SANREMO FESTIVAL 2024

FROM DIFFERENT LENSES

Course: Network Science 23-24
Group: Cristian Granchelli, Nour Al Housseini, Hazeezat Adebayo

OVERVIEW

WHY SANREMO?

Being one of the most important music festivals in Italy, it generates thousands of data thus attarcting attention.

- Economically, culturally, and touristically important.
- A hot topic among generations with different perspecrives, due to differences in homophily of preference.
- Local and international coverage.

DATA COLLECTION

Dataset retrieval started after authenticating with Reddit API. The process happened through three important parameters:

- client_id
- client_secret
- user_agent.

1

Reddit data was retrieved according to the following parameters: subreddit name, total posts, time filter, sort mode, keywords, and batch size.

DATA COLLECTION

1

Subreddits related to Sanremo were retrieved and filtered to keep only relevant information. Posts were collected from these subreddits according to these parameters:

total_posts_to_retrieve = 1000
time_filter_2024 = year
sort_mode_2024 = 'hot'
keywords_2024 = 'Sanremo 2024 OR Vincitore Sanremo 2024'

Comments of the posts were collected selecting only the posts with more than 20 upvotes.

DATA PROCESSING

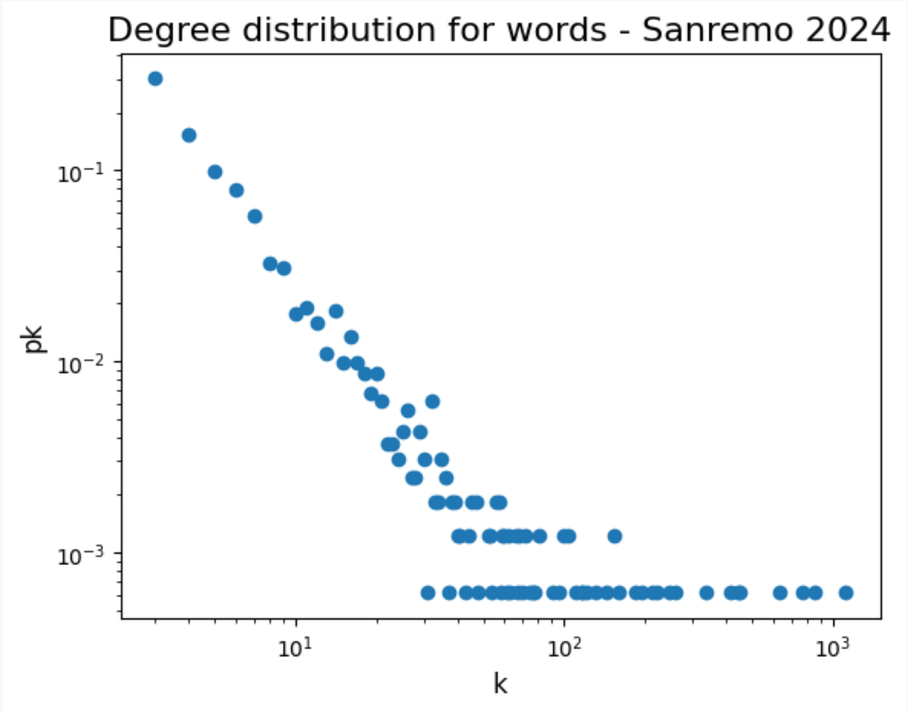
After obtaining the data they have been stored in a data frame and preprocessed using superficial and deep cleaning and hashtag removal.

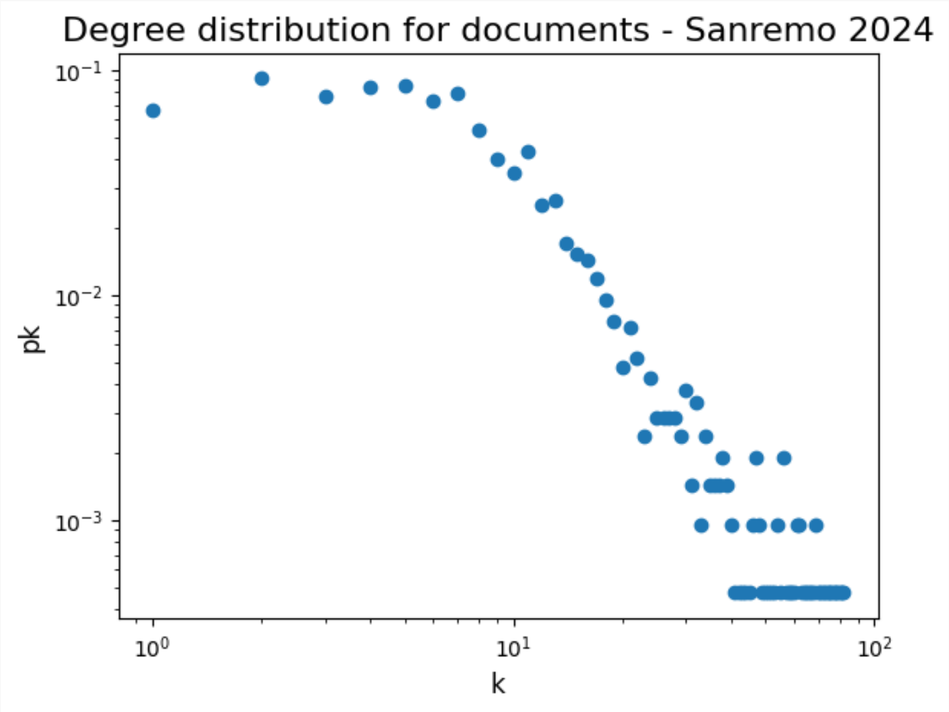
	id	created	body	text_sup_clean	text_deep_clean	text_deep_clean_pos	hashtags
0	kpkehh1	2024-02-09	Ore 01:36, 4611 commenti!\n\n**Risultati della...	ore commenti risultati della terza serata ha v...	ore commenti risultati della terza serata vota...	[ore NOUN, commenti PROPN, risultati PROPN, de...	[]
1	kpjfwdd	2024-02-08	Questo era Mr.Rain da giovane. Feel old yet?\n...	questo era mr rain da giovane feel old yet	questo era mr rain da giovane feel old yet	[questo PROPN, era NOUN, mr PROPN, rain NOUN, ...	[]
2	kpjvbo4	2024-02-08	RUSSEL PAZZO UOMO HA NOMINATO IL QUA QUA GATE	russel pazzo uomo ha nominato il qua qua gate	russel pazzo uomo nominato il qua gate	[russel NOUN, pazzo PROPN, uomo PROPN, nominat...	[]
3	kpj9o1y	2024-02-08	Ah si, Mameli - il twink che sognò l'Italia	ah si mameli il twink che sogno l italia	si mameli il twink che sogno l italia	[si PROPN, mameli PROPN, il PROPN, twink PROPN...	[]
4	kpjg6ii	2024-02-08	Per citare un saggio:\n\nhttps://preview.redd....	per citare un saggio	citare un saggio	[citare NOUN, un PROPN, saggio PROPN]	[]
...	...	...	...	...	...	...	...
2193	kp8wzzs	2024-02-06	Is anyone still watching lol??	is anyone still watching lol	still watch lol	[still ADV, watch VERB, lol NOUN]	[]

2

DEGREE DISTRIBUTIONS

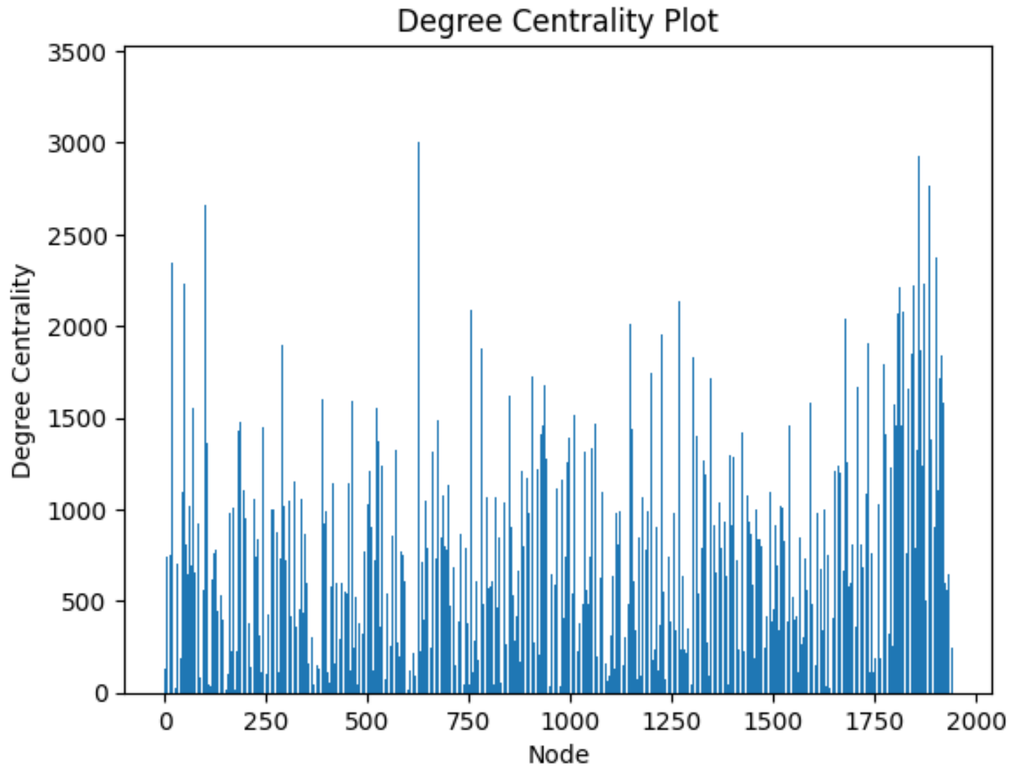
To retrieve the degree distributions, we need to build the semantic matrices for collections of words and hashtags and then join them together to form documents. After plotting word occurrences, we add a second step of processing involving the removal of one/two time appearing words.





CENTRALITY

Degree Centrality:



Top 10 Highest Degree Centrality Values:

3360

3210

3064

3048

3040

3000

2970

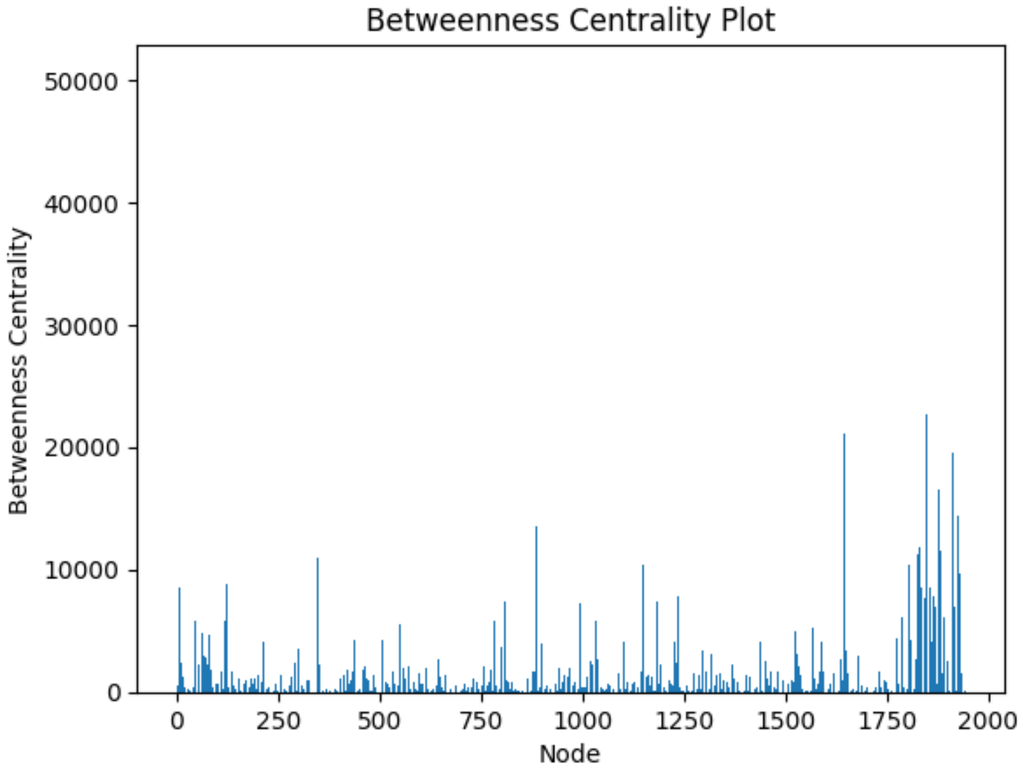
2934

2930

2926

CENTRALITY

Betweenness Centrality:



Top 10 Highest Betweenness Centrality Values:

50304.47113828503

41312.109544414154

32477.71634325816

31460.13706840102

30559.06396584049

30381.137785390943

28134.351218647498

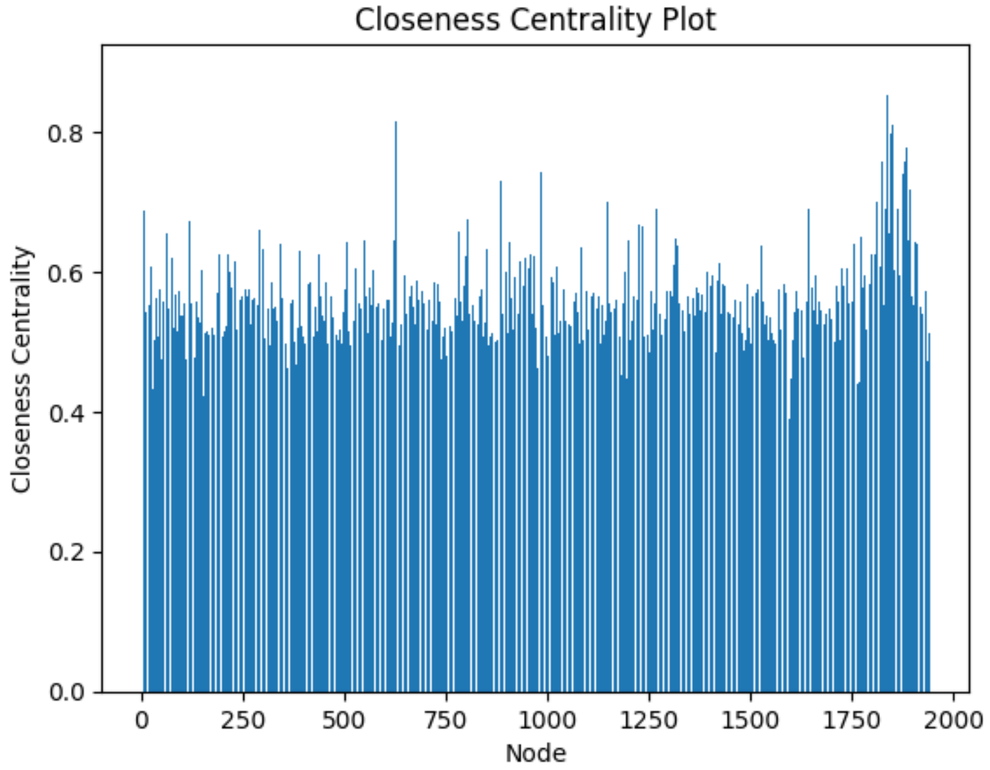
27162.9345800752

26124.32178384724

25869.06177732122

CENTRALITY

Closeness Centrality:

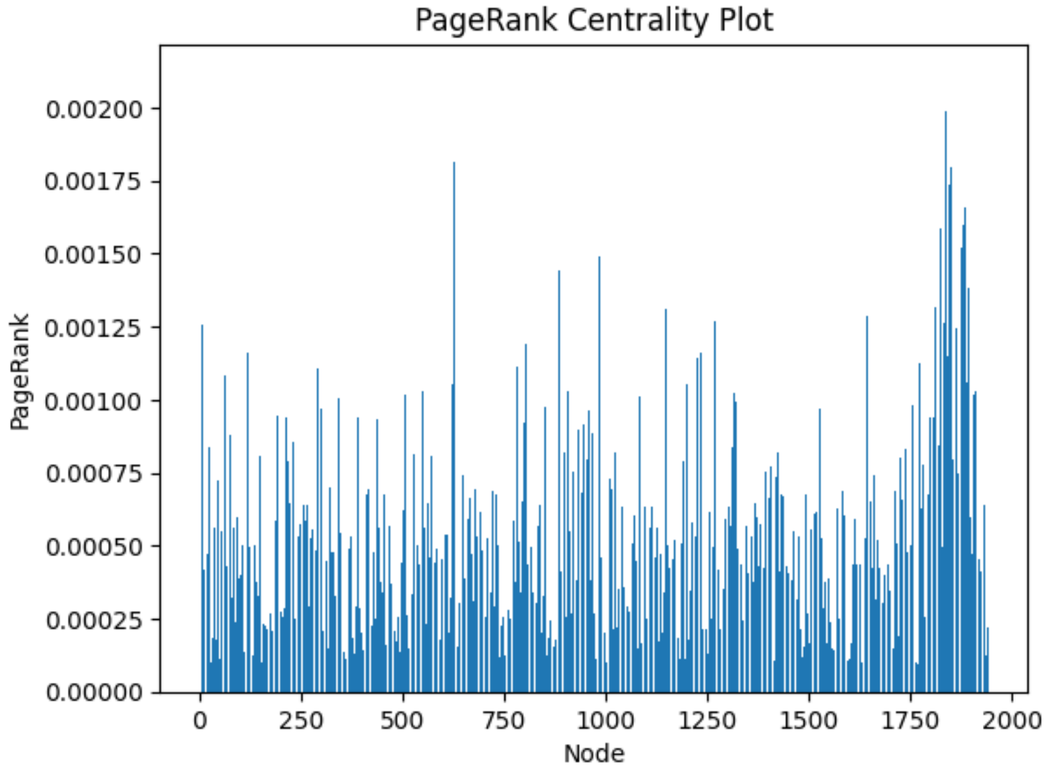


Top 10 Highest Closeness Centrality Values:

0.8807256235827664
0.8517543859649123
0.8253293667658309
0.8225328250741212
0.8211416490486257
0.8142557651991614
0.8091666666666667
0.8031430934656741
0.8024793388429752
0.8018166804293972

CENTRALITY

PageRankCentrality:



Top 10 Highest Pagerank Values:

0.002109832601702887

0.0019899519554752807

0.0018655067310633237

0.0018604381775092062

0.0018586848782031985

0.0018120582428785847

0.001795684560893403

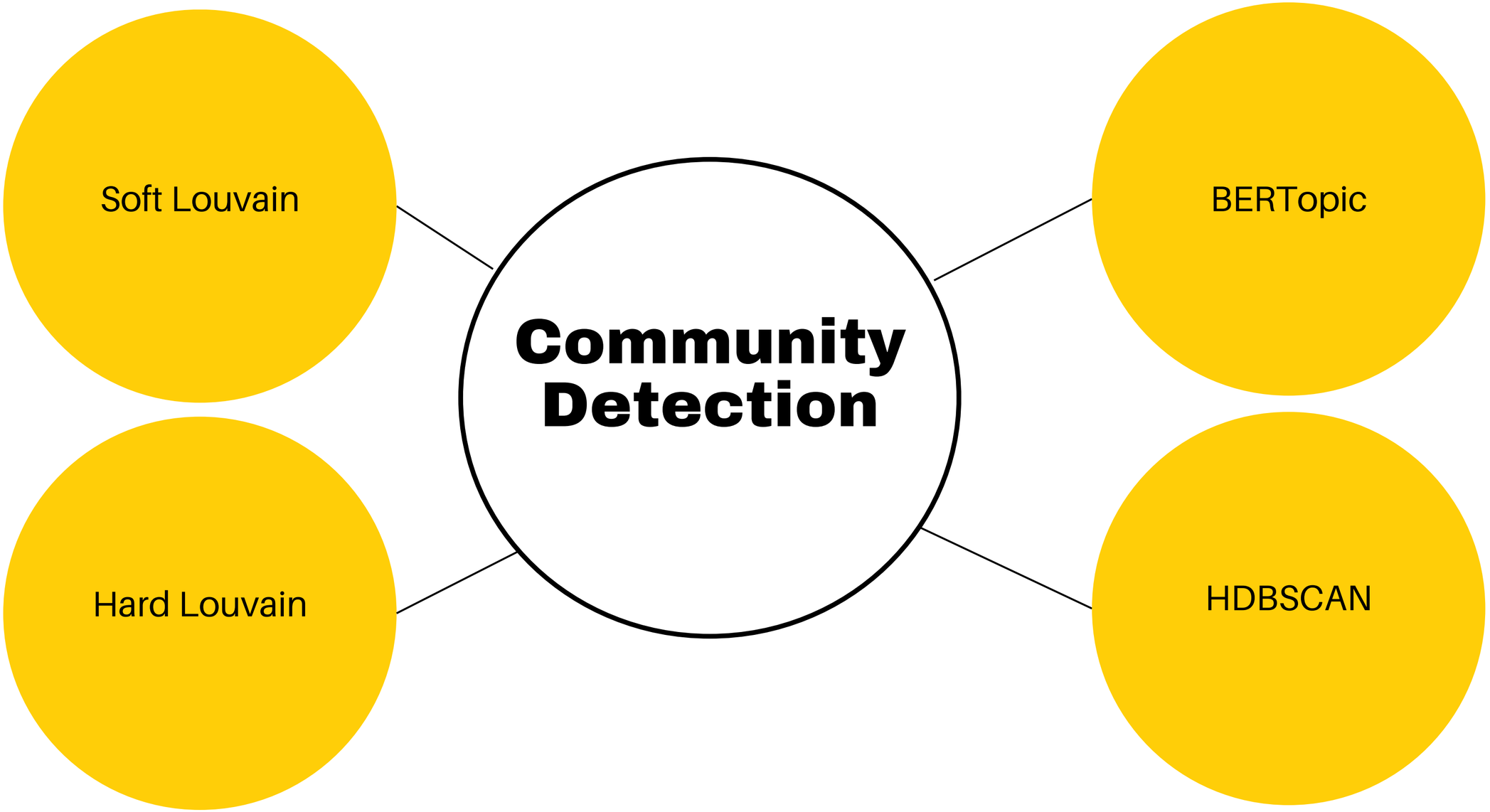
0.001791633435926497

0.0017650977557404943

0.00176366954855365

CENTRALITY

Most of the comments with high values for the centrality measures were related to the political statements given by singers during the competition and the consequences of these comments.


Soft Louvain
Hard Louvain
Community Detection
BERTopic
HDBSCAN

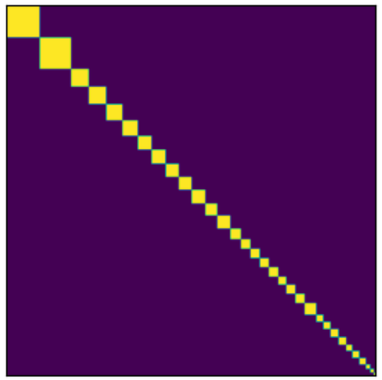
HDBSCAN

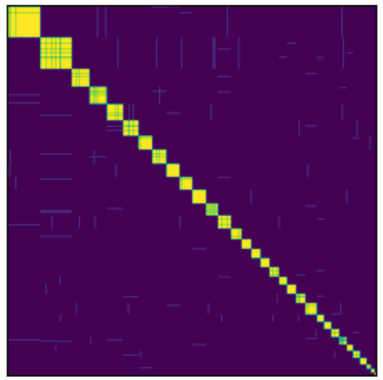
The HDBSCAN algorithm was implemented by:

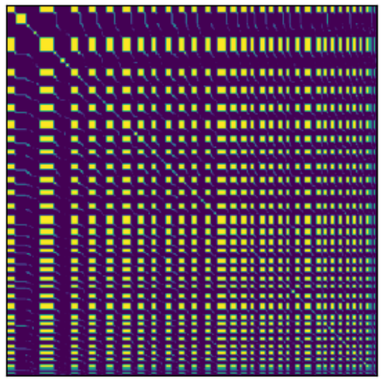
- Creating a feature matrix using the PageRank values
- Applying the HDBSCAN algorithm with multiple min_cluster_size values
- Filtering only the community assignment that gave less than 40 communities
- Choosing the best assignments according to the modularity and a composite index made by NMI, Modularity, Ncut and InfoMap.

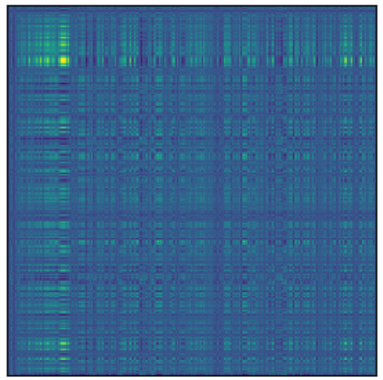
LOUVAIN & BERTOPIC

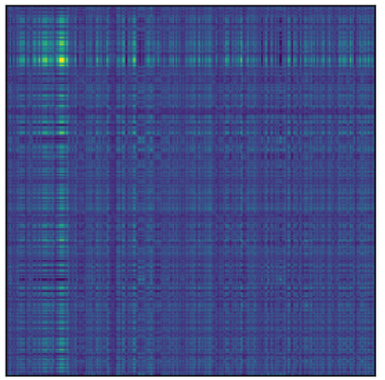
- We used the Louvain algorithm to find a partition of the nodes based on modularity.
- Hard Louvain produces non-overlapping communities, while soft Louvain refines these communities allowing nodes to belong to multiple groups. Both methods yield sparse matrices, representing the network’s community structure.
- The Bertopic algorithm was implemented to identify thematic communities in text data which reveals topics and discussions related to the Sanremo festival.











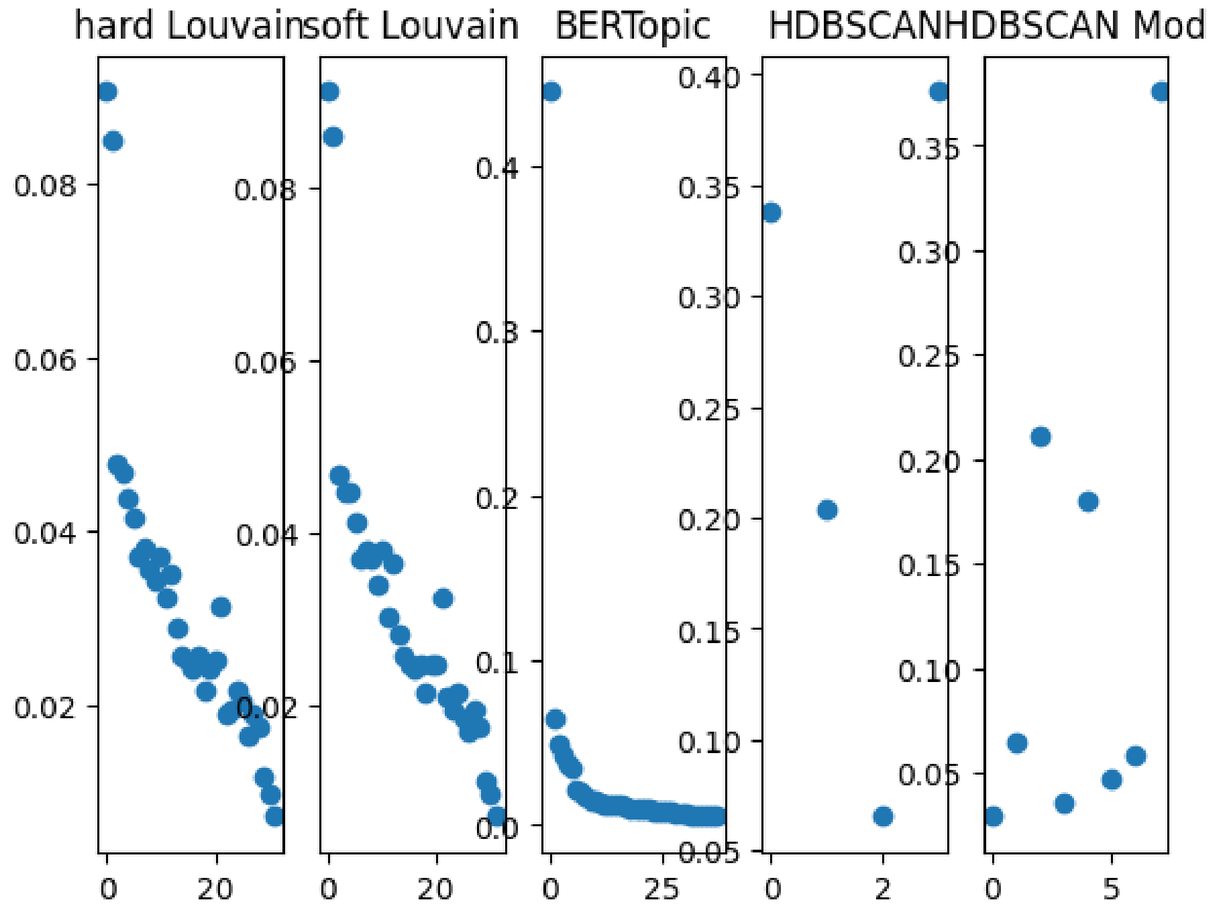
Hard Louvain
32 **communities**
NMI: 0.481053
Q: 0.320362
Ncut: 0.633845
InfoMap: 0.012399

Soft Louvain
31 **communities**
NMI: 0.480640
Q: 0.321070
Ncut: 0.630844
InfoMap: 0.012660

BERTopic
34 **communities**
NMI: 0.405393
Q: 0.161195
Ncut: 0.760728
InfoMap: 0.055721

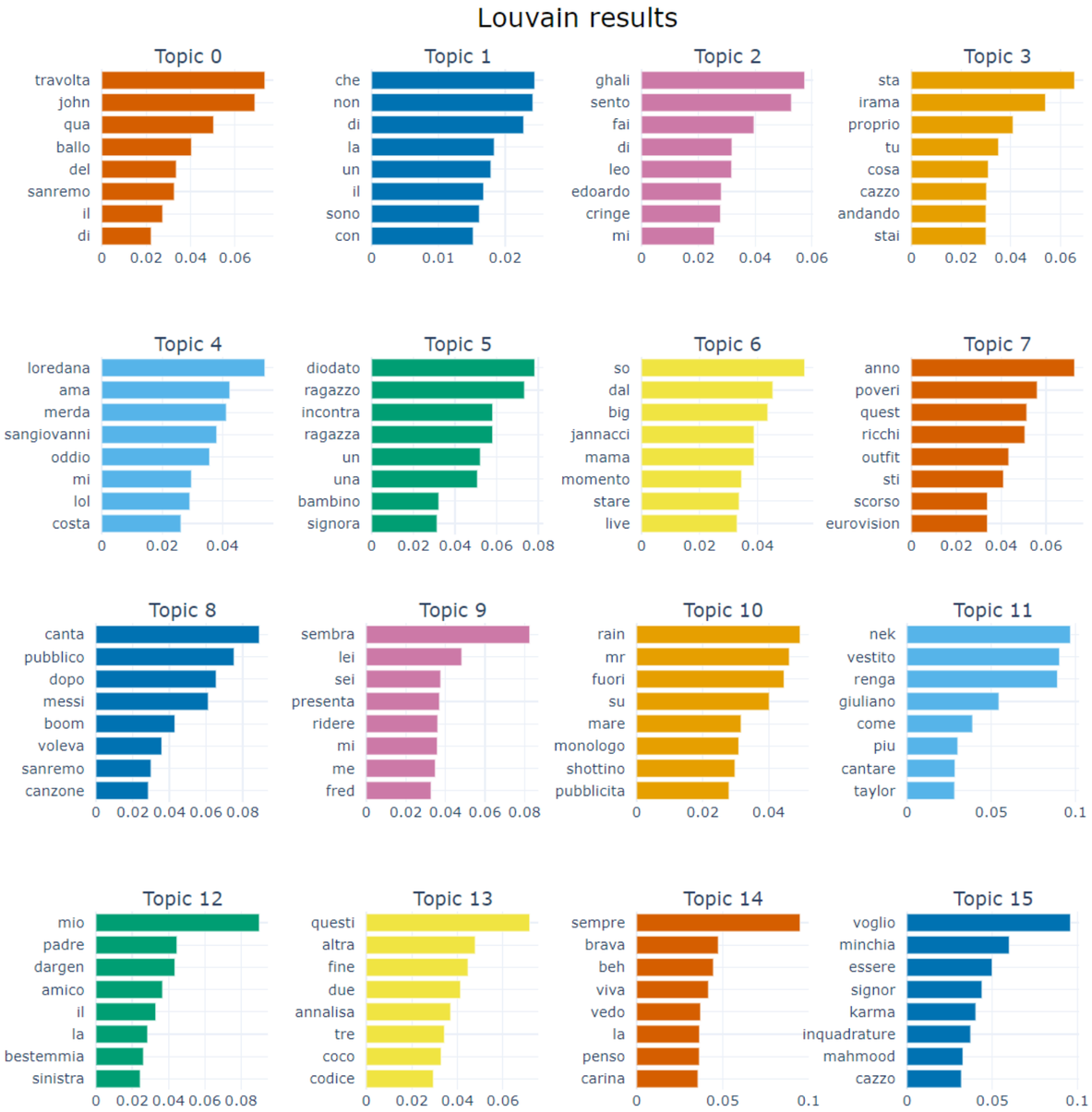
HDBSCAN
4 **communities**
NMI: 0.254881
Q: 0.135648
Ncut: 0.596759
InfoMap: 0.068682

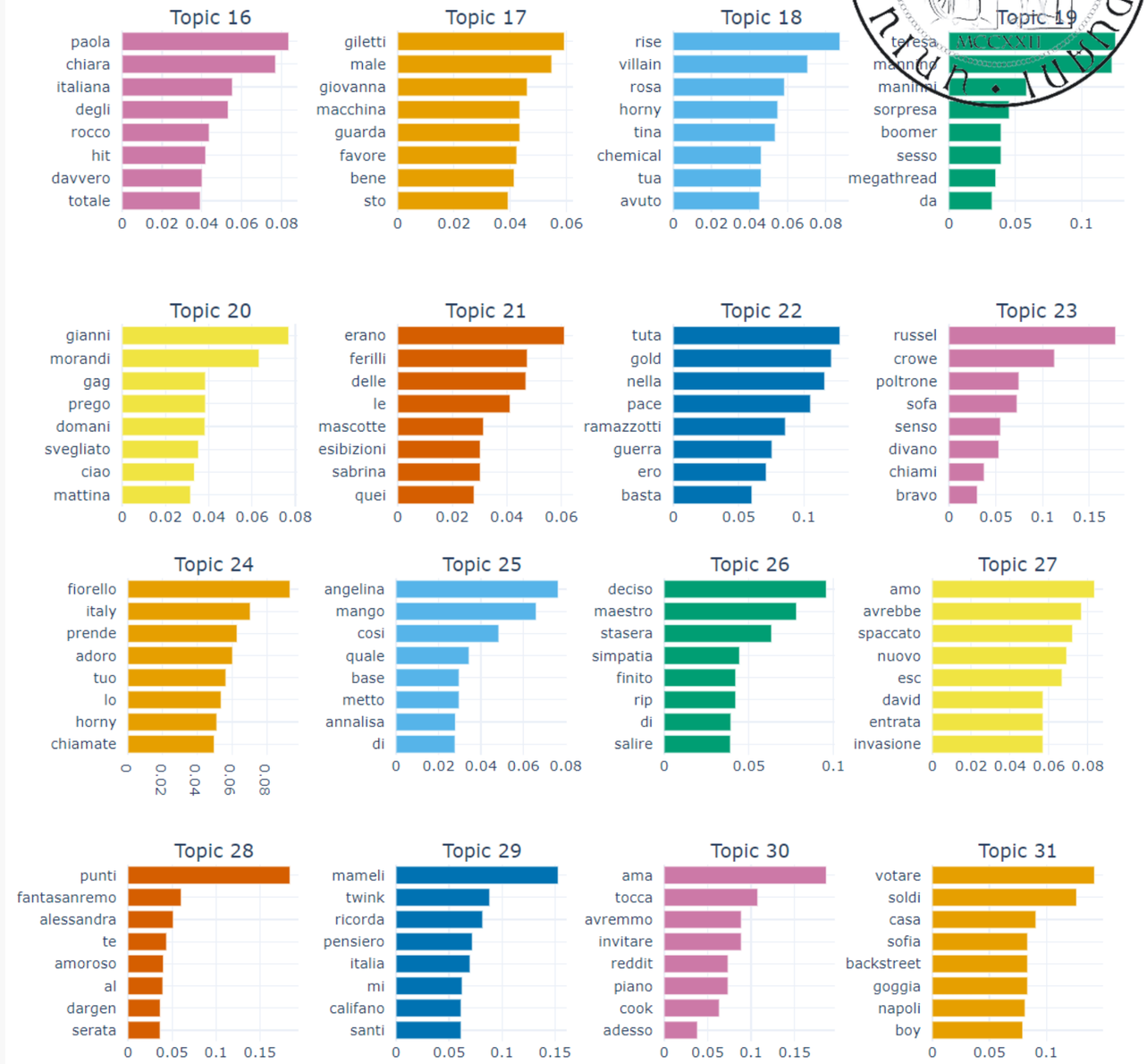
HDBSCAN Mod
8 **communities**
NMI: 0.275352
Q: 0.141757
Ncut: 0.742223
InfoMap: 0.076889



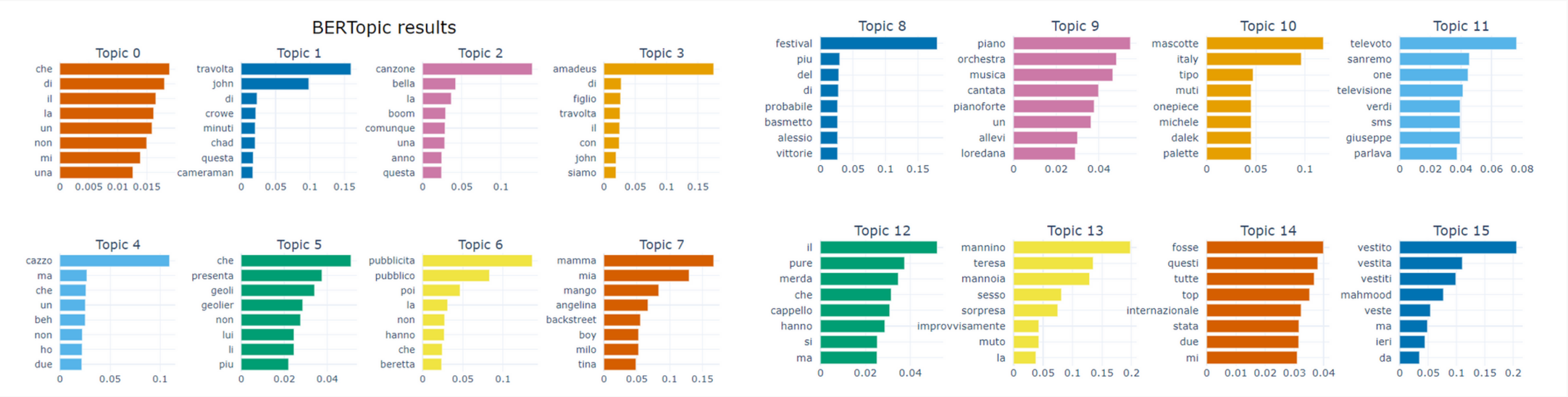
All the community detection algorithms found a few communities with many documents.

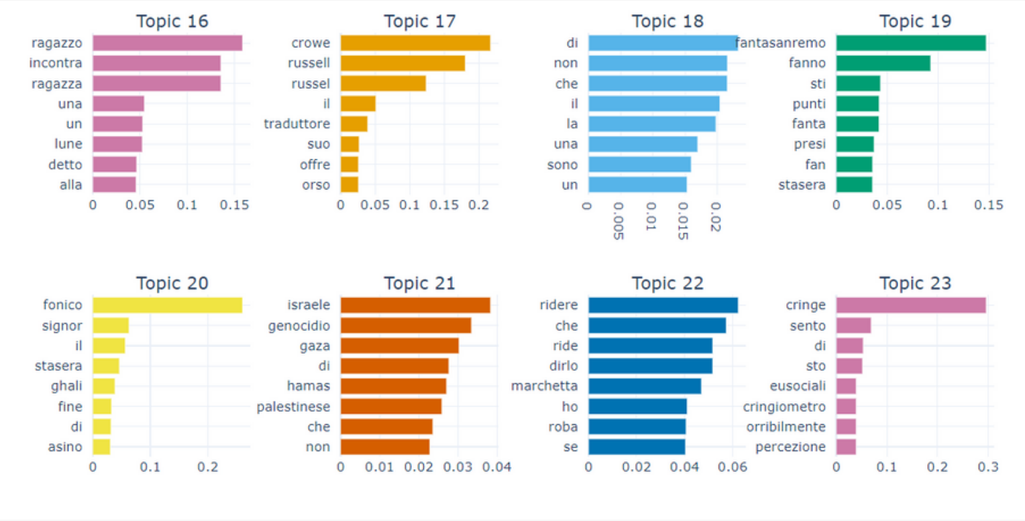
LOUVAIN


Louvain results
Topic 0
travolta
john
qua
ballo
del
sanremo
il
di
Topic 1
che
non
di
la
un
il
sono
con
Topic 2
ghali
sento
fai
di
leo
edoardo
cringe
mi
Topic 3
sta
irama
proprio
tu
cosa
cazzo
andando
stai
Topic 4
loredana
ama
merda
sangiovanni
oddio
mi
lol
costa
Topic 5
diodato
ragazzo
incontra
ragazza
un
una
bambino
signora
Topic 6
so
dal
big
jannacci
mama
momento
stare
live
Topic 7
anno
poveri
quest
ricchi
outfit
sti
scorso
eurovision
Topic 8
canta
pubblico
dopo
messi
boom
voleva
sanremo
canzone
Topic 9
sembra
lei
sei
presenta
ridere
mi
me
fred
Topic 10
rain
mr
fuori
su
mare
monologo
shottino
pubblicita
Topic 11
nek
vestito
renga
giuliano
come
piu
cantare
taylor
Topic 12
mio
padre
dargen
amico
il
la
bestemmia
sinistra
Topic 13
questi
altra
fine
due
annalisa
tre
coco
codice
Topic 14
sempre
brava
beh
viva
vedo
la
penso
carina
Topic 15
voglio
minchia
essere
signor
karma
inquadrature
mahmood
cazzo

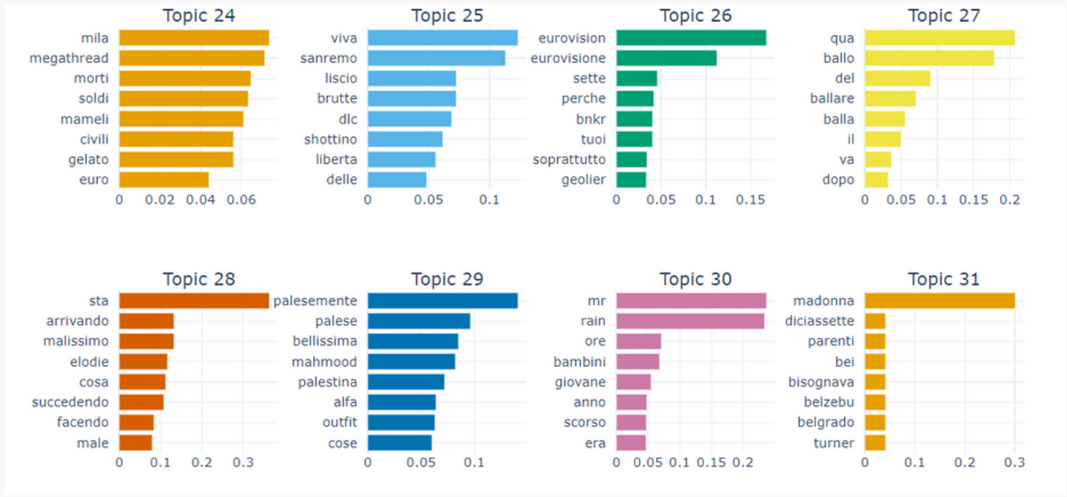

Topic 16
paola
chiara
italiana
degli
rocco
hit
davvero
totale
Topic 17
giletti
male
giovanna
macchina
guarda
favore
bene
sto
Topic 18
rise
villain
rosa
horny
tina
chemical
tua
avuto
Topic 19
teresa
sorpresa
boomer
sesso
megathread
da
Topic 20
gianni
morandi
gag
prego
domani
svegliato
ciao
mattina
Topic 21
erano
ferilli
delle
le
mascotte
esibizioni
sabrina
quei
Topic 22
tuta
gold
nella
pace
ramazzotti
guerra
ero
basta
Topic 23
russel
crowe
poltrone
sofa
senso
divano
chiami
bravo
Topic 24
fiorello
italy
prende
adoro
tuo
lo
horny
chiamate
Topic 25
angelina
mango
cosi
quale
base
metto
annalisa
di
Topic 26
deciso
maestro
stasera
simpatia
finito
rip
di
salire
Topic 27
amo
avrebbe
spaccato
nuovo
esc
david
entrata
invasione
Topic 28
punti
fantasanremo
alessandra
te
amoroso
al
dargen
serata
Topic 29
mameli
twink
ricorda
pensiero
italia
mi
califano
santi
Topic 30
ama
tocca
avremmo
invitare
reddit
piano
cook
adesso
Topic 31
votare
soldi
casa
sofia
backstreet
goggia
napoli
boy

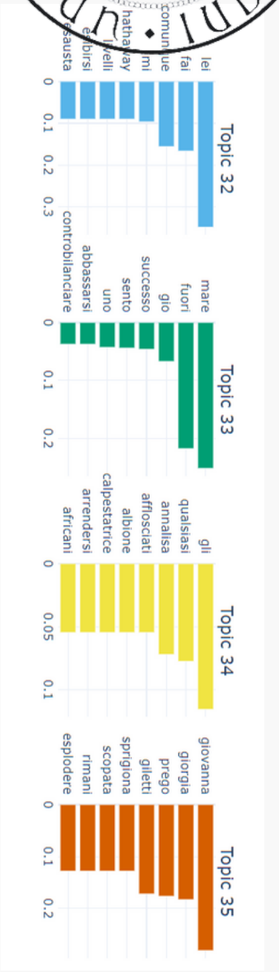
BERTopic


BERTopic results
Topic 0
che
di
il
la
un
non
mi
una
Topic 1
travolta
john
di
crowe
minuti
chad
questa
cameraman
Topic 2
canzone
bella
la
boom
comunque
una
anno
questa
Topic 3
amadeus
di
figlio
travolta
il
con
john
siamo
Topic 4
cazzo
ma
che
un
beh
non
ho
due
Topic 5
che
presenta
geoli
geolier
non
lui
li
piu
Topic 6
pubblicita
pubblico
poi
la
non
hanno
che
beretta
Topic 7
mamma
mia
mango
angelina
backstreet
boy
milo
tina

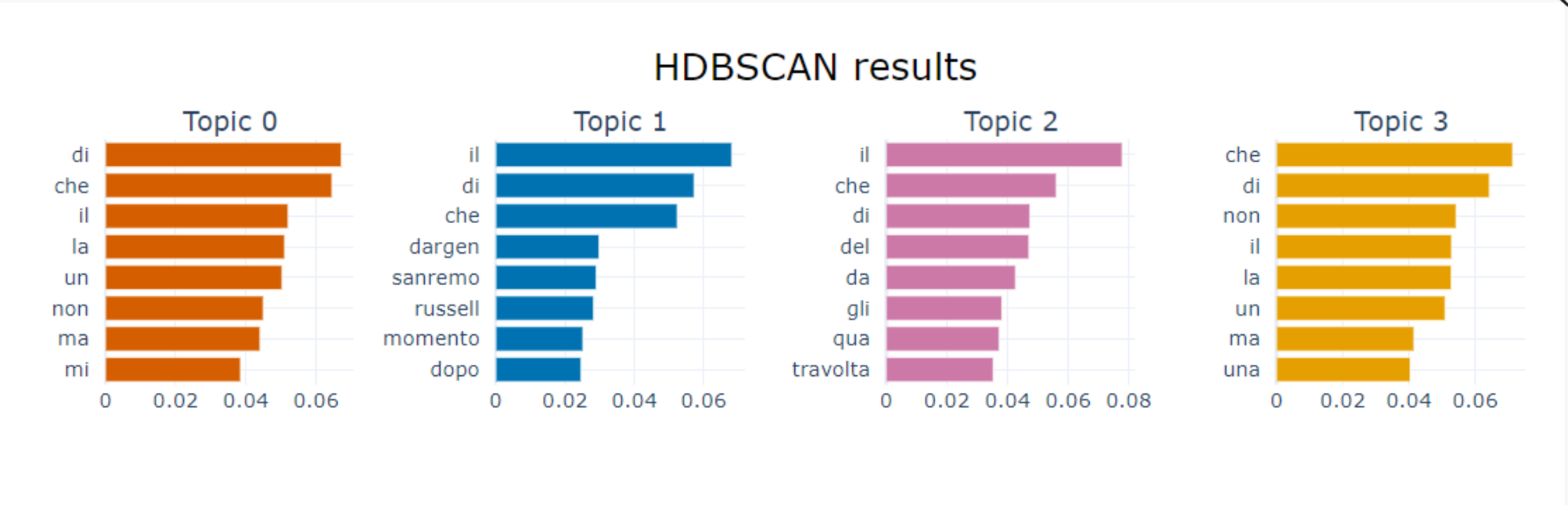

Topic 16
ragazzo
incontra
ragazza
una
un
lune
detto
alla
Topic 17
crowe
russell
russel
il
traduttore
suo
offre
orso
Topic 18
di
non
che
il
la
una
sono
un
Topic 19
fantasanremo
fanno
sti
punti
fanta
presi
fan
stasera
Topic 20
fonico
signor
il
stasera
ghali
fine
di
asino
Topic 21
israele
genocidio
gaza
di
hamas
palestinese
che
non
Topic 22
ridere
che
ride
dirlo
marchetta
ho
roba
se
Topic 23
cringe
sento
di
sto
eusociali
cringiometro
orribilmente
percezione

Topic 8
festival
piu
del
di
probabile
basmetto
alessio
vittorie
Topic 9
piano
orchestra
musica
cantata
pianoforte
un
allevi
loredana
Topic 10
mascotte
italy
tipo
muti
onepiece
michele
dalek
palette
Topic 11
televoto
sanremo
one
televisione
verdi
sms
giuseppe
parlava
Topic 12
il
pure
merda
che
cappello
hanno
si
ma
Topic 13
mannino
teresa
mannoia
sesso
sorpresa
improvvisamente
muto
la
Topic 14
fosse
questi
tutte
top
internazionale
stata
due
mi
Topic 15
vestito
vestita
vestiti
mahmood
veste
ma
ieri
da


Topic 24
mila
megathread
morti
soldi
mameli
civili
gelato
euro
Topic 25
viva
sanremo
liscio
brutte
dlc
shottino
liberta
delle
Topic 26
eurovision
eurovisione
sette
perche
bnkr
tuoi
soprattutto
geolier
Topic 27
qua
ballo
del
ballare
balla
il
va
dopo
Topic 28
sta
arrivando
malissimo
elodie
cosa
succedendo
facendo
male
Topic 29
palesemente
palese
bellissima
mahmood
palestina
alfa
outfit
cose
Topic 30
mr
rain
ore
bambini
giovane
anno
scorso
era
Topic 31
madonna
diciassette
parenti
bei
bisognava
belzebu
belgrado
turner

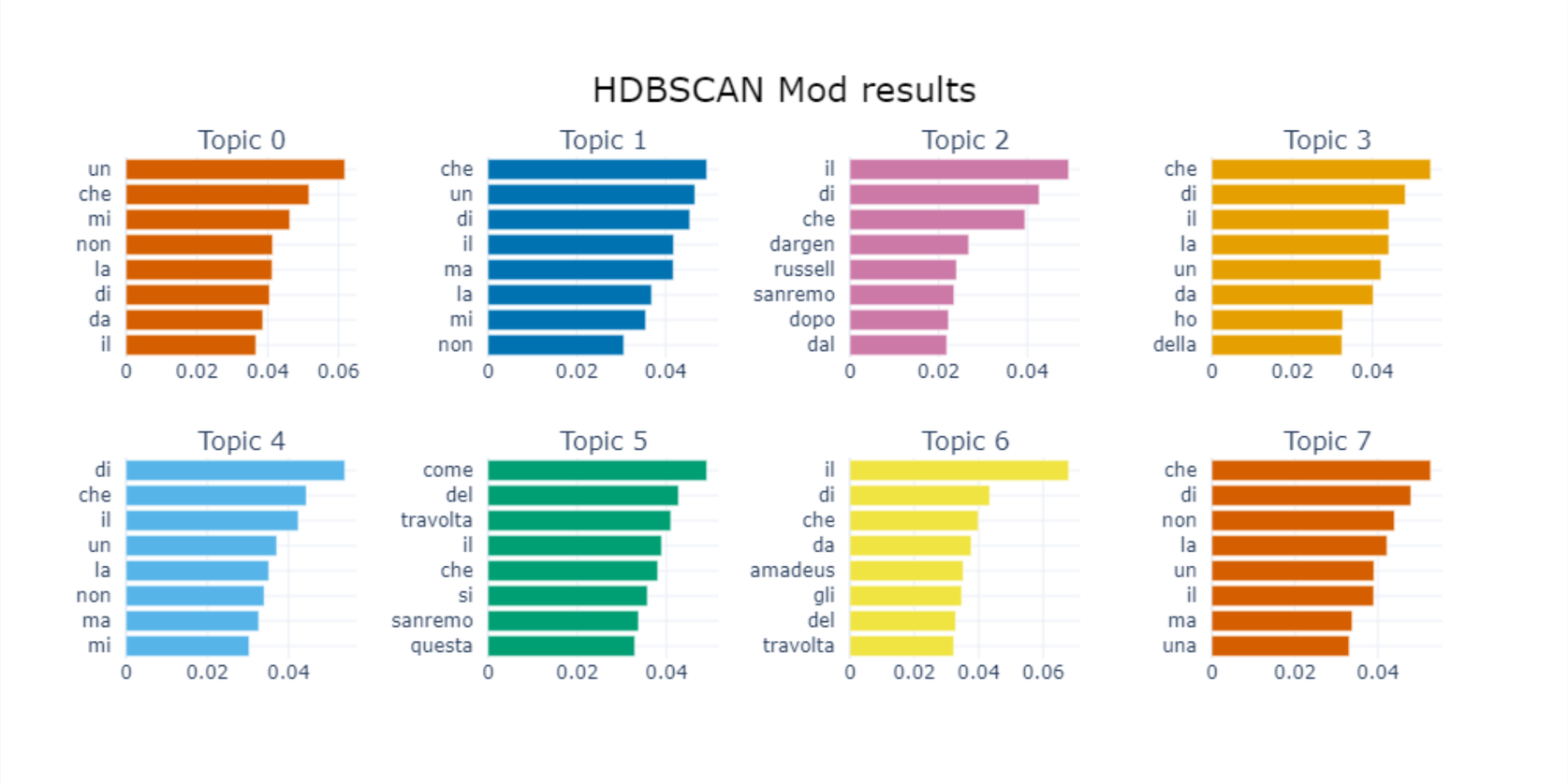

Topic 32
sausta
embrisi
elli
hathaway
mi
comunque
fai
lei
Topic 33
contrabilanciare
abbassarsi
uno
sento
successo
glo
fuori
mare
Topic 34
africani
arrendersi
calpestatrice
albione
affiosciati
amalisa
qualsiasi
gli
Topic 35
esplodere
rimani
scopata
sprigiona
giletti
prego
giorgia
giovanna

HBDSCAN

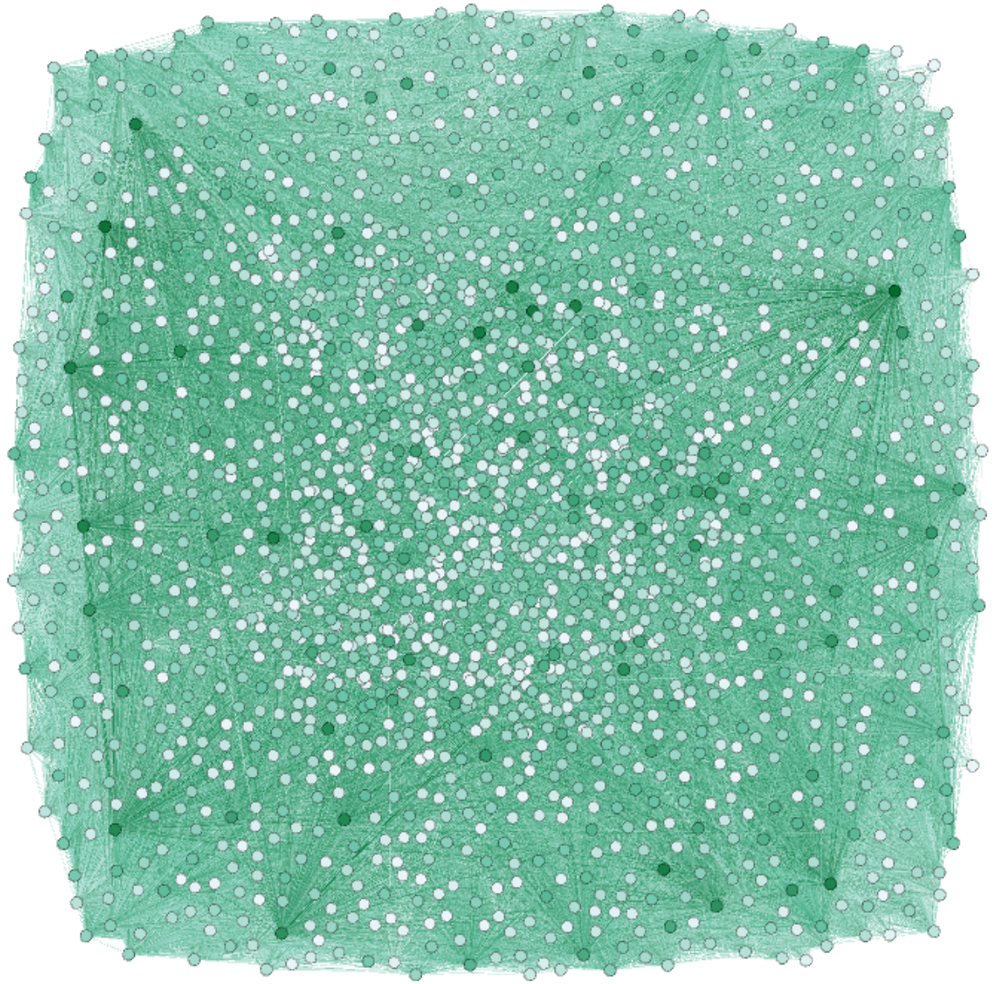


HDBSCAN gave a few topics, but the are not interpretable.

HBDSCAN MOD



HDBSCAN gave a few topics, but the are not interpretable.



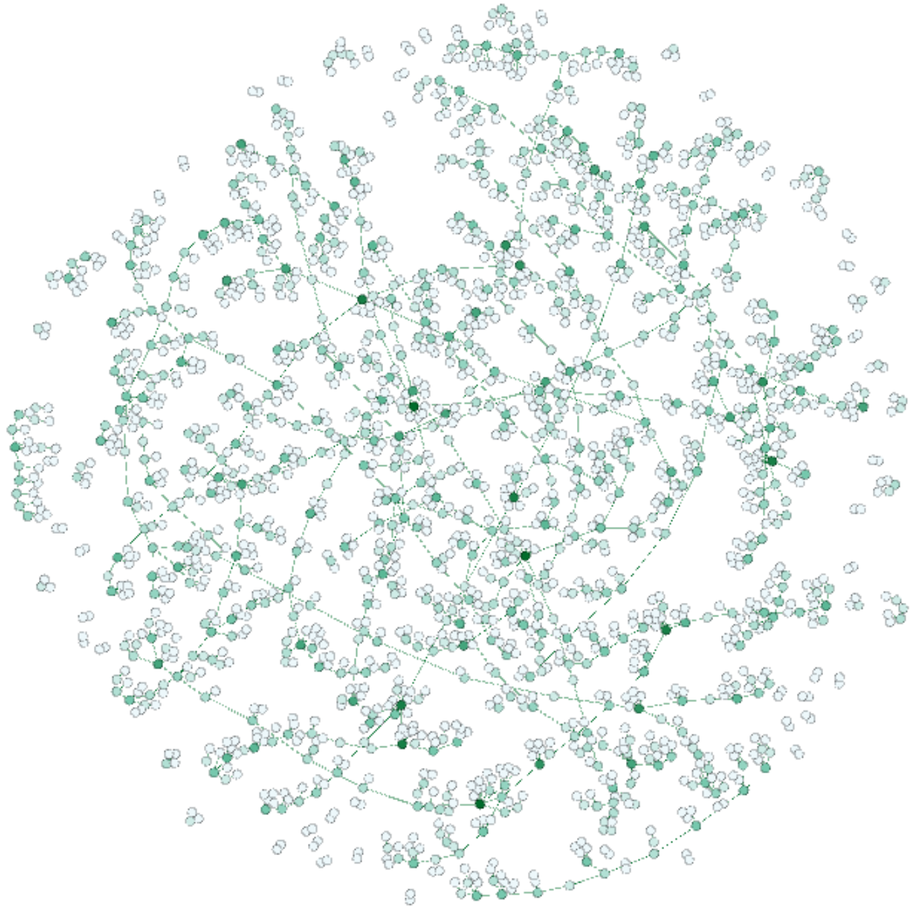
Gephi Visualization

Input:

- For nodes: composed of an excel file of two columns, one is the extracted list of nodes, totalling to 1943 after data cleaning as “id”, and two is the associated comment as “Label”.
- For edges: composed of an excel file of three columns, one is the “Source” node, two is the “Target” node, and three is the “Label”

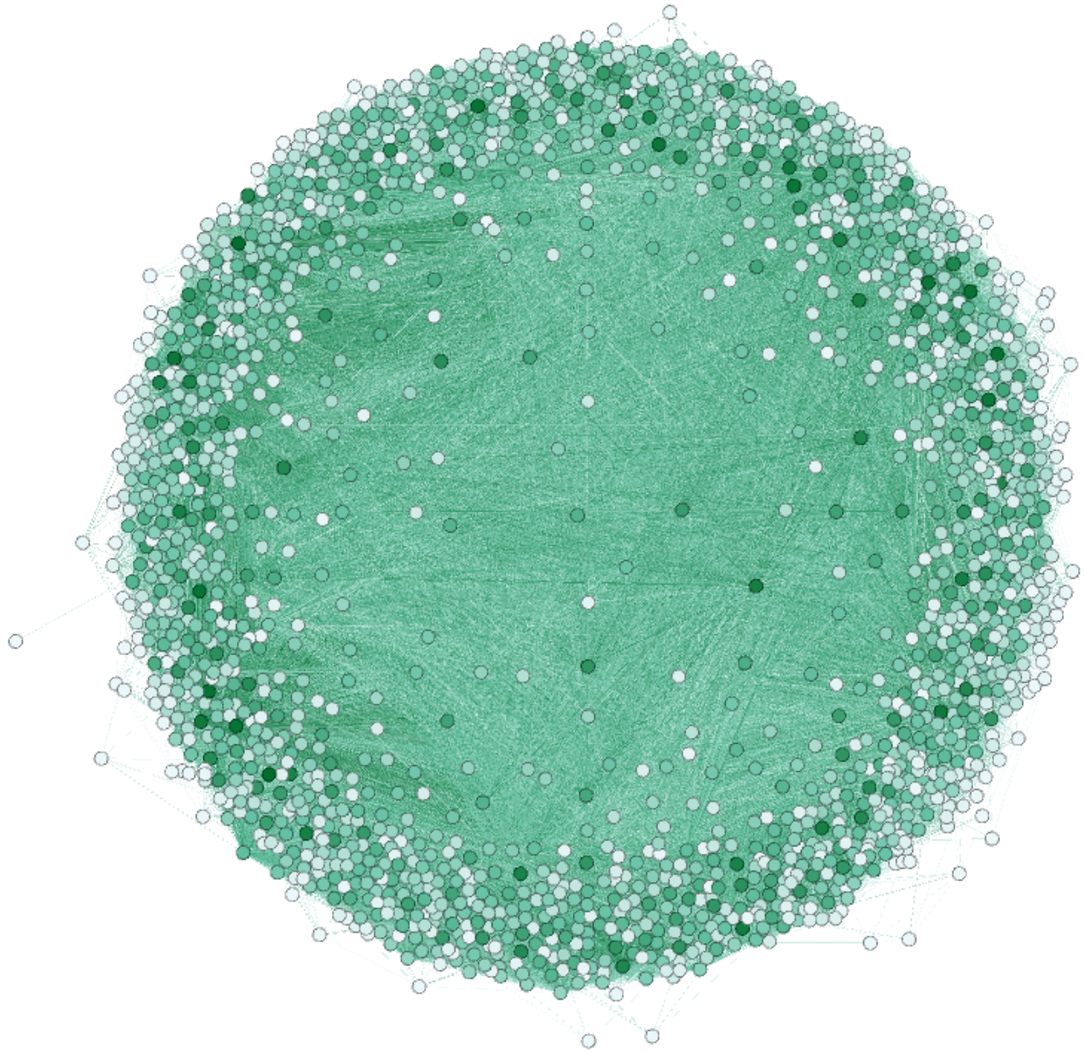
Output:

Graph visualizations of our dense network with some important parameters.



Degree:

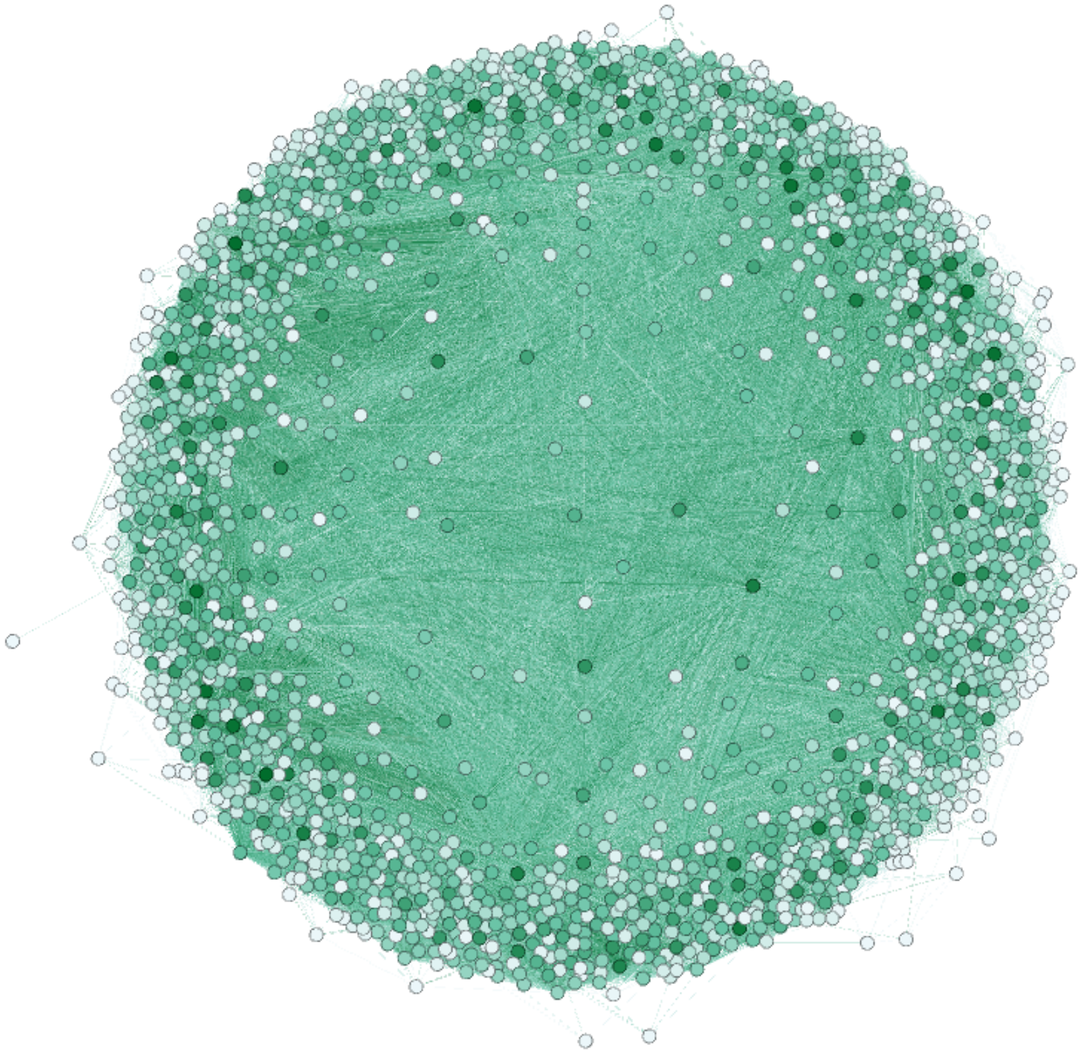
1. Agricoltura italiana in crisi, necessità di trasformazione e modernizzazione.
2. Critica alle esenzioni fiscali per gli agricoltori e richiesta di equità.
3. Problema della violenza nelle proteste e necessità di civiltà nel dissenso.
4. Censura mediatica sulle dichiarazioni politiche di un artista.
5. Racconto umoristico di un incontro con una celebrità al supermercato.
6. Maratona di proteste contro l'inequità fiscale e le esenzioni fiscali.
7. Discussione sulla protesta durante una trasmissione televisiva.
8. Canzone sulla pioggia come metafora dello stile di un artista.
9. Opinioni contrastanti sulle questioni agricole, inclusa la necessità di cambiamento.
10. Riflessioni sulla contestualità dei discorsi pubblici e l'importanza del rispetto nei confronti delle istituzioni.



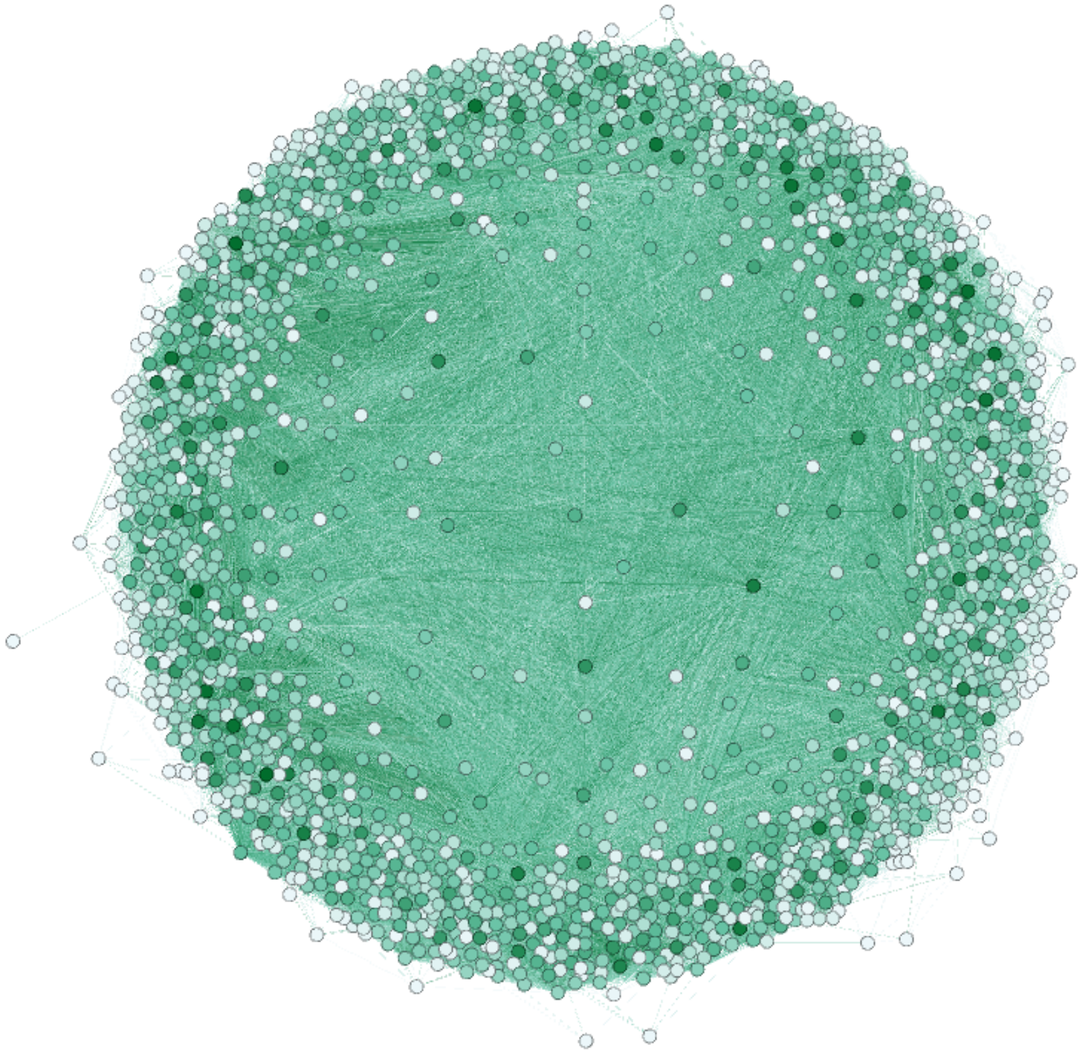
PageRank:

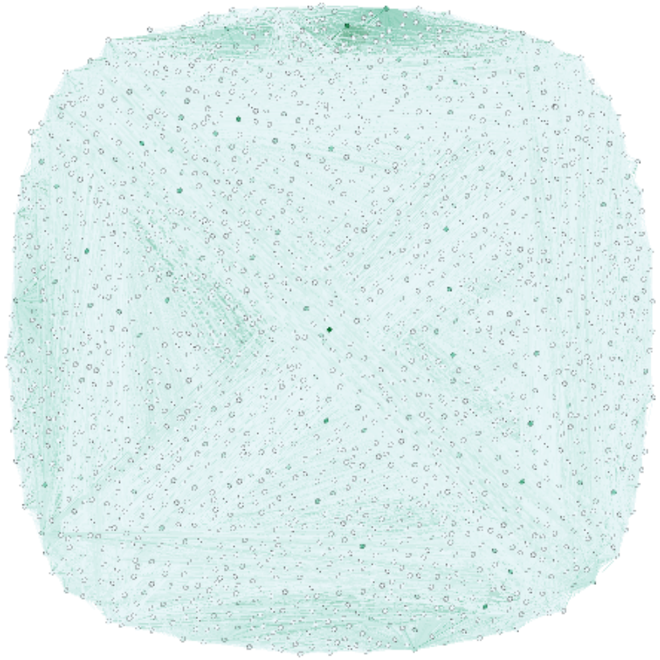
1. L'agricoltura necessita di trasformazioni a livello europeo per efficienza e modernizzazione.
2. Le aziende agricole italiane di medie dimensioni faticano contro concorrenti più grandi a causa di pratiche obsolete.
3. Il settore beneficia di esenzioni fiscali, ma ora affronta richieste di pagamento di imposte.
4. I manifestanti esprimono insoddisfazione, ma alcuni ricorrono a mezzi violenti.
5. La copertura mediatica di argomenti sensibili come il conflitto Israele-Palestina può suscitare controversie.
6. L'intervento di Mara Venier durante uno spettacolo televisivo ha scatenato il dibattito sulla censura.
7. Le azioni della polizia durante le proteste coinvolgono spesso misure di controllo della folla.
8. La canzone di Mr. Rain cattura la sua personalità e il suo stile, risuonando con il pubblico.
9. La legalità di certe affermazioni e contesti, come quelle fatte durante Sanremo, è messa in discussione.
10. Aneddoti personali e incontri, come uno con un famoso cantante, aggiungono un tocco personale.

Authorities



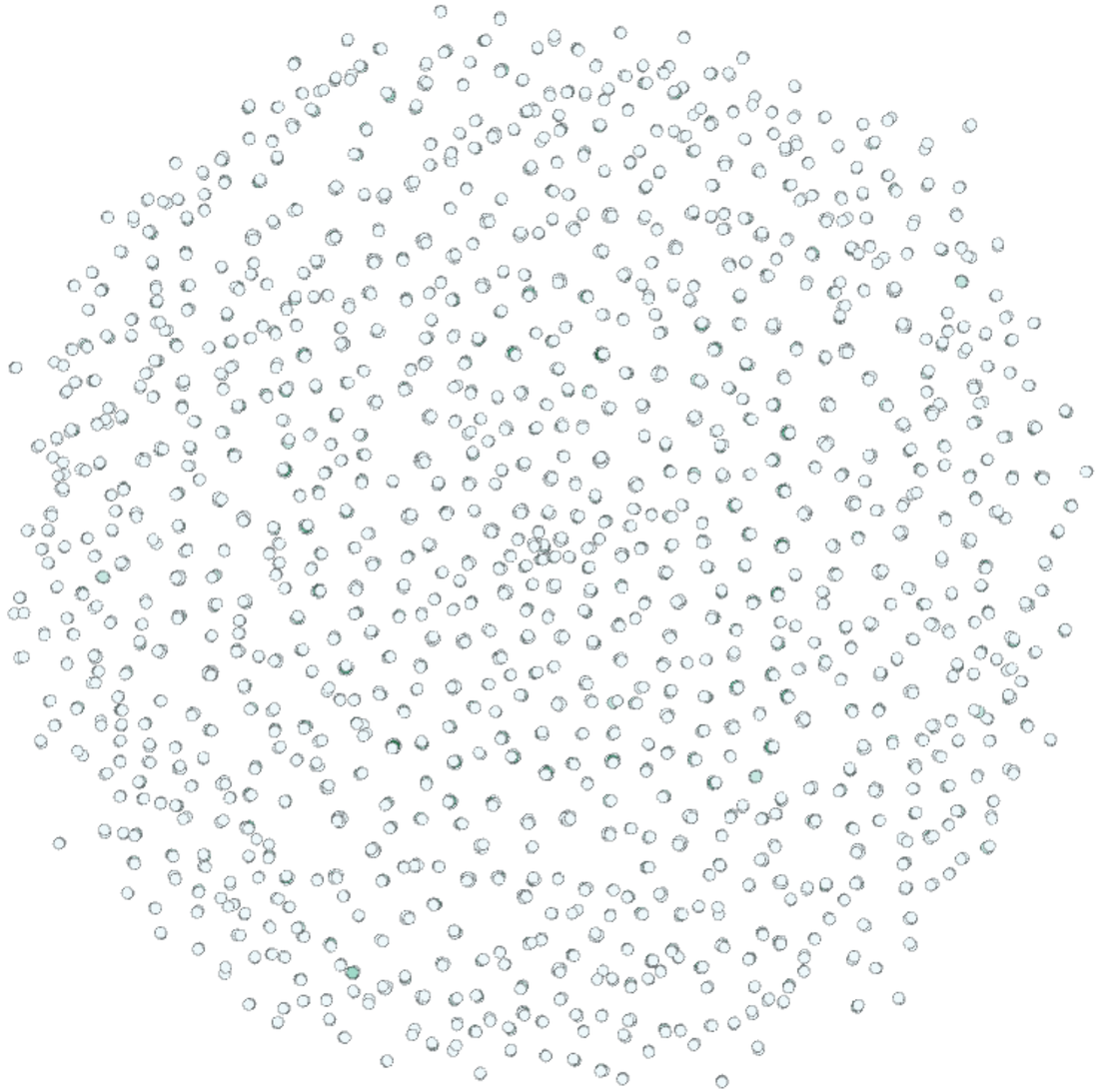
Hubs





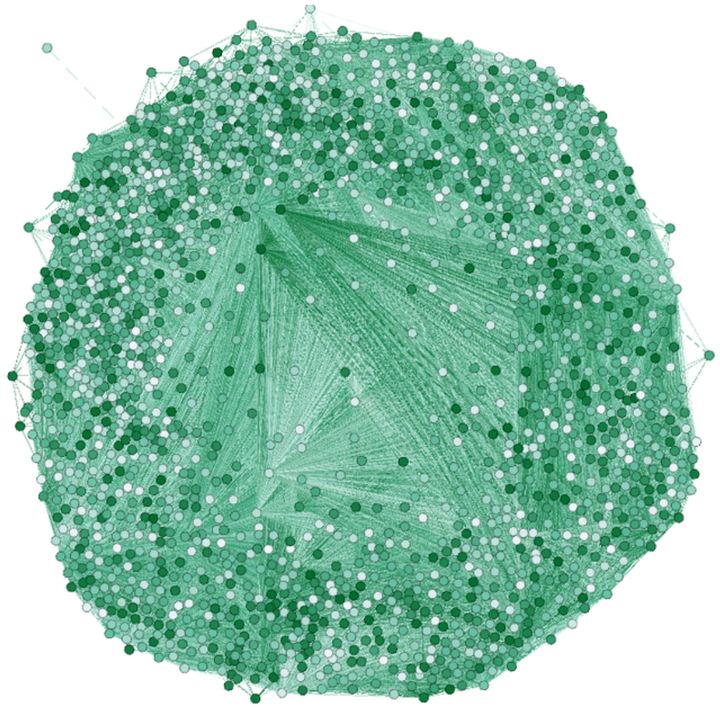
Betweenness:

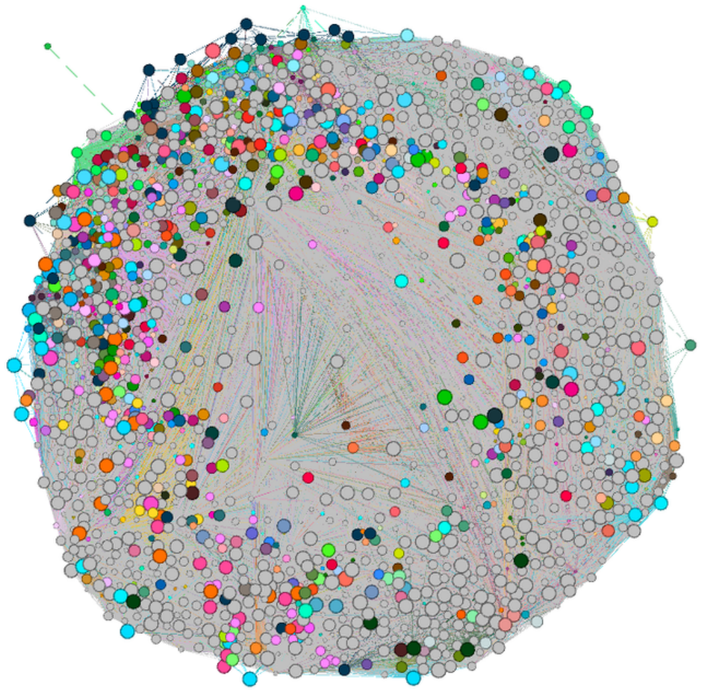
1. L'agricoltura europea necessita di modernizzazione.
2. Gli agricoltori italiani sono insoddisfatti delle tasse.
3. Vi è controversia sulle proteste agricole e l'operato della polizia.
4. Alcuni sostengono che la TV pubblica debba seguire la linea politica dello stato.
5. Si discute la libertà di espressione a Sanremo.
6. C'è disaccordo sulla definizione di genocidio nei conflitti geopolitici.
7. Reazioni varie a controversie a Sanremo.
8. Si discute la censura in contesti politici.
9. Dubbi sull'operato della polizia durante le proteste.
10. Accesso alle piattaforme di comunicazione e democratizzazione della voce pubblica sono temi di discussione.



Closeness:

1. L'agricoltura deve modernizzarsi a livello europeo.
2. È necessario cambiare la mentalità sull'agricoltura.
3. I contadini italiani protestano per tasse e restrizioni.
4. La disputa sul bilancio europeo riguarda l'esenzione fiscale.
5. Critiche sulla censura televisiva sono emerse.
6. La libertà di espressione a Sanremo è dibattuta.
7. Si discute l'uso del termine "genocidio" nei conflitti.
8. La risposta della polizia alle proteste agricole è stata valutata.
9. L'accesso alle piattaforme di comunicazione è una questione rilevante.
10. Le proteste agricole evidenziano la necessità di riforme.





Statistical Inference:

1 DEFINITION

Statistical inference of assortative community structures, typically operates by comparing the observed network with a null model to assess whether observed patterns or statistics in the network are statistically significant. It can be used to test hypothesises too.

2 PARAMETERS

Null model, number of trials, significance levels, and edge weight distributions.

3 PROCESS

Implemented in graph statistics in Gephi. After running several times on the same network, it was able to deduce between *844-883* communities of varying size distribution.

CONCLUSION

After observing the comments having highest centrality measures we can deduce that the most chosen comments are mostly related to the ongoing political climate, guests, and least regarding the songs of the festival themselves.

THANK YOU!

ANY QUESTIONS?