60% fra quest'ultima e la variabile di transitività.



Diversa polarizzazione fra i gruppi di utenti





Movimento QAnon Analisi del dibattito online