



Corso Intelligent System

ESONERO SOCIAL NETWORK ANALYSIS GEPHI

Luca Spaccini | Mat. 327488
Aniello Coscione | Mat. 327371

Università degli Studi di Perugia
Dipartimento Matematica e Informatica

Anno Accademico 2019/2020



Sommario

INTRODUZIONE.....	3
DATI GENERALI GRAFO	3
COMPOSIZIONE dei dataset	3
ANALISI PRIMO DATASET	4
Dati generali primo grafo:.....	4
<i>Grafo:</i>	4
<i>Minimo grado Nodi:</i>	4
<i>Massimo grado Nodi:</i>	4
<i>Media grado Nodi:</i>	4
<i>Diametro:</i>	5
<i>PageRank Massimo e Media:</i>	5
<i>Coefficiente di Clustering:</i>	6
<i>Numero Componenti connesse:</i>	6
LE COMUNITA' PRINCIPALI DEL PRIMO DATASET	8
Le 4 comunità principali	8
Le 2 comunità principali	9
ULTERIORI METRICHE DI ANALISI DEL PRIMO DATASET	10
GRADO.....	10
EIGENVECTOR CENTRALITY	10
BETWEENNESS CENTRALITY.....	11
CLOSENESS CENTRALITY.....	11
ANALISI SECONDO DATASET	12
Dati generali secondo grafo:	12
<i>Grafo:</i>	12
<i>Minimo grado Nodo:</i>	12
<i>Massimo grado Nodo:</i>	12
<i>Media grado Nodi:</i>	12
<i>Diametro:</i>	13
<i>PageRank Massimo e Media:</i>	13
<i>Coefficiente di Clustering:</i>	14
<i>Numero Componenti connesse:</i>	14
LE COMUNITA' PRINCIPALI DEL SECONDO DATASET	16
Le 4 comunità principali	16
Le 2 comunità principali	17
ULTERIORI METRICHE DI ANALISI DEL SECONDO DATASET	18

GRADO	18
EIGENVECTOR CENTRALITY	18
BETWEENNESS CENTRALITY.....	19
CLOSENESS CENTRALITY.....	19
ANALISI DATASET COMPLETO.....	20
Dati generali grafo completo:	20
<i>Grafo:</i>	20
<i>Minimo grado Nodi:</i>	20
<i>Massimo grado Nodi:</i>	20
<i>Media grado Nodi:</i>	20
<i>Diametro:</i>	21
<i>PageRank Massimo e Media:</i>	21
<i>Coefficiente di Clustering:</i>	22
<i>Numero Componenti connesse:</i>	22
LE COMUNITÀ PRINCIPALI DEL DATASET COMPLETO	24
Le 4 comunità principali	24
Le 2 comunità principali	25
ULTERIORI METRICHE DI ANALISI DEL DATASET COMPLETO	26
GRADO	26
EIGENVECTOR CENTRALITY	26
BETWEENNESS CENTRALITY.....	27
CLOSENESS CENTRALITY	27
TOP 25% NODI.....	28
CONCLUSIONI.....	29

INTRODUZIONE

La relazione seguente è relativa allo studio effettuato sui titoli del giornale “LaRepubblica”. L'obiettivo principale è stato quello di trovare, attraverso la pagina web apposita, tutti i titoli del giornale relativi alla parola “mascherine”. Lo studio è stato svolto utilizzando strumenti quali: Google Colab per creare i due file con estensione CSV contenenti uno l'insieme di tutti i nodi del grafo e l'altro l'insieme di tutti gli archi, Google Drive per il passaggio di dati, Gephi per manipolare il grafico estrapolando tutti i dati necessari, Python utilizzato come linguaggio principale, sia per l'estrazione dei titoli dalla pagina web, sia per la creazione dei CSV. Molto brevemente, la fase di pulizia dei titoli è stata composta dall'eliminazione di stopwords, stemming e come da richiesta due parole con iniziale maiuscola sono state unite tra di loro previ controlli espressi per la composizione “NomeMaiuscolo”. Dopo questa fase si è passati ad elaborare i titoli per generare, prima un CSV di tutte le parole univoche per rappresentare così i nodi, poi un CSV di tutti gli archi univoci con il relativo peso. Come da richiesta il peso è stato calcolato sommando il valore 1 ogni volta che si trovava un arco con la stessa composizione di uno già presente nella lista degli archi pesati. Successivamente si è passati all'importazione dei due file all'interno dell'applicativo Gephi, che ci ha permesso di rispondere alle domande richieste dallo studio. Di seguito vedremo tutti i risultati delle analisi effettuate.

DATI GENERALI GRAFO

L'analisi del nostro grafo si basa sulla parola “mascherine”:

- Periodo del primo dataset: Periodo dal 07/01/2020 al 07/02/2020
- Periodo del secondo dataset: Periodo dal 15/03/2020 al 14/04/2020
- Periodo del dataset completo: Periodo dal 07/01/2020 al 07/02/2020 - dal 15/03/2020 al 14/04/2020

I grafi sono composti da archi non orientati e pesati.

COMPOSIZIONE DEI DATASET

In questa parte troviamo il numero di nodi e archi dei tre grafi:

Grafo	Num. Nodi:	Num. Archi:
Grafo DataSet 1	428	2126
Grafo DataSet 2	2617	18753
Grafo Dataset completo	2762	20622

ANALISI PRIMO DATASET

DATI GENERALI PRIMO GRAFO:

Grafo:

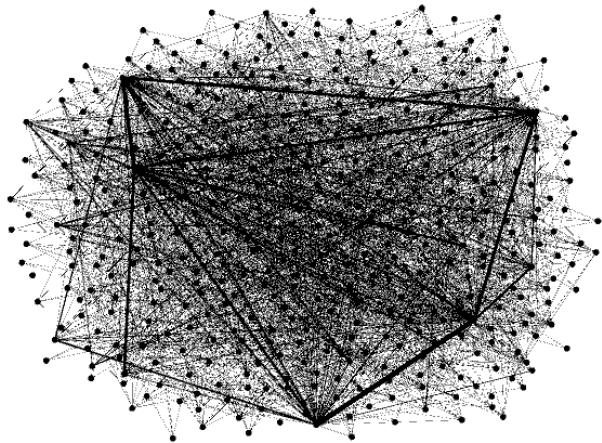


Figura 1: Prima visualizzazione del grafo relativo al dataset 1

Minimo grado Nodi:

Il nodo con il peso minore è la parola “Vietnam” poiché questo nodo non ha collegamenti con altre parole all’interno del nostro dataset.

Label:	Degree:	Weighted degree:
Vietnam	0	0.0

Massimo grado Nodi:

Il nodo con il peso maggiore è la parola “Coronavirus” poiché è una parola molto frequente nei titoli.

Label:	Degree:	Weighted degree:
Coronavirus	164	214.0

Media grado Nodi:

Average Degree	Avg. Weighted Degree
9,935	10,846

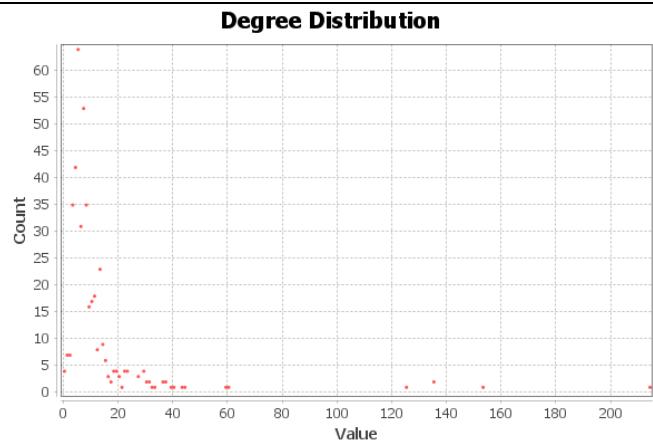
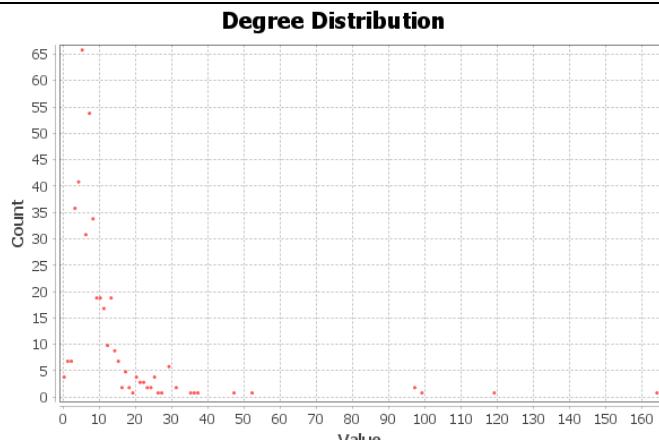
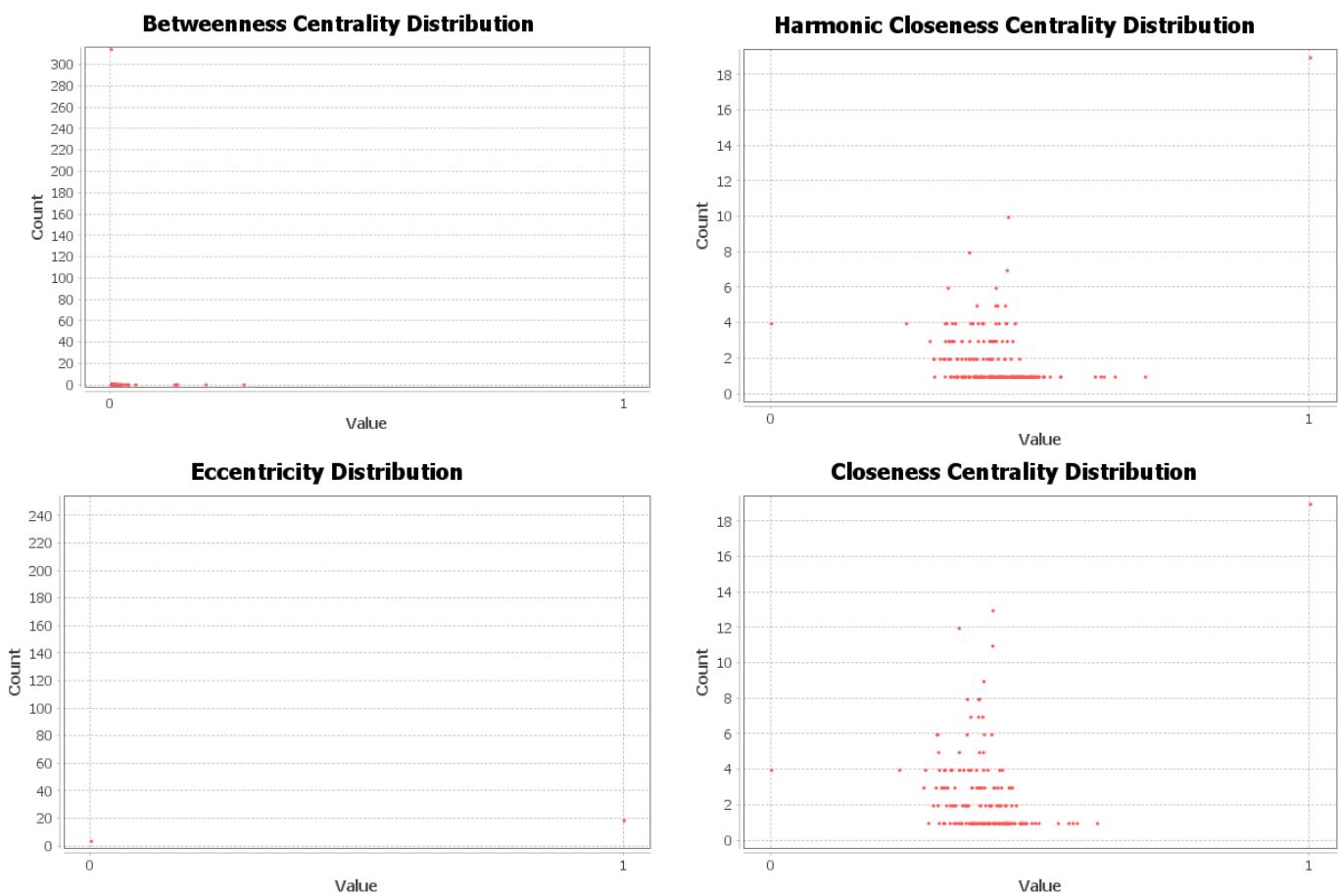


Figura 2: Average Degree e Avg. Weighted Degree

Diametro:

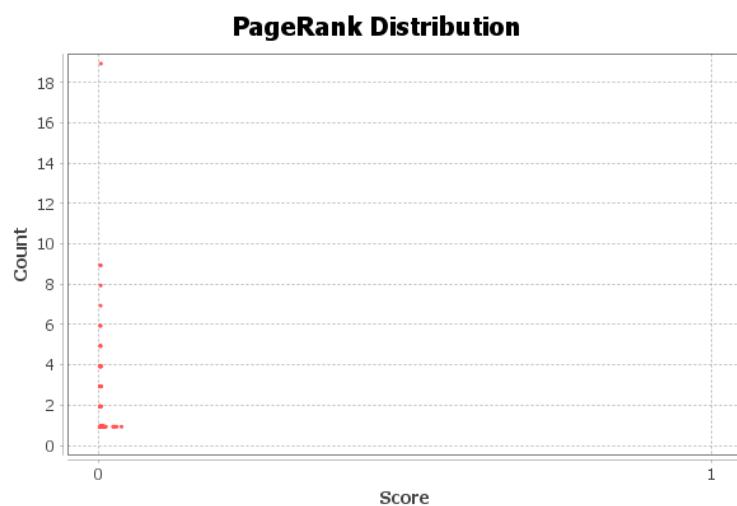
Calcolando il diametro, con l'opzione Normalize Centralites [0;1]), il risultato è 5.



PageRank Massimo e Media:

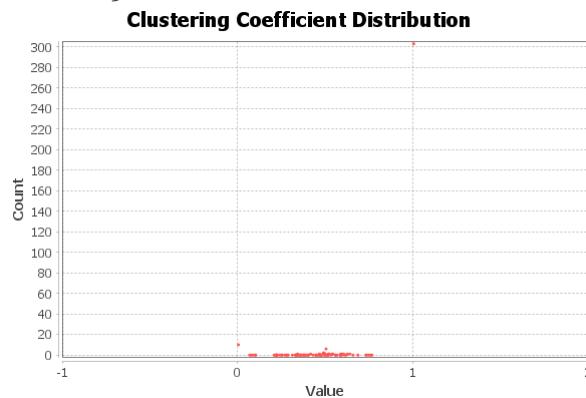
Page Rank indica il modo in cui a un nodo della rete può essere assegnato un "punteggio di importanza", in breve è un "voto" su quanto sia importante un nodo. I valori di PageRank sono quindi i seguenti:

Max. PageRank	Avg. PageRank
0,035933	0,0023364



Coefficiente di Clustering:

Il coefficiente di Clustering è la misura del grado in cui i nodi di un grafo tendono ad essere connessi fra loro.
Il nostro grafo ha il seguente valore: "0,851".



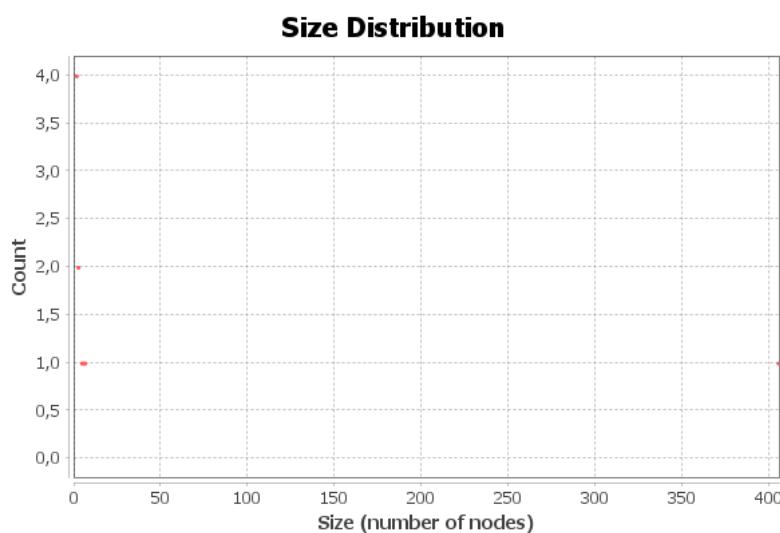
Numero Componenti connesse:

Il numero delle componenti connesse del nostro grafo è: 10



Figura 3: Colori e Percentuali relative alle componenti connesse del grafo

Come vediamo nella Figura 3, abbiamo una componente连通的 che include il 94,63% dei nodi. Le restanti componenti connesse hanno una percentuale molto bassa di inclusione dei nodi quindi sono anche poco significative ai fini dello studio.



Screen componenti connesse grafo:

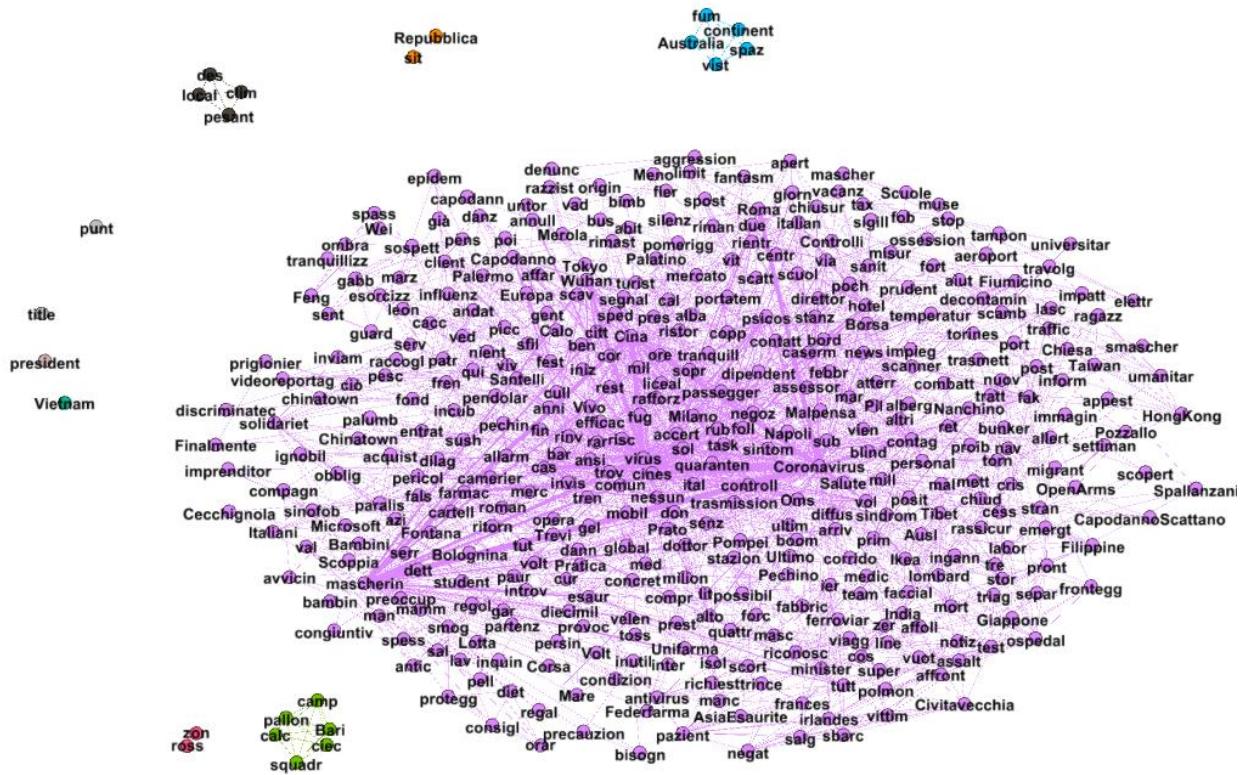


Figura 4: Componenti connesse con label

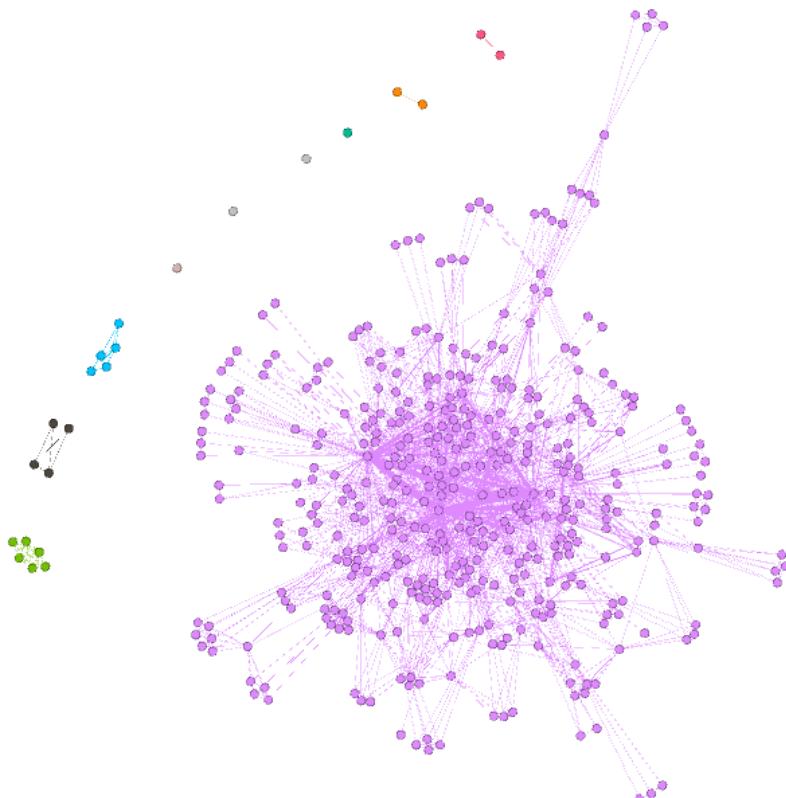


Figura 5: Componenti connesse

LE COMUNITA' PRINCIPALI DEL PRIMO DATASET

LE 4 COMUNITÀ PRINCIPALI

Per trovare le quattro comunità principali abbiamo usato una modularità con risoluzione “1,5”. Il totale dei nodi ricoperti da tutte le comunità è dell’82,01%.

E abbiamo ottenuto il seguente risultato:

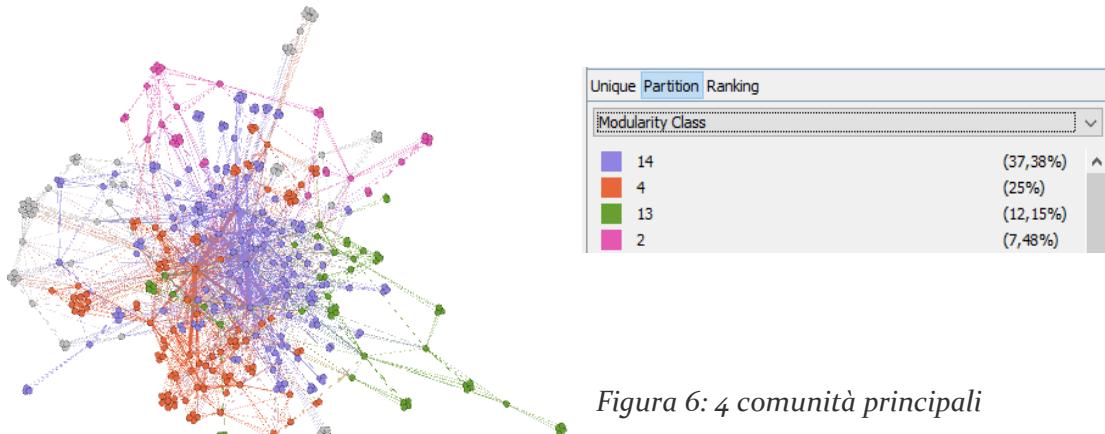


Figura 6: 4 comunità principali

Nel nostro caso non c'è una comunità che include il 50% dei nodi. Come vediamo dalla figura abbiamo una comunità (14) che include il 37,38% dei nodi.

In figura (7) vediamo la comunità più grande delle 4 selezionate che contiene più del 30% dei nodi del nostro grafo.

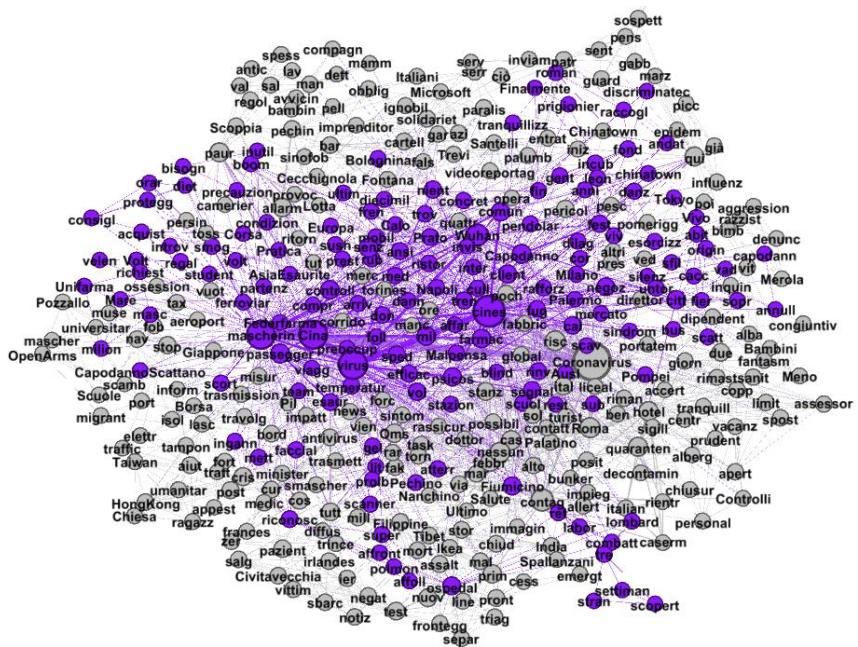


Figura 7: comunità più grande delle 4 principali

Poiché essa contiene parole rilevanti come “mascherine”, “Cina”, “virus” e “Wuhan”. Tali nodi ricoprono le parole che sono state di uso frequente durante il periodo analizzato, confermando così la bontà dello studio fatto.

LE 2 COMUNITÀ PRINCIPALI

Per trovare le due comunità principali abbiamo usato una modularità con risoluzione “2.o”. Il totale dei nodi ricoperti da tutte le comunità è dell’73,06%.

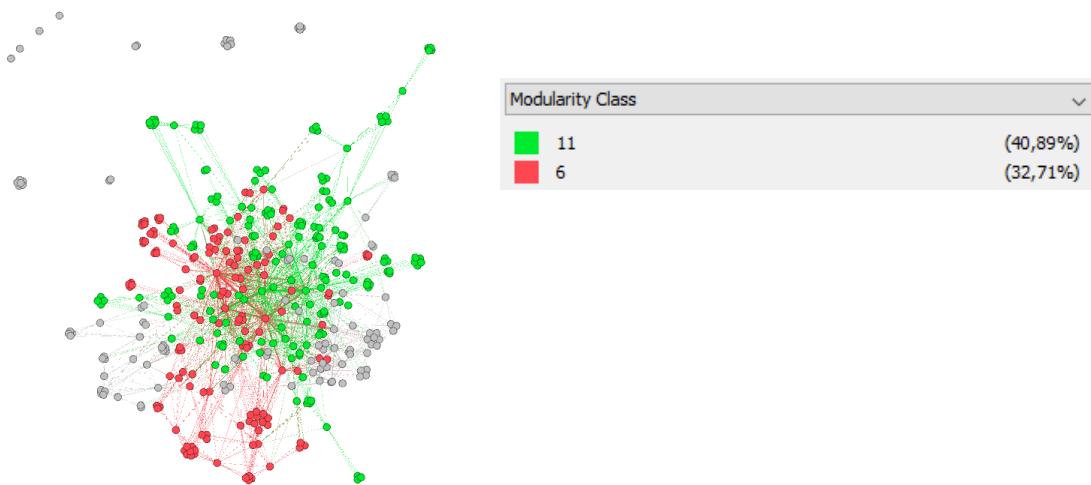


Figura 8: 2 principali comunità

Come possiamo vedere in Figura 8 il grafo può essere suddiviso in due comunità che ricoprono rispettivamente 40,89% e 32,71%. Anche in questo caso non abbiamo una comunità che include almeno il 50% dei nodi. Invece entrambe superano il 30% di nodi inclusi. Questo risultato è visibile in modo più chiaro attraverso la seguente figura.

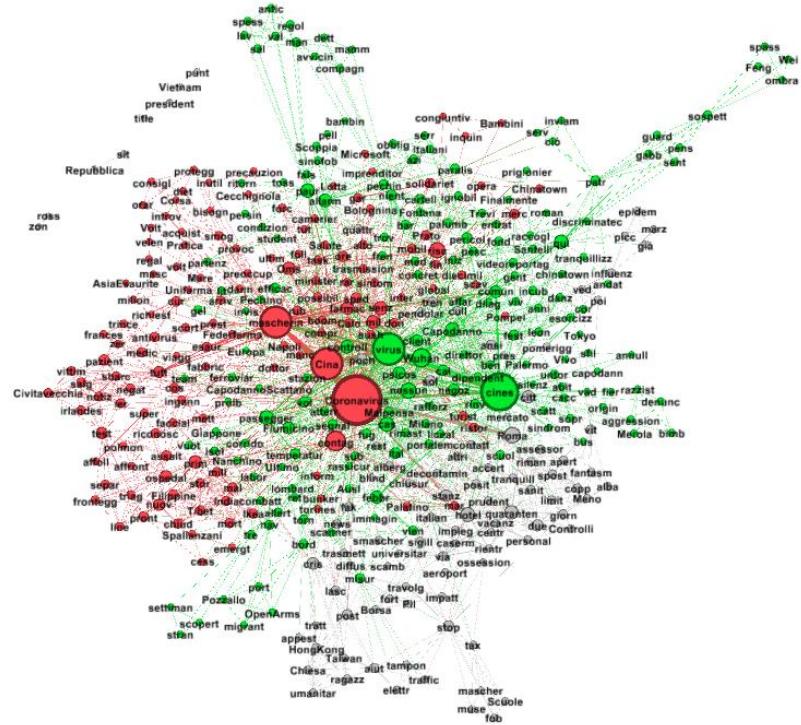


Figura 9: 2 principali comunità con label

Come possiamo vedere sulla figura 9, le due principali comunità sono composte da parole rilevanti che sono state presenti in molti titoli del giornale “LaRepubblica”. In particolare la comunità 11 contiene nodi che rappresentano le parole “mascherine”, “Cina”, “Coronavirus” e “contagi”; questi nodi infatti hanno un alto Degree. Invece la comunità 6 contiene nodi che rappresentano le parole “virus”, “Wuhan”, “cines*” e “Milano”.

ULTERIORI METRICHE DI ANALISI DEL PRIMO DATASET

GRADO

Il grado di un nodo, ha un'importanza nello studio del grafo poiché è uguale al numero di archi che incidono sul nodo stesso.

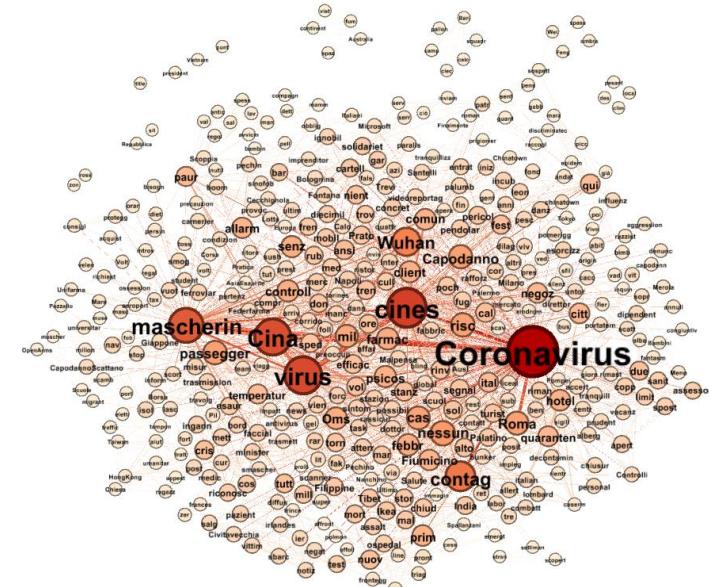


Figura 10: Ranking con il grado

Nella figura seguente si può vedere la grandezza delle label e dei nodi proporzionale alla dimensione del grado di ogni nodo. Infatti il nodo “Coronavirus”, che è quello con grado massimo, ha dimensione del testo e del nodo più grande di tutti. Inoltre è stata aggiunta una colorazione a scala per far risaltare ancora di più questa caratteristica. Perciò a colori più chiari corrispondono gradi bassi e viceversa.

EIGENVECTOR CENTRALITY

Eigenvector Centrality misura l'importanza e l'influenza di un nodo in una rete infatti assegna alti punteggi ai nodi collegati a quelli con punteggi elevati. Indica le tipologie di relazione esistenti e l'importanza tra un singolo nodo e i circostanti. Misura la reputazione di un nodo e il riconoscimento della stessa che gli viene dato dagli altri nodi. Quindi la figura seguente indica che sempre “Coronavirus” è il nodo più importante.

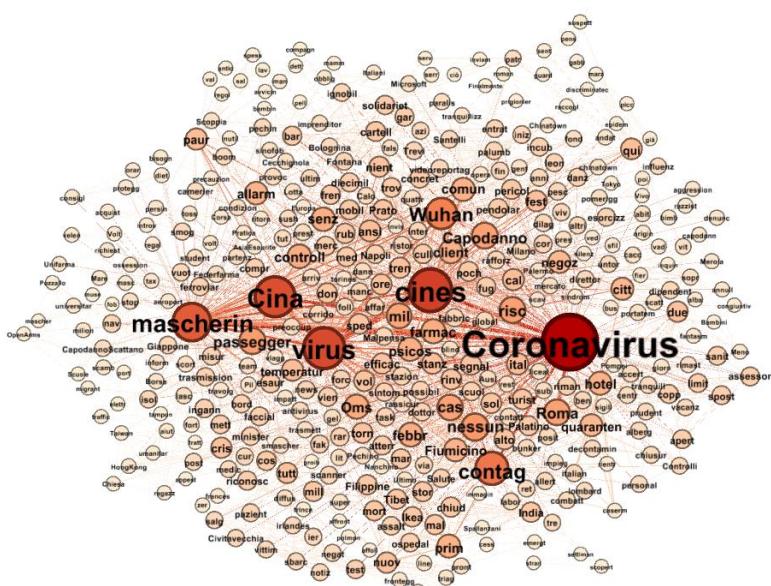


Figura 11: Eigenvector Centrality

BETWEENNESS CENTRALITY

La Betweenness centrality è tra coppie di nodi: è legato al numero di volte in cui un nodo si ritrova lungo lo shortest path tra le altre coppie di nodi della rete (descrive la capacità di diffusione delle informazioni nella rete attraverso quell'utente). Misura la strategicità di un nodo nella rete tra due aree importanti della stessa (concetto di betweenness). Un nodo con una elevata betweenness centrality ha una grande influenza nel flusso di informazioni. I concetti collegati a tale metrica sono la reciprocità (La misura in cui un nodo si trova tra gli altri nodi della rete) e bridge (quanti elementi causerebbero interruzioni ai flussi).

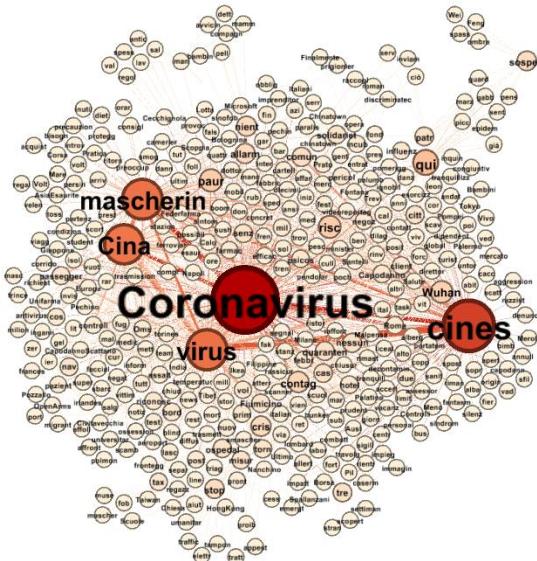


Figura 12 Betweenness Centrality

Come da definizione i nodi più grandi, più scuri e con testo più grande corrispondono a parole che hanno avuto una grande influenza nel flusso delle informazioni. Questa è una prova che durante il periodo in esame principalmente i titoli erano composti da tali parole.

CLOSENESS CENTRALITY

La Closeness Centrality o vicinanza è legato alla distanza media geodesica – soluzione al problema dello shortest path – tra un nodo e tutti gli altri da lui raggiungibili (esprime il grado di prossimità di un nodo agli altri della rete). Misura il numero di connessioni di un membro. Esprime la vicinanza di un nodo alle componenti interconnesse rimanenti della rete. Attraverso questa metrica sono stati, per la prima volta, evidenziati nodi di componenti connesse estranee ai nodi che prima erano i principali. La motivazione è proprio data dal fatto che questi nodi sono molto interconnessi tra di loro.

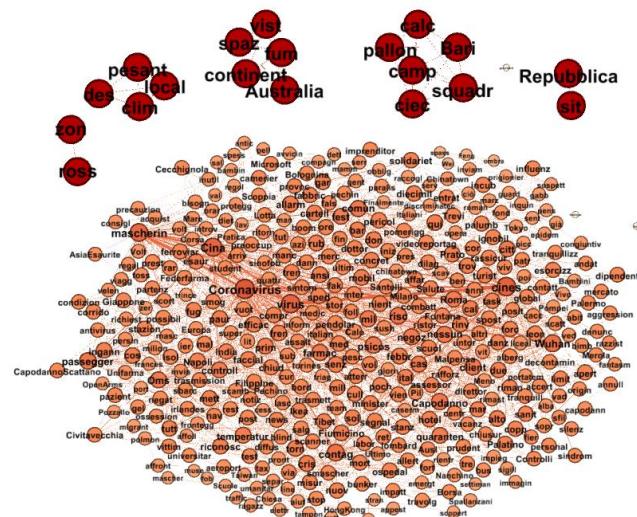
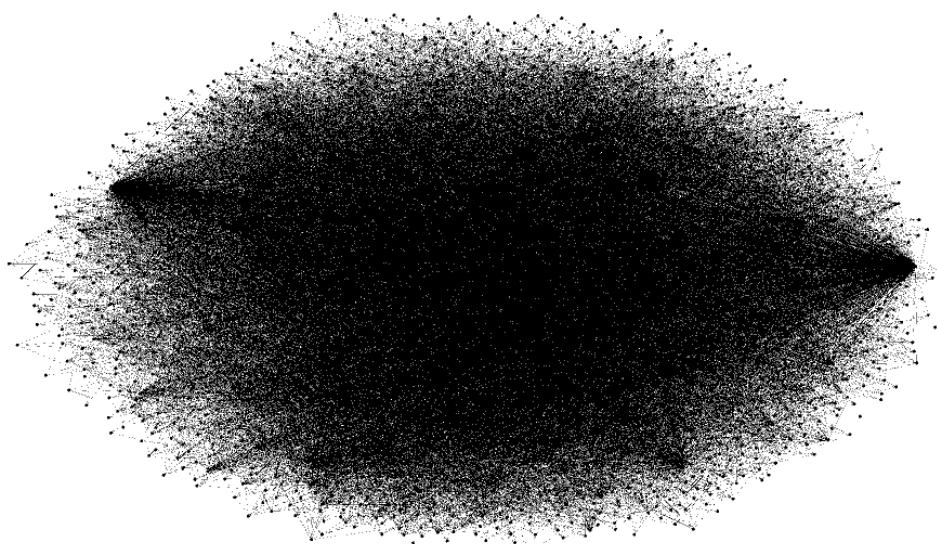


Figura 13: Closeness Centrality

ANALISI SECONDO DATASET

DATI GENERALI SECONDO GRAFO:

Grafo:



Minimo grado Nodo:

Il nodo con il peso minore è la parola “AlbertoBargioni” poiché, questo nodo non ha collegamenti con altre parole all’interno del nostro dataset.

Label	Degree	Weighted Degree
AlbertoBargioni	0	0.0

Massimo grado Nodo:

Il nodo con il peso maggiore è la parola “Coronavirus” poiché è una parola di uso frequente nei titoli.

Label	Degree	Weighted Degree
Coronavirus	1130	2387.0

Media grado Nodi:

Average Degree	Avg. Weighted Degree
14,332	17,383

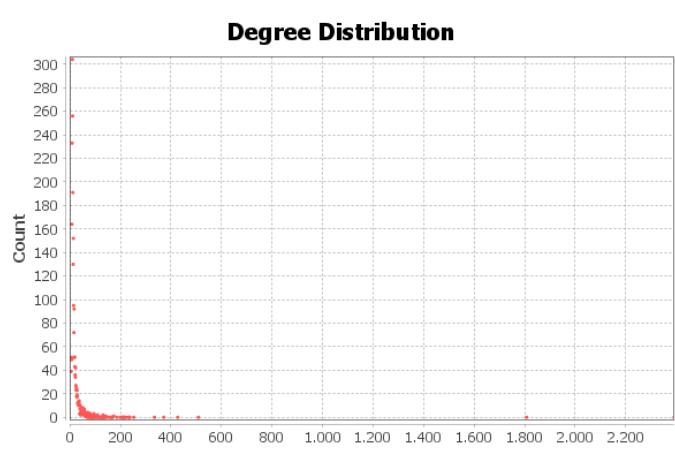
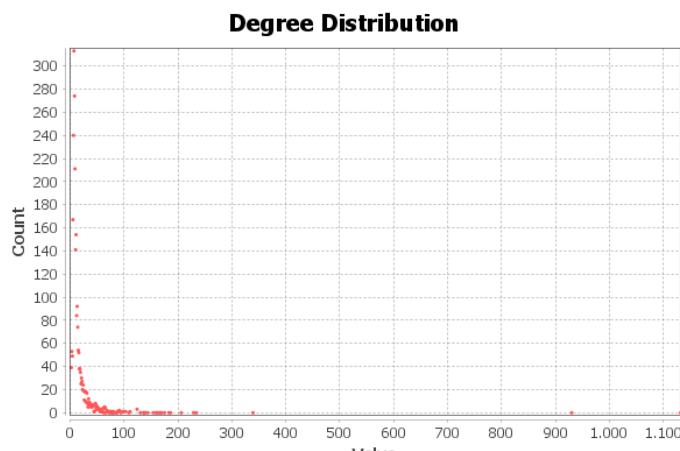
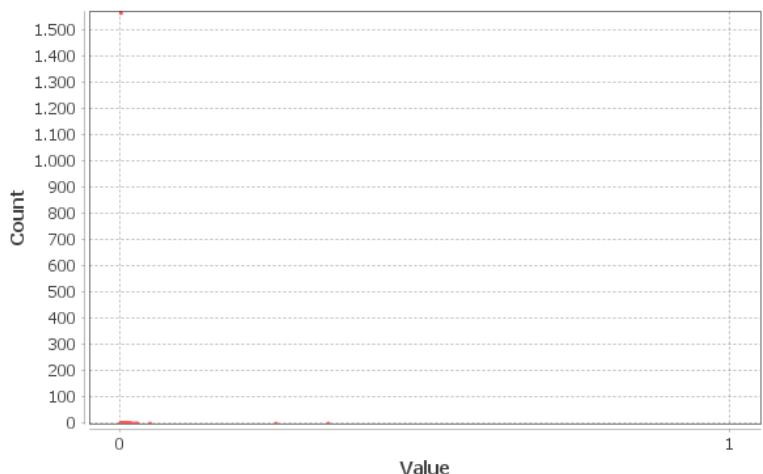


Figura 14: Average Degree e Avg:Weighted Degree

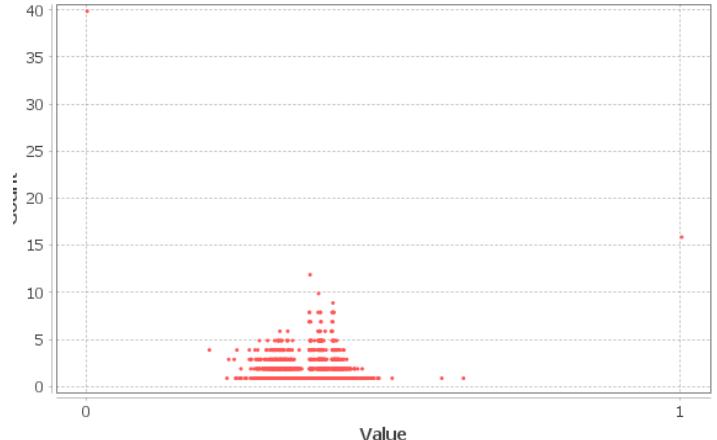
Diametro:

Calcolando il diametro, con l'opzione Normalize Centralites [0;1]), il risultato è 6.

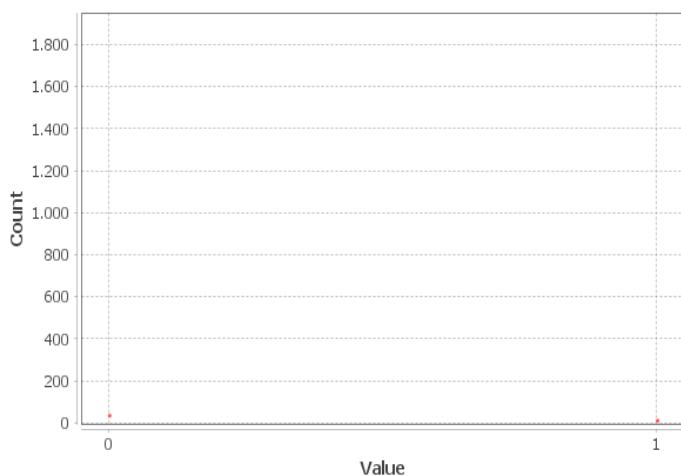
Betweenness Centrality Distribution



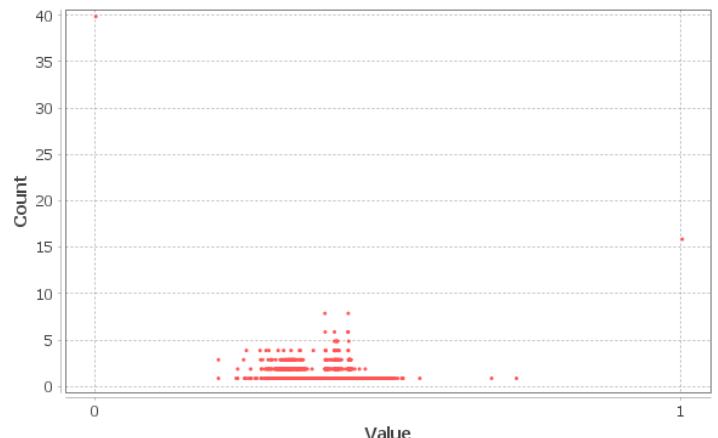
Closeness Centrality Distribution



Eccentricity Distribution



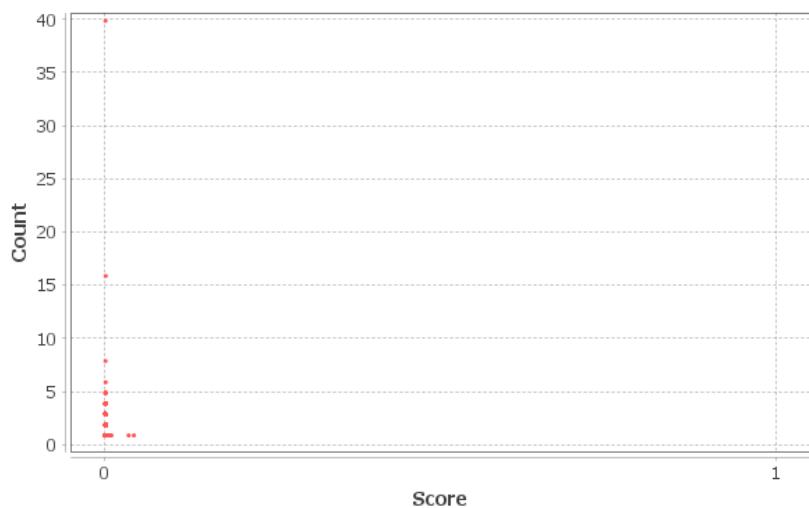
Harmonic Closeness Centrality Distribution



PageRank Massimo e Media:

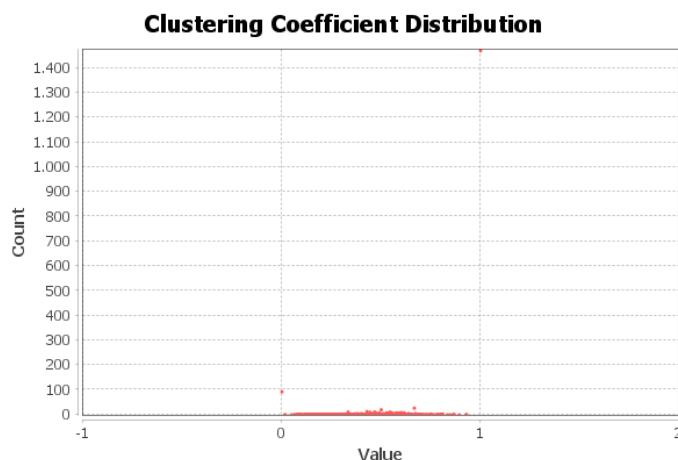
Max. PageRank	Avg. PageRank
0,042263	0,0003821

PageRank Distribution



Coefficiente di Clustering:

Il coefficiente di Clustering è la misura del grado in cui i nodi di un grafo tendono ad essere connessi fra loro.
Il nostro grafo ha il seguente valore: "0,756".



Numero Componenti connesse:

Il numero delle componenti connesse del nostro grafo è: 49

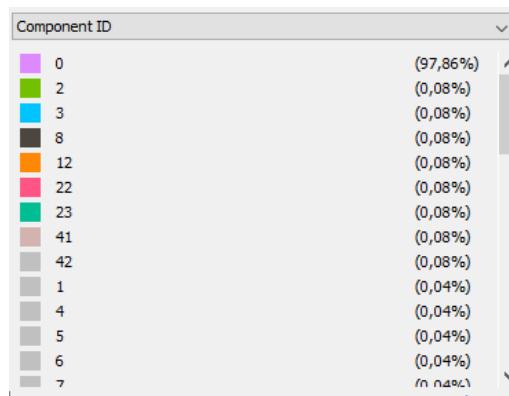
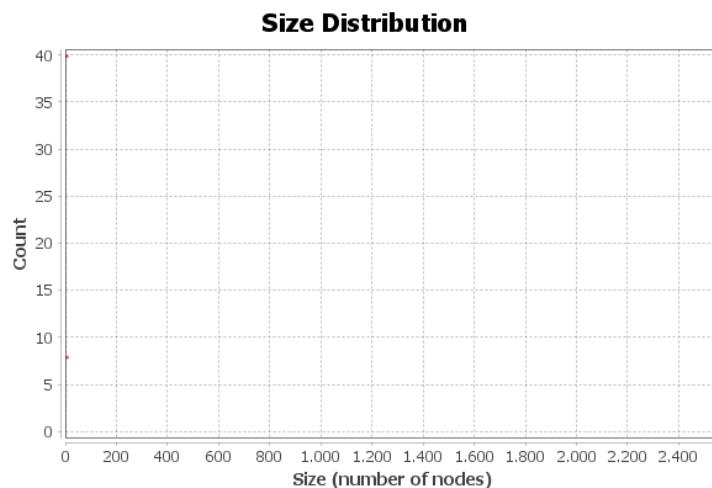


Figura 15 Colori e Percentuali relative alle componenti connesse del grafo

Come vediamo nella Figura 15, abbiamo una componente连通的 che include il 97,86% dei nodi. Le restanti componenti connesse hanno una percentuale molto bassa di inclusione dei nodi quindi sono anche poco significative ai fini dello studio.



Screen componenti connesse del grafo:

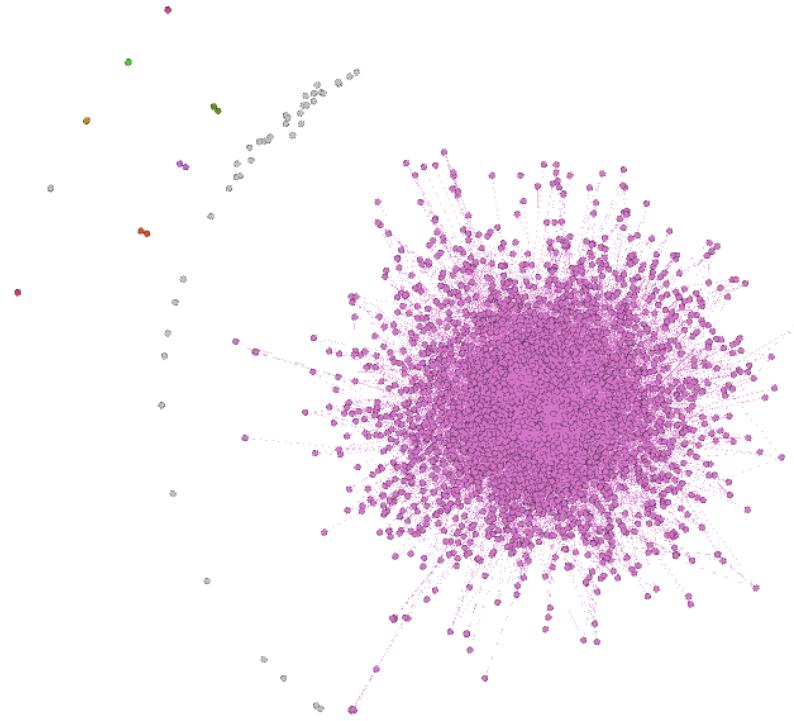


Figura 16: Componenti connesse

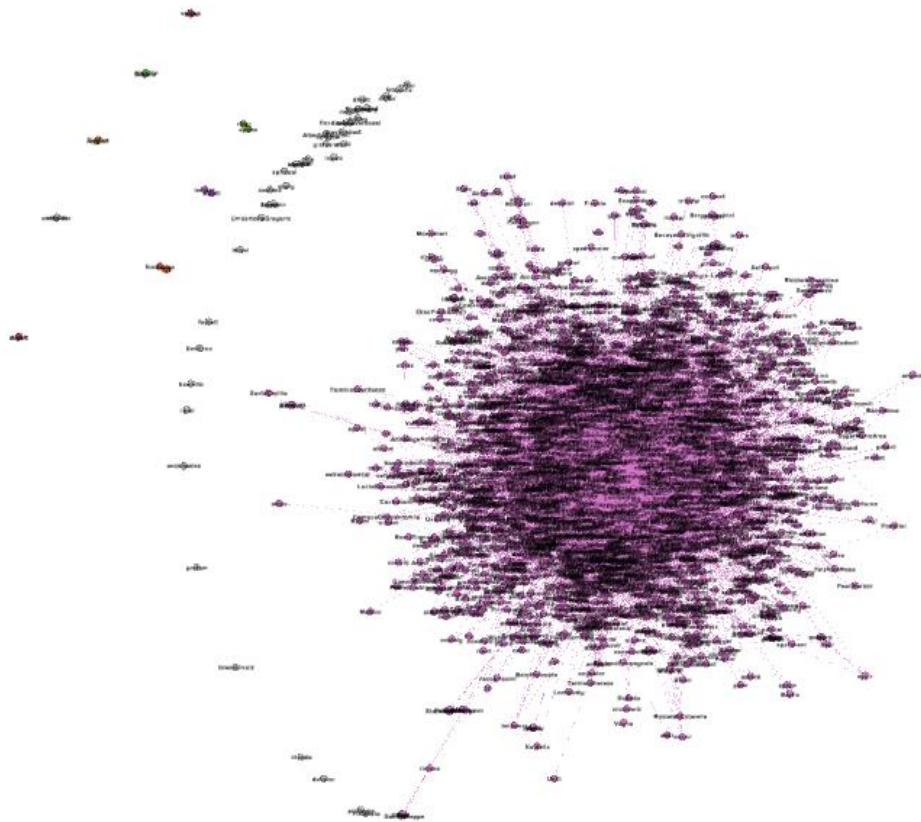


Figura 17: Componenti connesse con label

LE COMUNITÀ PRINCIPALI DEL SECONDO DATASET

LE 4 COMUNITÀ PRINCIPALI

Per trovare le quattro comunità principali abbiamo usato una modularità con risoluzione “1,5”. Il totale dei nodi ricoperti da tutte le comunità è dell’84,56%.

E abbiamo ottenuto il seguente risultato:

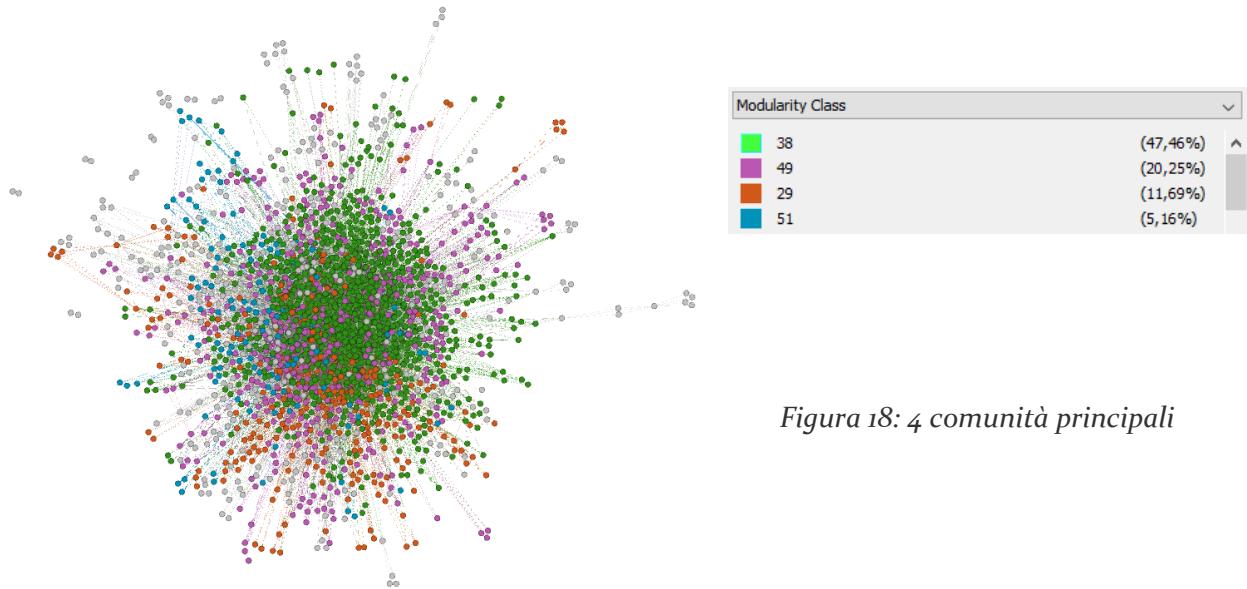


Figura 18: 4 comunità principali

Nel nostro caso non c’è una comunità che include il 50% dei nodi. Come vediamo dalla figura abbiamo una comunità (38) che include il 47,46% dei nodi.

In figura (18) vediamo la comunità (38) cioè la più grande delle 4 selezionate che contiene più del 30% dei nodi del nostro grafo.

Poiché essa contiene parole rilevanti come “mascherine”, “Coronavirus”. Tali nodi ricoprono le parole che sono state di uso frequente durante il periodo analizzato, confermando così la bontà dello studio fatto.

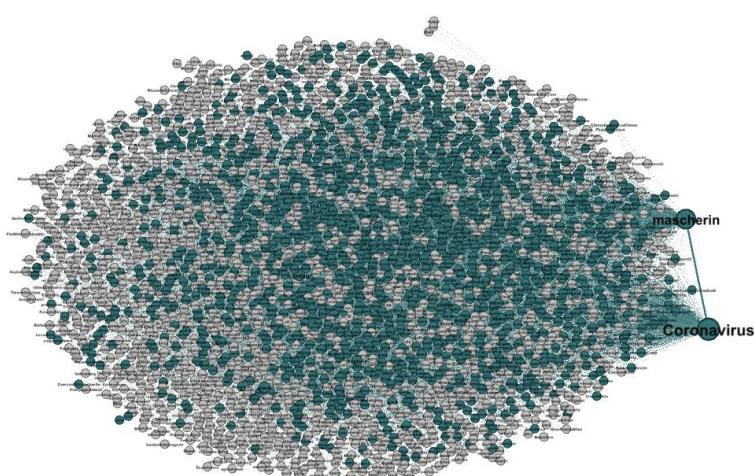


Figura 19: comunità più grande delle 4 principali

LE 2 COMUNITÀ PRINCIPALI

Per trovare le due comunità principali abbiamo usato una modularità con risoluzione “1.7”. Il totale dei nodi ricoperti da tutte le comunità è dell’88,92%.

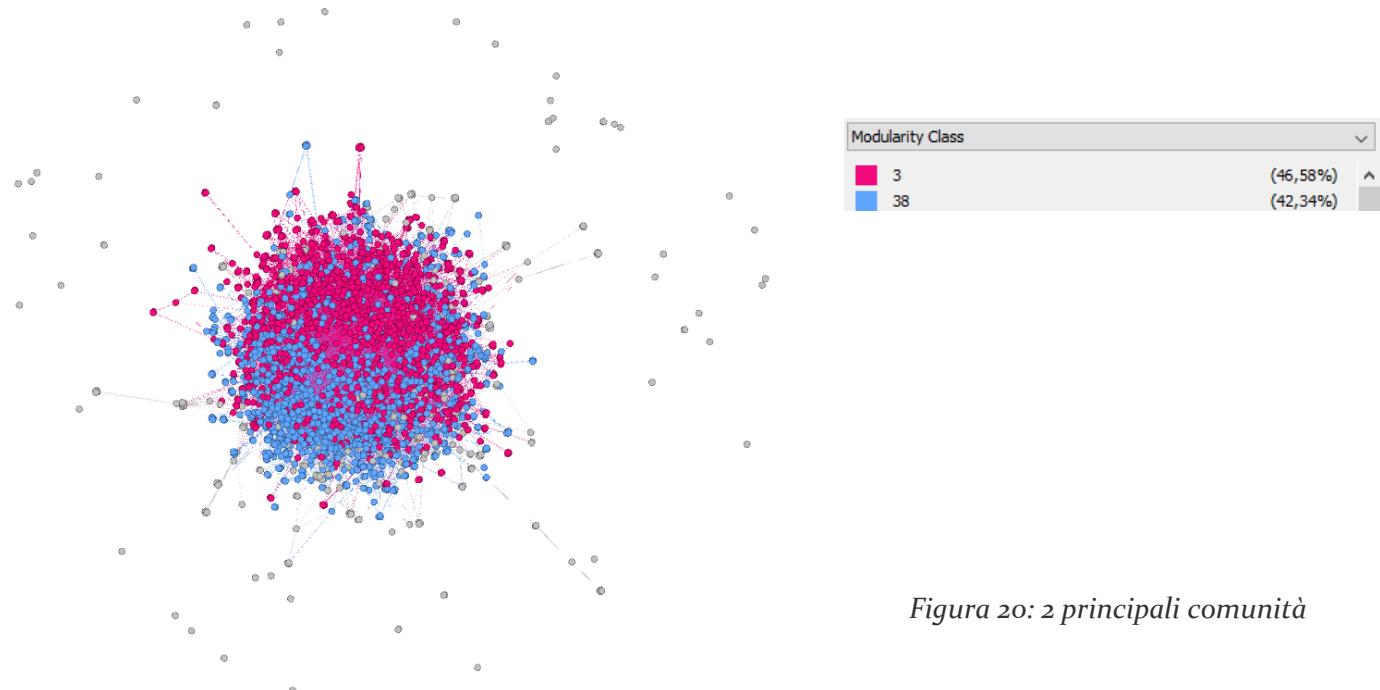


Figura 20: 2 principali comunità

Come possiamo vedere in Figura 20 il grafo può essere suddiviso in due comunità che ricoprono rispettivamente 46,58% e 42,34%. Anche in questo caso non abbiamo una comunità che include almeno il 50% dei nodi. Invece entrambe superano il 30% di nodi inclusi. Questo risultato è visibile in modo più chiaro attraverso la seguente figura.

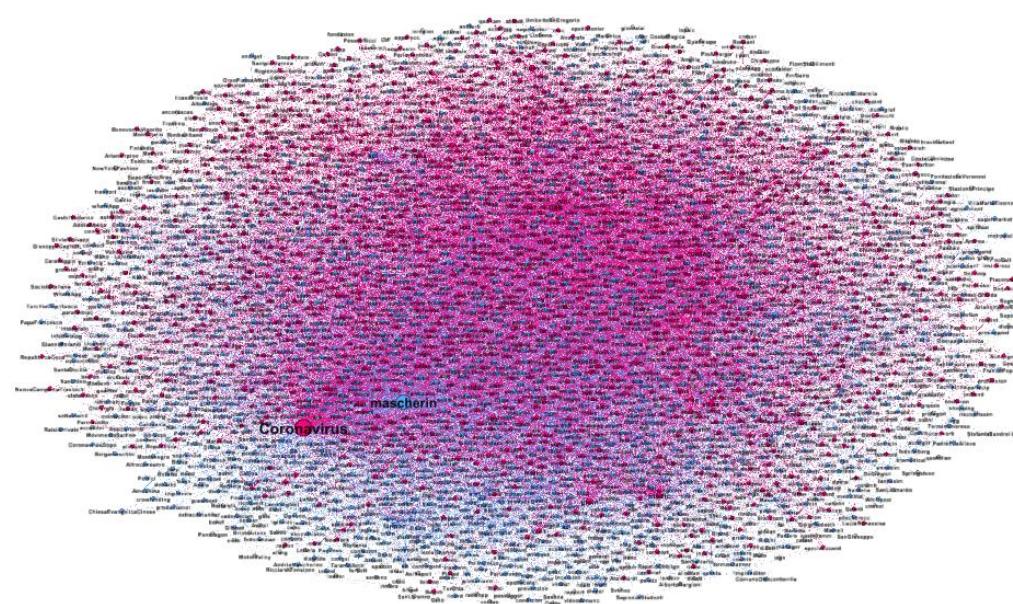


Figura 21: Le due principali comunità con label

Come possiamo vedere sulla Figura 21, le due principali comunità sono composte da parole rilevanti che sono state presenti in molti titoli del giornale “LaRepubblica”. In particolare a differenza della precedente suddivisione in 4 comunità, abbiamo come risultato una divisione netta in due comunità composte rispettivamente da due nodi molto rilevati sotto il punto di vista del loro grado. In particolare il grado nella figura 17 viene rappresentato attraverso la grandezza del nodo. Infatti la comunità 38 ha come nodo “mascherin*” che rappresenta la parola presa in analisi per la ricerca dei titoli. Invece la seconda comunità, cioè la 3, ha come nodo principale “Coronavirus” a cui si collegano la maggior parte dei nodi.

ULTERIORI METRICHE DI ANALISI DEL SECONDO DATASET

GRADO

Il grado di un nodo, ha un'importanza nello studio del grafo poiché è uguale al numero di archi che incidono sul nodo stesso.

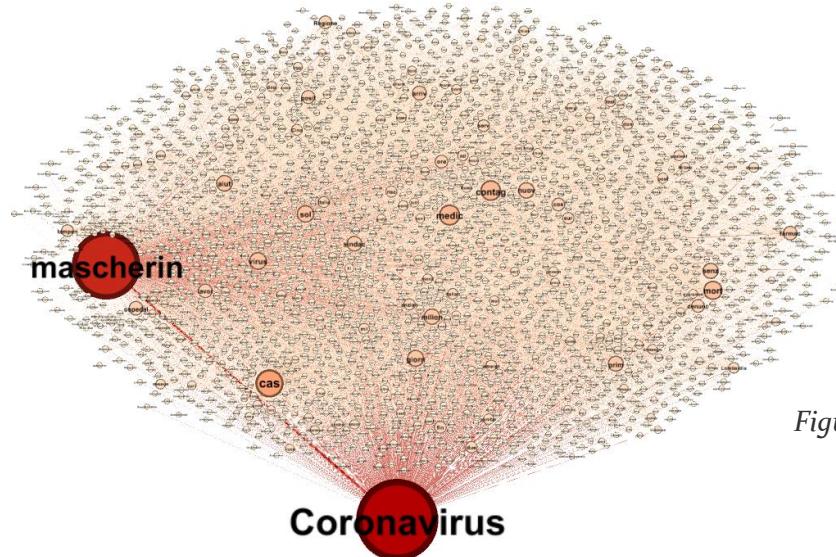


Figura 22: Ranking con grado

Nella figura seguente si può vedere la grandezza delle label e dei nodi proporzionale alla dimensione del grado di ogni nodo. A differenza del grafo del primo dataset possiamo notare che non abbiamo un solo nodo principale ma in questo caso abbiamo due nodi principali, ovvero con grado maggiore degli altri, che sono le parole “mascherin*” e “Coronavirus”.

EIGENVECTOR CENTRALITY

Come possiamo vedere dalla Figura 23 abbiamo una differenza netta tra le parole più importanti e altre parole, che nel primo dataset erano rilevanti adesso non lo sono più. Rimangono comunque importanti le parole “mascherin*” e “Coronavirus”. Dato l'aumentare dei dati rispetto al primo dataset abbiamo di conseguenza un grado di attendibilità maggiore nei risultati dell'analisi con questa metrica.

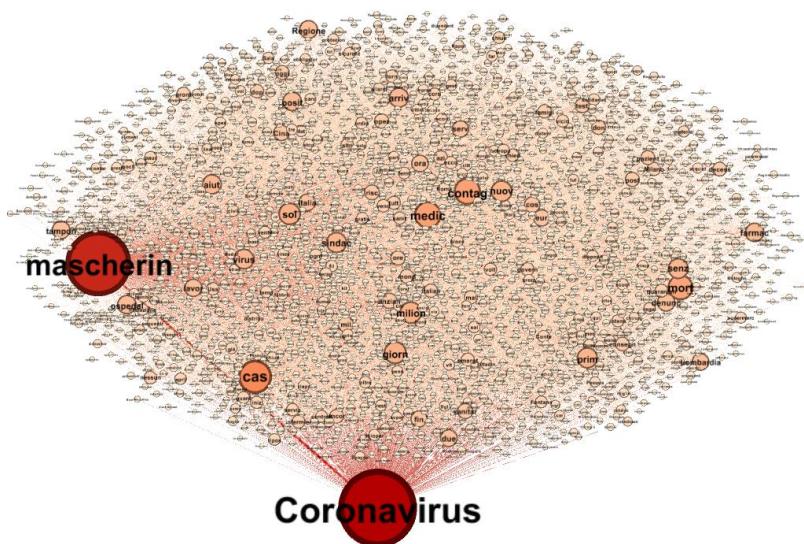


Figura 23: Eigenvector Centrality

BETWEENNESS CENTRALITY

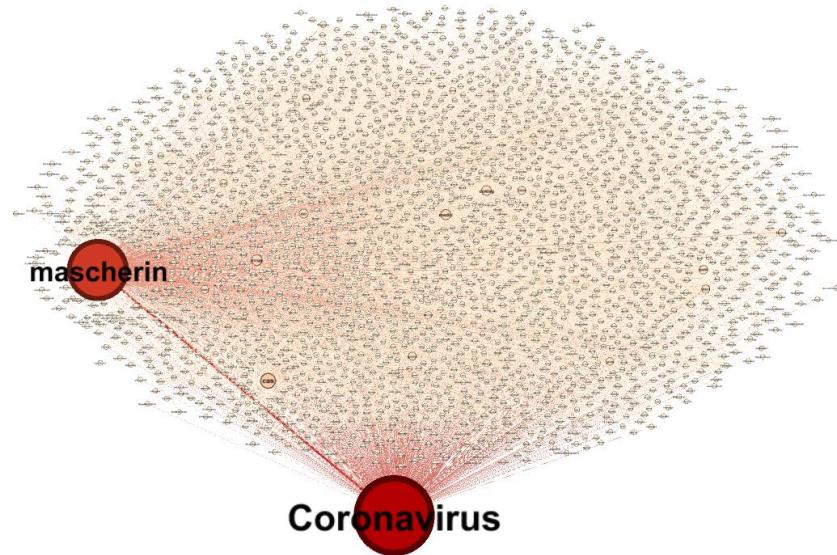


Figura 24: Betweeness Centrality

Con questa metrica vediamo che i nodi con un'elevata betweeness centrality sono “Coronavirus”, “mascherin*” poiché nel periodo del secondo dataset hanno avuto una grande influenza nella creazione dei titoli giornalistici. Infatti questi due nodi fanno da collegamento tra altri nodi della rete.

CLOSENESS CENTRALITY

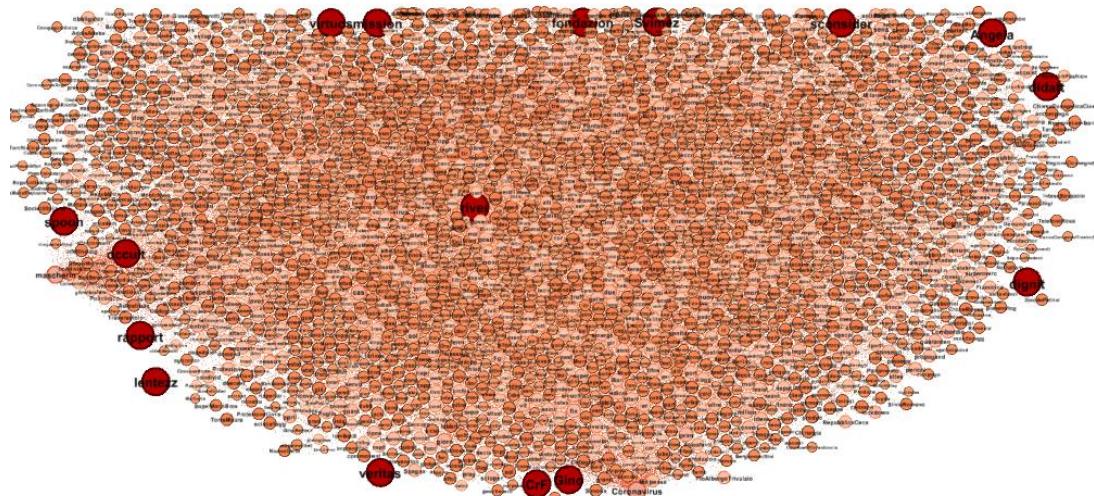


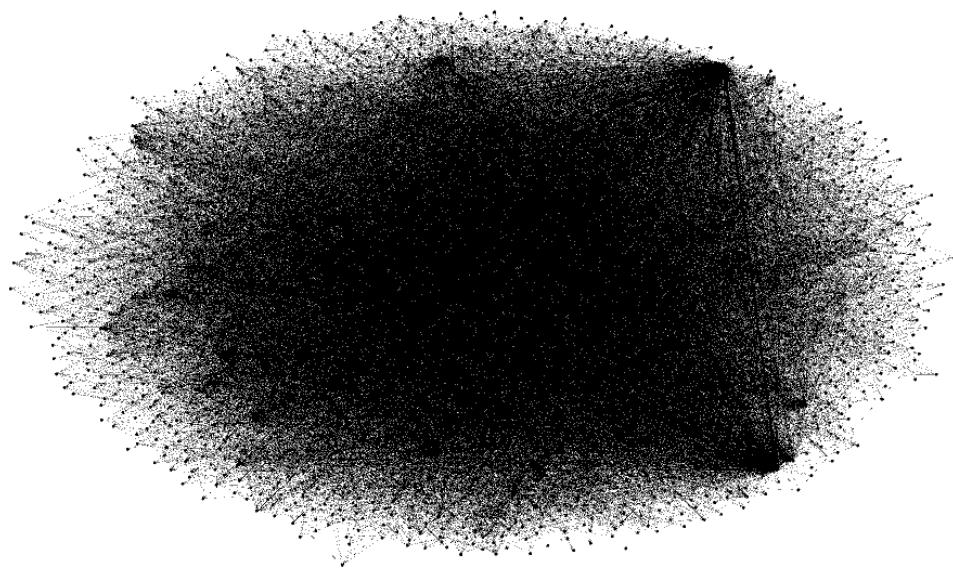
Figura 25: Closeness Centrality

Come vediamo dalla figura la closeness centrality è maggiore nelle componenti connesse con pochi nodi. Data la quantità di dati maggiore a quella del primo dataset abbiamo nuovi nodi che hanno una maggiore prossimità rispetto a quelli analizzati precedentemente.

ANALISI DATASET COMPLETO

DATI GENERALI GRAFO COMPLETO:

Grafo:



Minimo grado Nodi:

Il nodo con il peso minore è la parola “AlbertoBargioni” poiché, questo nodo non ha collegamenti con altre parole all’interno del nostro dataset.

Label:	Degree:	Weighted degree:
AlbertoBargioni	0	0.0

Massimo grado Nodi:

Il nodo con il peso maggiore è la parola “Coronavirus” poiché è una parola di uso frequente nei titoli.

Label:	Degree:	Weighted degree:
Coronavirus	1207	2601.0

Media grado Nodi:

Average Degree	Avg. Weighted Degree
14,933	18,151

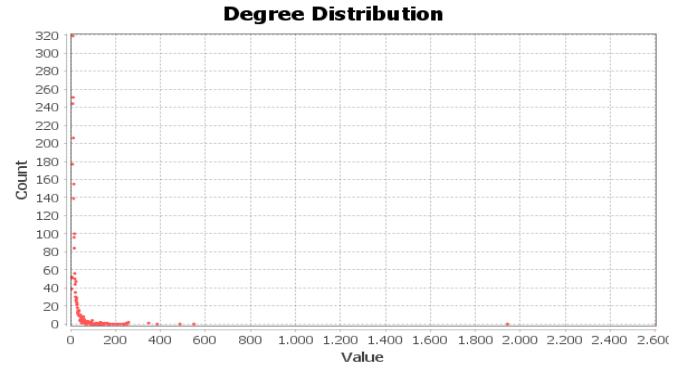
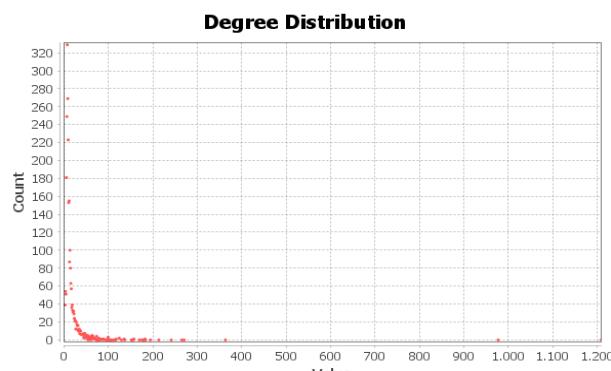
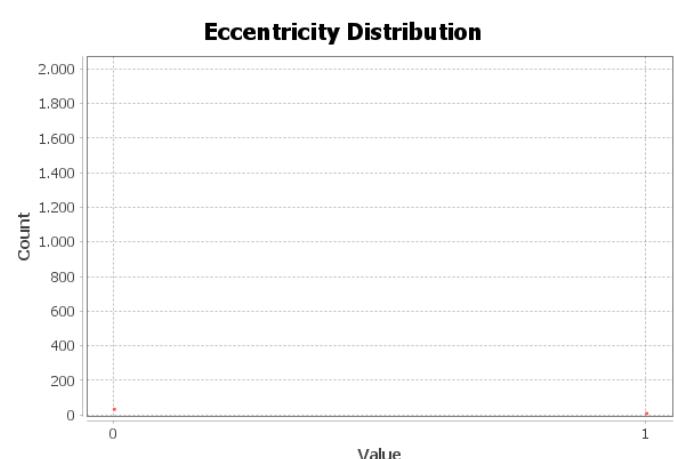
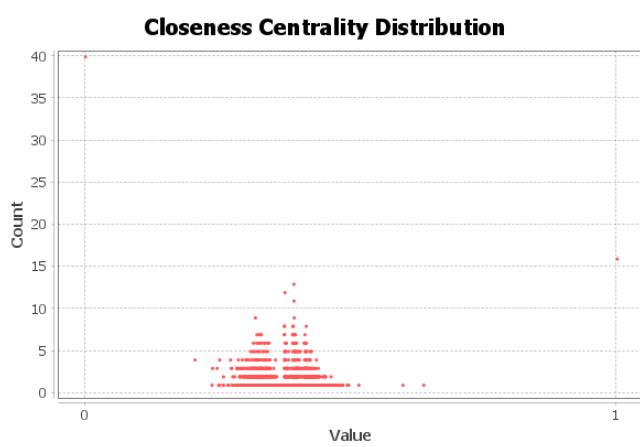
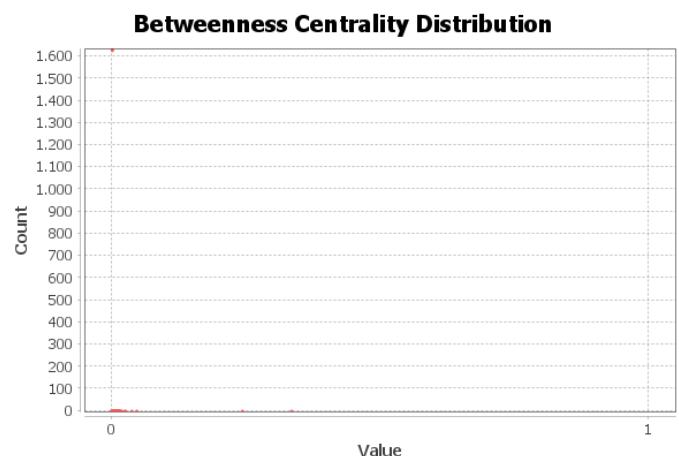
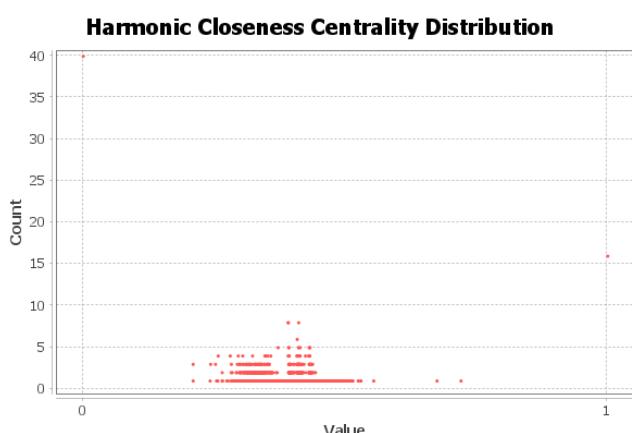


Figura 26: Average degree e Avg. Weighted Degree

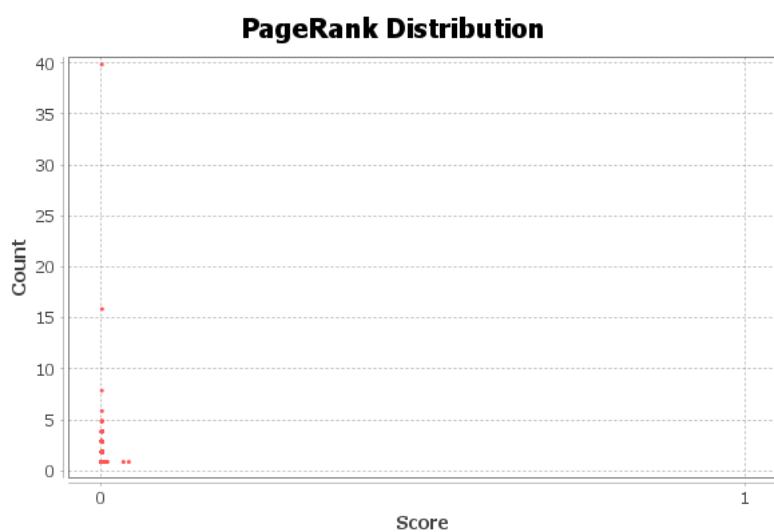
Diametro:

Calcolando il diametro, con l'opzione Normalize Centralites [0;1]), il risultato è 6.



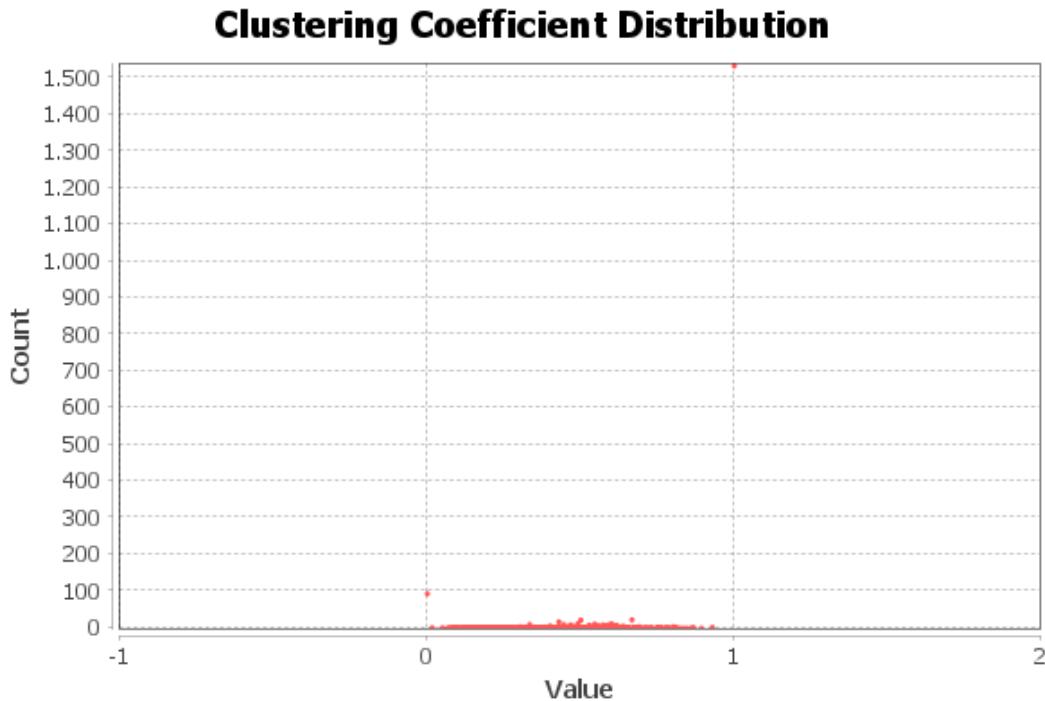
PageRank Massimo e Media:

Max. PageRank	Avg. PageRank
0,041925	0,0003621



Coefficiente di Clustering:

Il coefficiente di Clustering è la misura del grado in cui i nodi di un grafo tendono ad essere connessi fra loro. Il nostro grafo ha il seguente valore: "0,752".



Numero Componenti connesse:

Il numero delle componenti connesse del nostro grafo è: 49

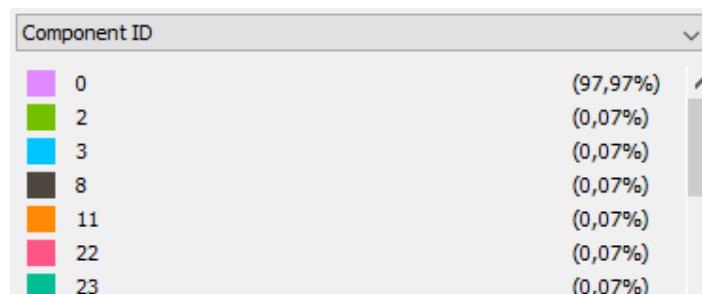
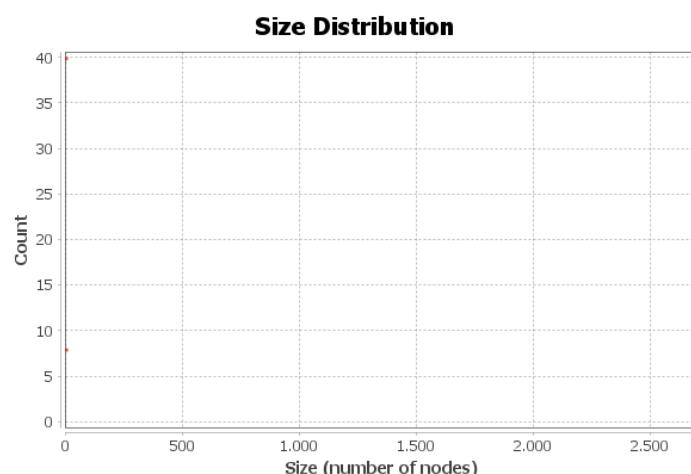


Figura 27: Colori e Percentuali relative alle componenti connesse del grafo

Come vediamo nella Figura 27, abbiamo una componente connessa che include il 97,97% dei nodi. Le restanti componenti connesse hanno una percentuale molto bassa di inclusione dei nodi quindi sono anche poco significative ai fini dello studio.



Screen componenti connesse grafo:

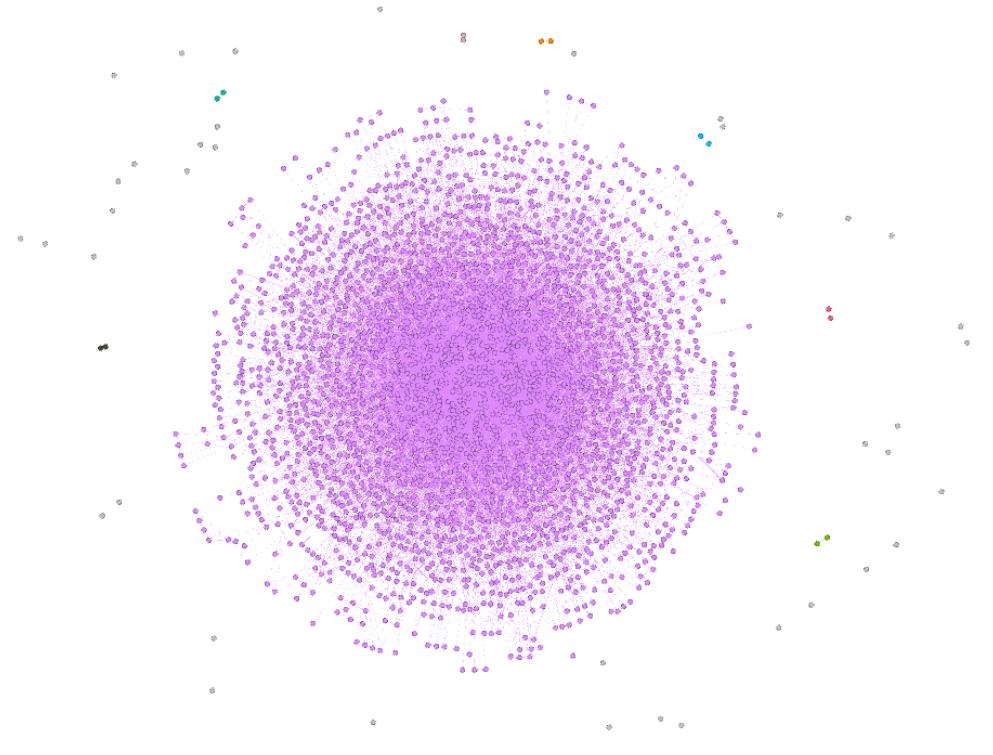


Figura 29: Componenti connesse no label

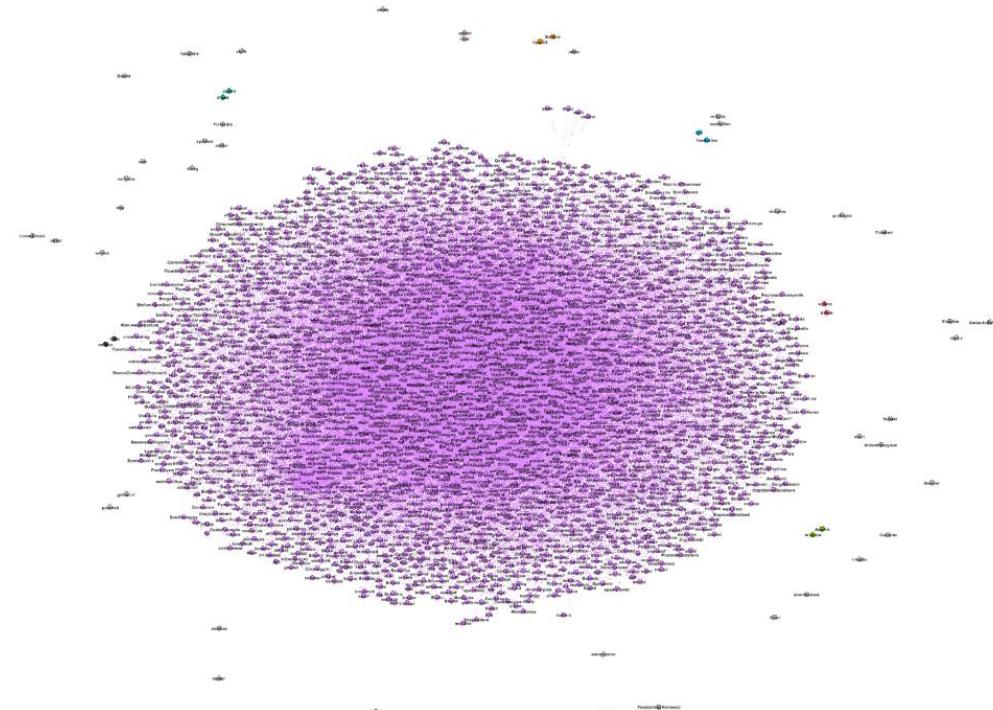


Figura 28: Componenti connesse

LE COMUNITÀ PRINCIPALI DEL DATASET COMPLETO

LE 4 COMUNITÀ PRINCIPALI

Per trovare le quattro comunità principali abbiamo usato una modularità con risoluzione “1,22”. Il totale dei nodi ricoperti da tutte le comunità è dell’75,56%. E abbiamo ottenuto il seguente risultato:

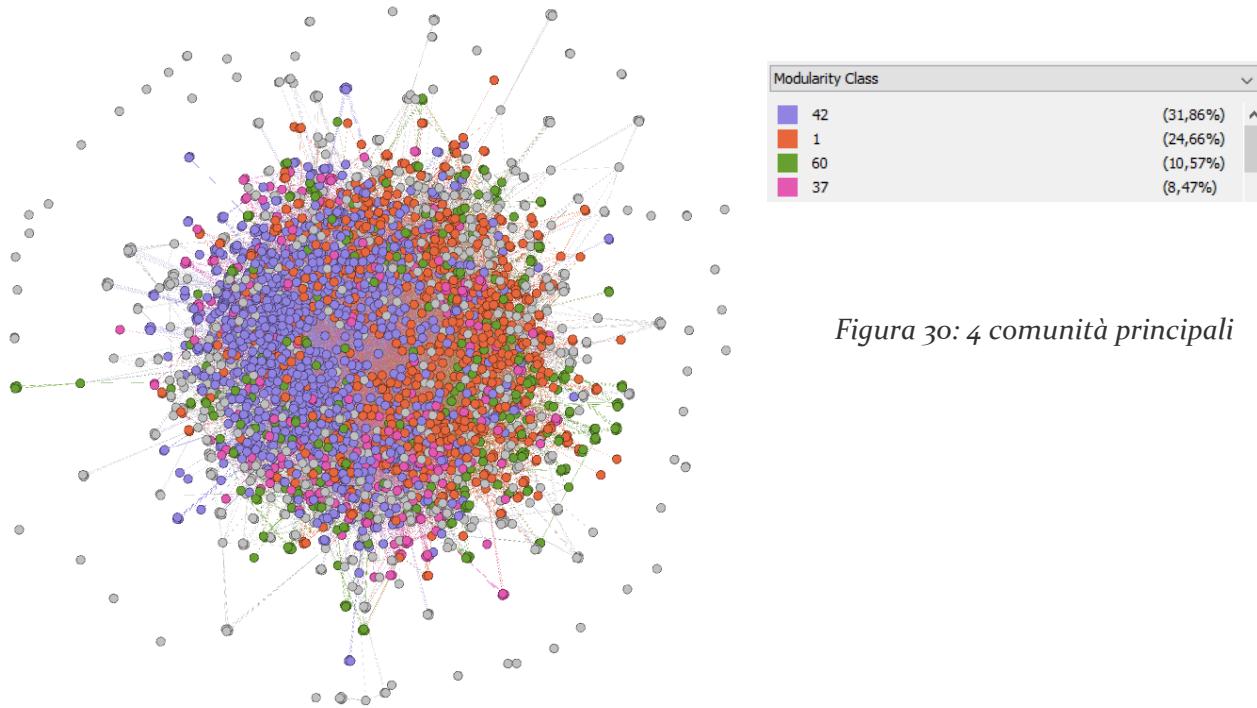


Figura 30: 4 comunità principali

Nel nostro caso non c’è una comunità che include il 50% dei nodi. Come vediamo dalla figura abbiamo una comunità (42) che include il 31,86% dei nodi.

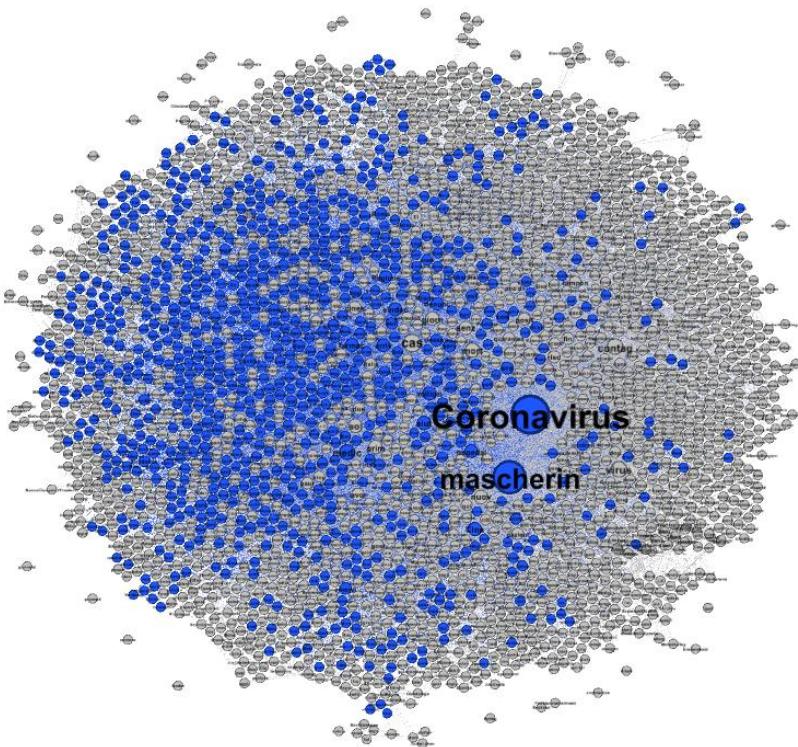


Figura 31: comunità più grande delle 4 principali

In figura (31) vediamo la comunità (42) cioè la più grande delle 4 selezionate che contiene più del 30% dei nodi del nostro grafo.

Poiché essa contiene parole rilevanti come “mascherin*” e “Coronavirus”. Tali nodi ricoprono le parole che sono state di uso frequente durante il periodo analizzato, confermando così la bontà dello studio fatto.

LE 2 COMUNITÀ PRINCIPALI

Per trovare le due comunità principali abbiamo usato una modularità con risoluzione “1.7”. Il totale dei nodi ricoperti da tutte le comunità è dell’88,92%.

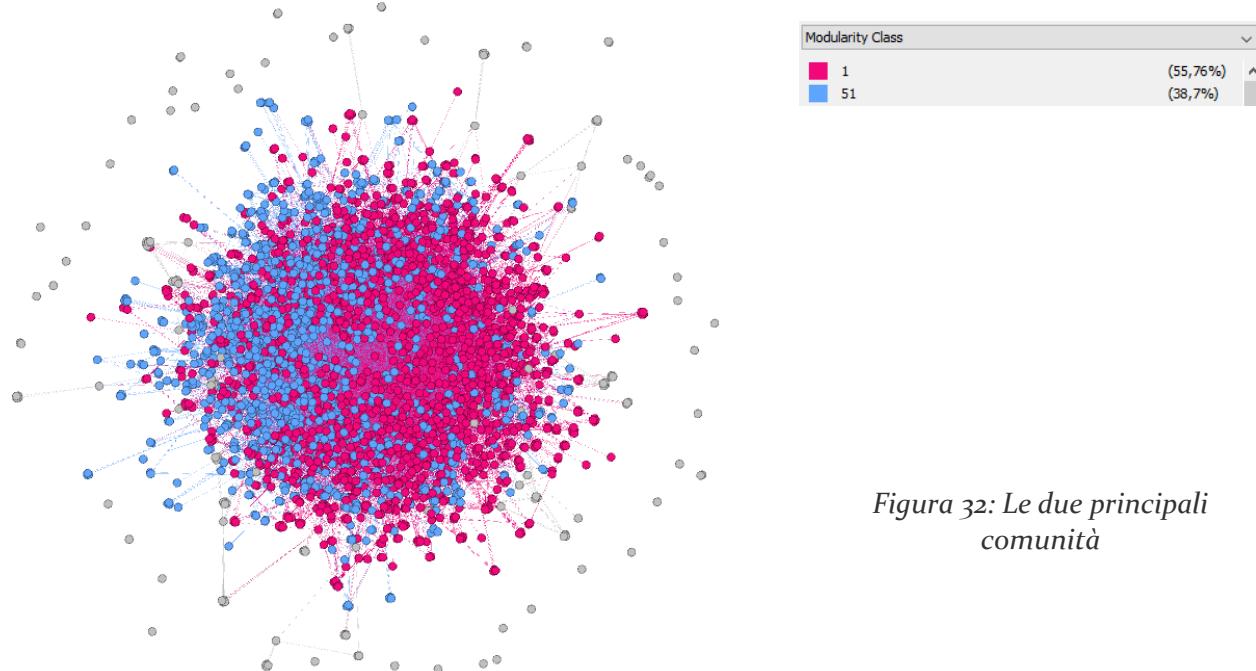


Figura 32: Le due principali comunità

Come possiamo vedere in Figura 32 il grafo può essere suddiviso in due comunità che ricoprono rispettivamente 55,76% e 38,70%. In questo caso abbiamo una comunità che include almeno il 50% dei nodi. Ed inoltre entrambe superano il 30% di nodi inclusi. Questo risultato è visibile in modo più chiaro attraverso la seguente figura.

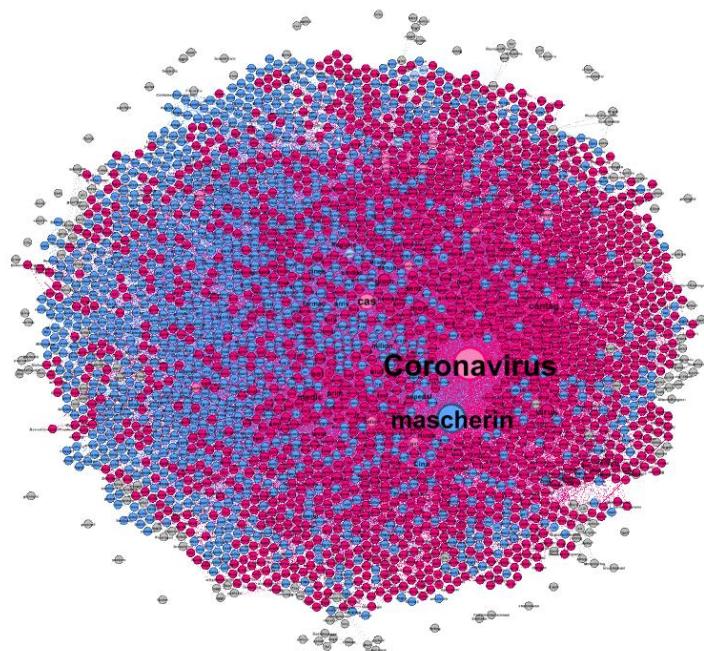
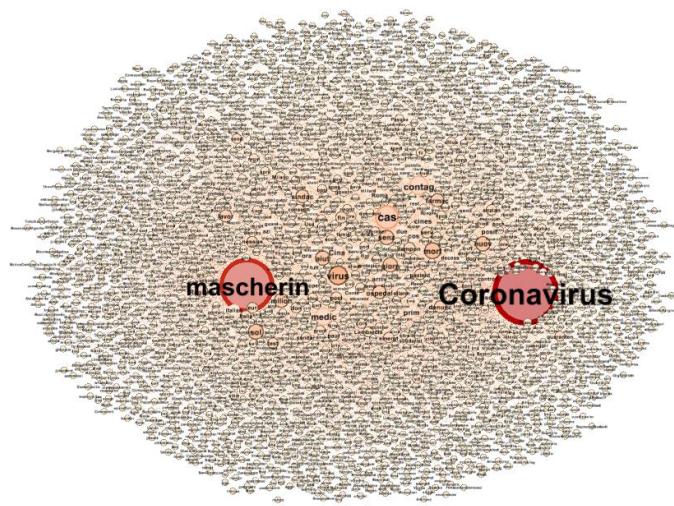


Figura 33: le due principali comunità con label

Come possiamo vedere sulla figura 33, le due principali comunità sono composte da parole rilevanti che sono state presenti in molti titoli del giornale “LaRepubblica”. Per quanto riguardo la comunità 1, che include più del 50% dei nodi, attraverso un’analisi più approfondita si nota che il nodo principale, con il più alto grado, è per l’appunto “Coronavirus”. Come per gli altri grafici si crea una distinzione netta con il secondo componente cioè il 51 che vede come nodo principale, un’altra volta, “mascherin*”.

ULTERIORI METRICHE DI ANALISI DEL DATASET COMPLETO

GRADO



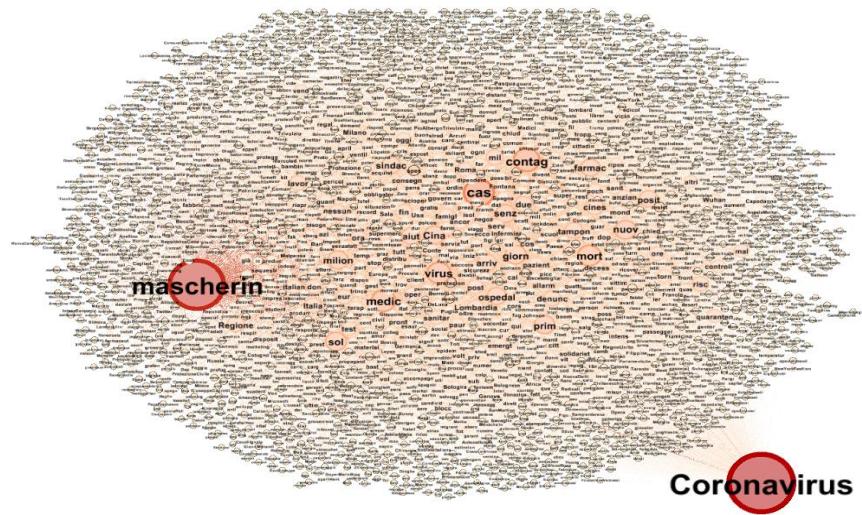
Come possiamo vedere nella figura 34 il grado dei nodi non si scosta di molto dal grado dei nodi del grafo del secondo dataset. Anche qui infatti vediamo che i due gradi maggiori vengono dai nodi “mascherin*” e “Coronavirus”.

Inoltre, anche i nodi più piccoli continuano ad affermarsi all’interno del grafo come nel secondo dataset.

EIGENVECTOR CENTRALITY

Attraverso questa metrica l’importanza e l’influenza di un nodo è data dai suoi collegamenti. Infatti vediamo che i nodi più influenti sono “Coronavirus” e “mascherin*” dato che hanno più collegamenti rispetto agli altri nodi del grafo.

A differenza del secondo vediamo anche una diminuzione dei nodi più piccoli che di conseguenza diventano meno influenti in questo dataset. I nodi centrali sono nodi che acquistano più importanza e influenza.



BETWEENNESS CENTRALITY

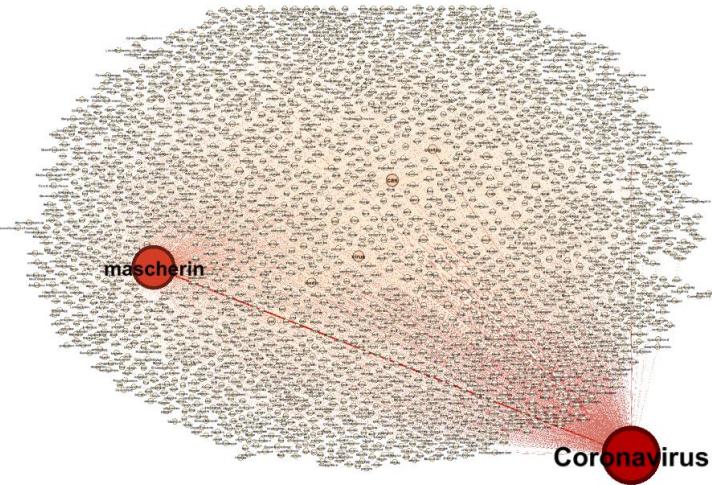
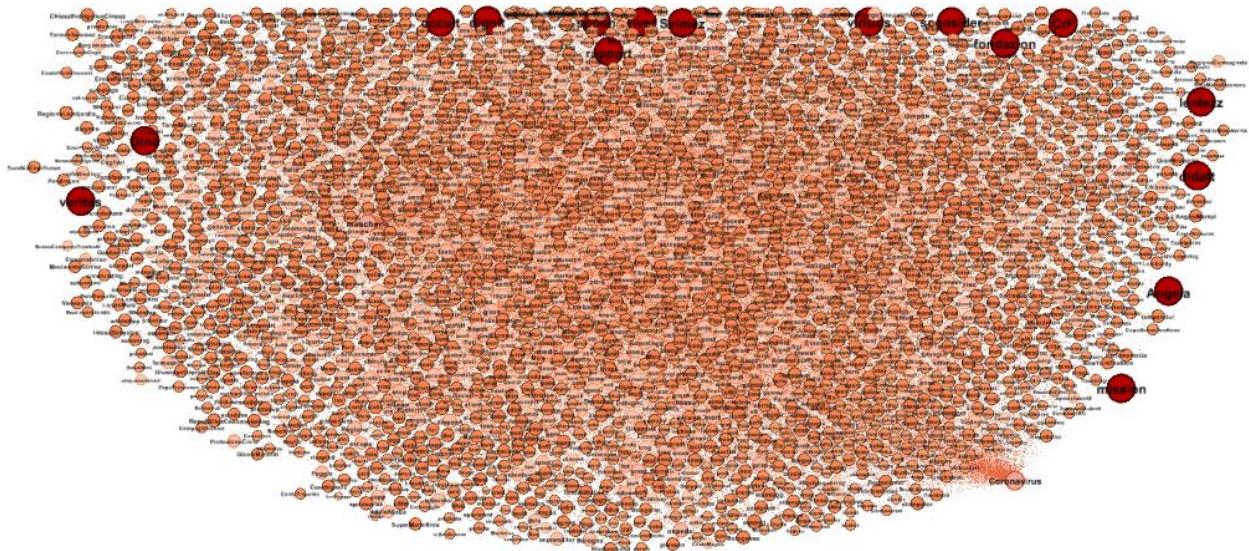


Figura 36: Betweenness Centrality

Anche in questo caso la betweenness centrality ci mostra che i nodi con elevata influenza nel flusso di informazioni sono “maschern*” e “Coronavirus” proprio come nel secondo dataset poiché esse sono parte fondamentale dei titoli.

CLOSENESS CENTRALITY

La Closeness Centrality nel dataset completo rimane simile a quella analizzata nel secondo dataset.



TOP 25% NODI

DATASET	NODO GRADO MINIMO	NODO GRADO MASSIMO	TOP 25%
PRIMO DATASET	o (Vietnam)	164 (Coronavirus)	123 -> 164
SECONDO DATASET	o (AlbertoBargioni)	1130 (Coronavirus)	847 -> 1130
DATASET COMPLETO	o (AlbertoBargioni)	1207 (Coronavirus)	905 -> 1207

Grado minimo top 25% primo dataset = $(100: 164 = 75: x) = 123$

Id	Label	Interval	Degree
16	Coronavirus		164
121	cines		119
418	virus		99

Coronavirus

Grado minimo top 25% secondo dataset = $(100: 1130 = 75: x) = 847$

Id	Label	Interval	Degree
168	Coronavirus		1130
1672	mascherin		928
949	cas		337

mascherin

Coronavirus

Grado minimo top 25% dataset completo = $(100: 1207 = 75: x) = 905$

Id	Label	Interval	Degree
179	Coronavirus		1207
1766	mascherin		975
1003	cas		361

Nel primo dataset la comunità dei gradi della top 25 % è composta dal nodo “Coronavirus”.

Come si può vedere nei dati questo nodo ha un Degree molto maggiore rispetto al secondo nodo più grande.

Mentre nel secondo dataset è subentrato il nodo “mascherin*” nella top 25%.

Di conseguenza sono aumentati i nodi della comunità top 25% da uno a due. Abbiamo la medesima situazione con il dataset completo.

CONCLUSIONI

In conclusione, dopo le varie analisi, possiamo dire che c'è una componente che è ricorrente in ogni dataset e all'aumentare dei dati cresce proporzionalmente. La componente include il nodo "Coronavirus" che è il nodo che si afferma come principale in ogni dataset.

Analizzando la parola a noi assegnata, "mascherine", vediamo come nel primo dataset, avendo meno dati, essa non sia rilevante poiché nel primo periodo ancora non si parlava molto dell'oggetto "mascherine". Mentre se confrontiamo i dati del secondo dataset e del dataset completo, che si riferiscono al secondo periodo analizzato e quindi a pieno periodo pandemia vediamo come il termine "mascherine" prende il sopravvento su altri termini e diventa uno dei due nodi principali insieme a "Coronavirus".

Vediamo che nel primo periodo quando ancora non era stato dichiarato lo stato di lockdown i nodi più rilevanti erano "Coronavirus", "Cina" e "Wuhan" possiamo facilmente denotare che prima si parlava di questo fenomeno solamente come cronaca estera. Mentre nel secondo periodo, a pieno lockdown, possiamo notare che si evidenziano nodi relativi a come comportarsi e proteggersi in questa situazione. Vediamo infatti, nel secondo periodo, che avremo dei nodi nuovi rilevanti come "casa", "contagi" e la nostra parola cercata "mascherine". Tali affermazioni si possono notare anche quando analizziamo più approfonditamente la rete con le metriche quali **grado, eigenvector, betweeness, closeness**.

Confrontando le comunità principali del primo dataset con quelle del secondo emergono i seguenti risultati:

Inizialmente possiamo notare che il componente più grande all'interno delle quattro comunità principali non contiene la parola "Coronavirus" ma contiene parole come "cina", "Wuhan", "cinesi" e "virus"; Nelle due comunità principali vediamo come nodi rappresentanti "Coronavirus" e "cines*".

Il secondo dataset, a differenza del primo, ha nel componente più grande, all'interno delle quattro comunità principali, la parola "Coronavirus" affiancata dalla parola "macherin*"; Nelle due comunità principali vediamo che i nodi rappresentanti sono "Coronavirus" e "mascherin*" che va a sostituire "cines*" che riguardava il primo periodo.

Grazie a questo studio abbiamo potuto analizzare ancora più approfonditamente il periodo della pandemia identificando così la composizione dei titoli del giornale "LaRepubblica".