

UNIVERSITÀ DI PISA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Laurea Triennale in Ingegneria Informatica

**Analisi degli effetti della quarantena sugli
utenti del subreddit The_Donald**

Candidata:

Asja Attanasio

Relatori:

Prof. Marco Avvenuti

Ing. Lorenzo Cima

Abstract

Tramite gli “Online Social Networks” (OSNs), gli utenti hanno la possibilità di pubblicare e scambiare contenuti riguardanti gli argomenti e gli interessi più popolari all’interno della comunità. Proprio per quest’ultimo motivo, le piattaforme online si sono dovute munire di sistemi di vigilanza per limitare il comportamento scorretto al loro interno. Questi sistemi prendono il nome di interventi di moderazione e, nonostante il lavoro preliminare per analizzare i loro effetti, c’è ancora da fare chiarezza sulla loro efficacia.

Attraverso l’analisi di un dataset di Reddit di circa 2,62M di commenti effettuati nel subreddit The_Donald e strumenti per la network analysis, questo studio si pone l’obiettivo di esaminare gli effetti dell’applicazione della quarantena - come intervento di moderazione - sugli utenti del suddetto subreddit.



Figura 1: Logo di Reddit nel 2023

Indice

1	Introduzione	1
1.1	Interventi di moderazione	1
1.2	Target dell'analisi	2
1.3	Lavoro svolto	2
2	Related work	3
2.1	Letteratura	3
2.2	Costruzione del grafo di coordinazione	4
3	Dataset	6
3.1	Contesto	6
3.2	Descrizione del dataset	7
3.3	Grafi di coordinazione	7
4	Community detection	9
4.1	Rappresentazione delle comunità coordinate	9
4.1.1	Modularity	10
4.1.2	Louvain	10
5	Analisi numerica delle comunità	15
5.1	Analisi per modularity	15
5.2	Analisi per Louvain	20
6	Calcolo della similarità	22
6.1	Matrici di similarità normalizzate	22
6.2	Matrici di similarità diagonalizzate	24
6.3	Coefficiente di similarità	26
7	Conclusioni	27

Capitolo 1

Introduzione

Reddit è una piattaforma online estremamente popolare che svolge un ruolo significativo nella creazione di comunità online intorno ad una vasta gamma di argomenti e interessi. Nel contesto del panorama statunitense è classificato come il sesto sito web più visitato fino a marzo 2023. Questa piattaforma si distingue per la sua organizzazione in una serie di comunità virtuali denominate " subreddit ", ciascuna delle quali è dedicata a un particolare tema.

All'interno di questi subreddit, gli utenti che ne fanno parte partecipano attivamente alle discussioni, condividono contenuti pertinenti e, in generale, interagiscono tra loro. Questa struttura permette agli utenti di connettersi con persone che hanno interessi simili o affini. Tuttavia, data l'ampia base di utenti di Reddit e la natura aperta della piattaforma, possono emergere comportamenti scorretti o contenuti problematici all'interno di questi subreddit.

Per affrontare questi problemi, sono state introdotte misure di moderazione e contenimento per proteggere gli utenti da contenuti dannosi o inappropriati e per garantire che questi possano interagire in modo positivo sulla piattaforma.

1.1 Interventi di moderazione

Su Reddit, ci sono tre tipi di interventi di moderazione che possono essere applicati: quarantena, restrizione e ban. Nel contesto di questo studio, ci concentriamo in particolare sulla quarantena, che rappresenta l'obiettivo principale della nostra analisi. La quarantena è una misura di moderazione utilizzata per ridurre la visibilità di un subreddit e scoraggiare nuovi utenti dall'iscriversi allo stesso.

Quando un subreddit viene posto in quarantena, diversi effetti si verificano. Innanzitutto, il subreddit viene rimosso dai risultati di ricerca e dal feed principale dei non iscritti, il che significa che diventa meno accessibile e meno visibile per chi non è già un membro. Tuttavia, i non iscritti possono ancora accedere al subreddit.

dit, ma prima devono visualizzare un messaggio di avvertimento che ne sottolinea il contenuto controverso o problematico.

In questo modo, la quarantena limita l'accesso e la diffusione del contenuto del subreddit, ma non lo chiude completamente, consentendo agli utenti esistenti di continuare a partecipare e discutere all'interno della comunità.

1.2 Target dell'analisi

Questo studio è incentrato sull'analisi degli impatti derivanti dall'intervento di moderazione - la quarantena - introdotto il 26 giugno 2019 sul subreddit The_Donald, specificamente sugli schemi di interazione e coordinazione tra gli utenti che vi partecipano.

Questa analisi mira ad esaminare i possibili effetti delle politiche di moderazione sulle interazioni degli utenti all'interno delle comunità e tra diverse comunità, consentendo una comprensione più completa delle conseguenze delle decisioni prese dalle piattaforme di social media.

1.3 Lavoro svolto

La prima fase di questo studio si concentra sulla creazione di una rappresentazione delle interazioni tra gli utenti, che viene effettuata mediante la costruzione di grafi di coordinazione. Questi grafi sono una sorta di mappa che mostra come gli utenti sono collegati tra loro attraverso le loro interazioni all'interno del subreddit preso in esame.

Dopo aver creato questi grafi, utilizziamo algoritmi di *community detection* per identificare e isolare comunità di utenti all'interno di queste reti. Questo processo ci permette di individuare insiemi di utenti che tendono a interagire più frequentemente tra loro rispetto ad altri, rivelando così la struttura della rete e le relazioni tra i partecipanti all'interno delle diverse comunità.

Una volta individuati i *cluster* prima e dopo l'intervento di moderazione, ne verranno analizzate delle metriche di *network analysis* - come assortività, densità, coefficiente di clustering medio e grado medio - per comprendere in che modo la quarantena abbia influenzato queste comunità dal punto di vista delle loro dinamiche interne e delle relazioni tra i membri.

La seconda sezione dello studio si dedica principalmente ad esaminare la similità tra le comunità che sono state rilevate precedentemente. Questo ci permette di ottenere una misura quantitativa della variazione avvenuta all'interno di tali comunità, e, di conseguenza, capire quanto la loro composizione interna sia cambiata a causa della quarantena.

Capitolo 2

Related work

2.1 Letteratura

Negli ultimi anni, diversi gruppi di ricercatori hanno esaminato varie modalità per l'individuazione di comunità altamente coordinate e sulla valutazione di strategie di moderazione sui social media. Tra questi:

- **Trujillo**, 2023 [16]: analizza gli effetti degli interventi di moderazione sugli utenti del subreddit "rThe_Donald". Lo studio cerca di valutare come tali interventi influenzino il comportamento degli utenti all'interno della comunità online; in particolare, come le misure di moderazione influiscano sulla tonalità e il contenuto dei commenti e delle discussioni nel subreddit.
- **Trujillo**, 2022 [15]: esamina gli impatti delle azioni di moderazione sulla comunità di "rThe_Donald" su Reddit. L'obiettivo è comprendere come tali interventi influiscano sulla dinamica e sulla qualità delle discussioni all'interno di questa comunità online.
- **Weber**, 2021 [17]: esplora come gruppi di utenti possano collaborare in modo strategico per amplificare la diffusione di contenuti o idee, fornendo un metodo di identificazione di comunità che evidenziano la coordinazione tra diversi account.
- **Nizzoli**, 2021 [12]: il metodo proposto in questo testo va oltre la semplice classificazione binaria dei comportamenti coordinati e non coordinati, permettendo di identificare diversi schemi di coordinazione e di valutare il grado di coordinazione tra le comunità.
- **Horta Ribeiro**, 2021 [13]: in questo studio vengono valutati gli impatti di un intervento di moderazione in termini di migrazione di utenti da un subreddit all'altro, concludendo che questo trasferimento può causare problemi nel controllo sulla diffusione di contenuti problematici.

- **Jhaver**, 2021 [6]: in questo studio viene analizzato l'impatto della rimozione degli utenti problematici dalla piattaforma presa in esame. La conclusione traibile da questa analisi è che il deplatforming può essere considerato come un metodo valido per la moderazione e la limitazione dei comportamenti indesiderati sulla piattaforma.
- **Chandrasekharan**, 2020 [3]: analizza gli impatti di un intervento di moderazione su larga scala all'interno di una comunità su Reddit. Esamina come questa quarantena influenzi il comportamento degli utenti, la diffusione di contenuti controversi e la dinamica complessiva della comunità.
- **Haji**, 2018 [14]: lo studio esplora gli effetti sulla dispersione dei membri di una comunità, dopo che questa viene sciolta; in particolare, la possibile creazione di nuovi spazi simili e il cambiamento nei comportamenti online dopo disgregazione degli utenti che ne facevano parte. La ricerca cerca di comprendere come la rimozione di comunità dannose possa influire sulla loro riorganizzazione e sulle dinamiche online.
- **Chandrasekharan**, 2017 [4]: gli autori di questo studio esplorano se il ban - come intervento di moderazione - ha portato a una riduzione significativa di contenuti offensivi e discriminatori o se sono stati solo spostati altrove su Reddit; quindi, valutandone l'efficacia in termini di mitigazione di discorsi d'odio sulla piattaforma.
- **Cao**, 2015 [2]: attraverso l'analisi del comportamento degli utenti nella condivisione di link, che può essere spontanea o pianificata, lo studio cerca di identificare una possibile organizzazione tra gruppi di utenti. Può essere utile per l'individuazione di schemi di condivisione coordinata.

2.2 Costruzione del grafo di coordinazione

La tesi, nel suo complesso, si fonda sull'analisi di un *grafo di coordinazione*, che è stato generato da un dataset di Reddit tramite l'applicazione del metodo di *coordination detection* sviluppato da Weber [17]. La pipeline è sommariamente descritta nei seguenti passaggi:

- **Co-actions**: il primo passo per il rilevamento di comportamenti coordinati è la scelta delle azioni che contribuiscono alla costruzione del modello di coordinazione tra gli utenti. Nel caso di Reddit, possono essere commenti alla stessa sottomissione.
- **Reti di coordinazione**: fatto ciò, si realizzano delle reti di coordinazione, ovvero dei *grafti non orientati* in cui i nodi rappresentano i diversi utenti e gli archi che li uniscono corrispondono alle co-azioni tra loro. Il peso di ciascun arco è proporzionale al numero di co-azioni che hanno coinvolto due utenti che sono collegati tramite esso.

- **Filtraggio:** vengono applicate procedure di filtraggio e aggregazione al fine di ridurre le dimensioni del grafo al fine di rendere possibile un'analisi quantitativa.
- **Comunità coordinate:** vengono individuate comunità coordinate di utenti tramite l'applicazione di algoritmi di *community detection*.

Capitolo 3

Dataset

Questo studio sfrutta come punto di partenza due grafi di coordinazione ottenuti applicando il metodo descritto nel paragrafo precedente a due dataset di Reddit, uno prima della quarantena e uno dopo. Tutte le rappresentazioni grafiche dei grafi mostrate successivamente sono state realizzate sfruttando il software di *network analysis* Gephi [5].

3.1 Contesto

Tutti i dati utilizzati in questa analisi sono stati raccolti dagli archivi mensili di Pushshift [1], un servizio che fornisce uno storico dei dati di Reddit.

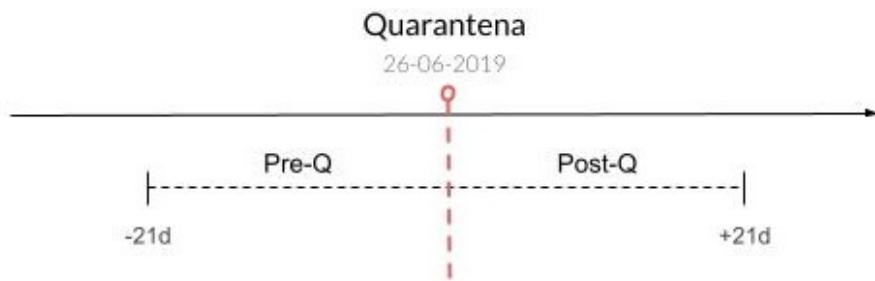


Figura 3.1: Timeline che rappresenta la finestra temporale, intorno alla quarantena, utilizzata per la raccolta e l'analisi dei dati.

Rispetto al totale dei dati forniti nel dataset di Reddit, lo studio si è concentrato su quelli estratti da 21 giorni prima a 21 giorni dopo la quarantena, ovvero nel periodo tra il 5 Giugno e 17 Luglio 2019.

Dall'insieme dei dati ottenuti sono stati filtrati i bot in modo da semplificare l'identificazione di comunità genuine. I criteri per il filtraggio sono i seguenti:

- il nome dell'utente contiene "auto" o "bot" nel nome;

- lo stesso utente ha postato più di un commento nello stesso istante.

Nel corso delle operazioni di filtraggio, mi sono trovata di fronte a una sfida nella gestione di un volume di dati così elevato. L'inesperienza nell'elaborazione di dataset di queste dimensioni mi ha portata all'implementazione di strategie poco efficienti nel codice di filtraggio, risultando in un consumo eccessivo di memoria e, di conseguenza, in prestazioni notevolmente rallentate.

La soluzione adottata è stata quella di sostituire le righe di codice

```
1 new_row = pd.DataFrame({"author": [author], "ts": [timestamp]})  
2 df = pd.concat([new_row, df], ignore_index = True)
```

Listing 3.1: codice precedente

con le seguenti

```
1 comments_instants.append(timestamp)  
2 authors[author] = comments_instants
```

Listing 3.2: codice modificato

Il problema del codice precedente è che la funzione *concat* in Python carica in memoria tutto il dataframe esistente e poi aggiunge una nuova riga e, all'aumentare della dimensione del dataframe, questo caricamento diventa sempre più oneroso. Quindi, al posto di utilizzare un dataframe, si sfruttano i dizionari, associando così ad ogni nome utente una lista degli istanti in cui ha postato un commento.

3.2 Descrizione del dataset

Il dataset consiste in una lista di commenti in formato JSON, ognuno dei quali ha come campi più rilevanti:

- **author_fullname**: identificatore univoco dell'autore del commento;
- **created_utc**: istante di creazione del commento;
- **subreddit_id**: identificatore univoco del subreddit;
- **subreddit**: nome del subreddit.

3.3 Grafi di coordinazione

I grafi ottenuti dall'applicazione della pipeline descritta in precedenza presentano le seguenti caratteristiche:

- **10879 nodi** prima dell'intervento e **6713 nodi** dopo;
- **41551 archi** prima dell'intervento e **49791 archi** dopo.

I grafi risultano inizialmente compatti da un punto di vista grafico. Sarà infatti necessario applicare degli algoritmi di visualizzazione per dare una rappresentazione grafica alle informazioni da essi contenuti.

Non possiamo distinguere i due grafi compatti, quindi ne viene riportato di seguito soltanto uno.

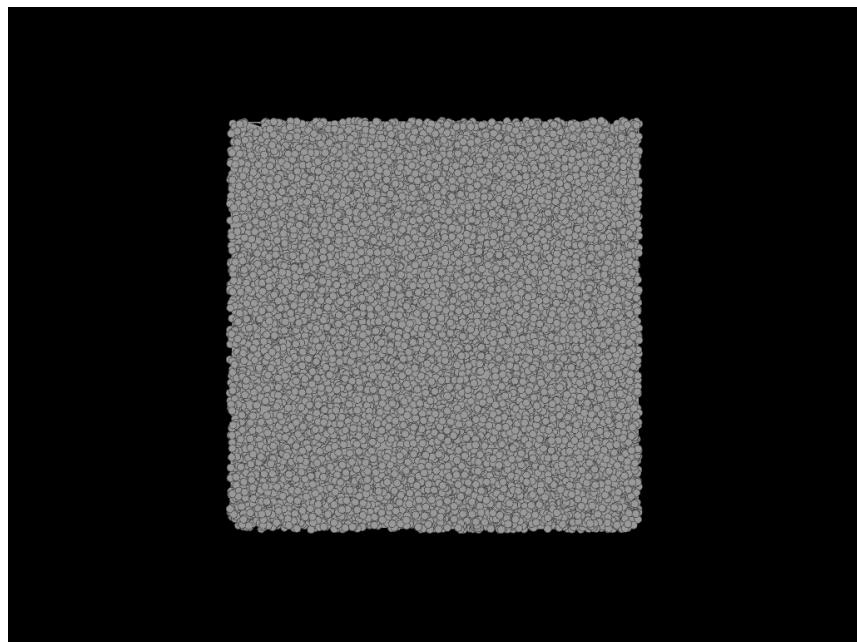


Figura 3.2: Grafo compatto pre-quarantena.

Capitolo 4

Community detection

In questo capitolo viene fatto uno studio dei grafi utilizzando la libreria Python NetworkX [11], la quale consente di attuare un'analisi della dinamica e della struttura delle reti complesse.

Per poter individuare le comunità coordinate all'interno dei grafi ottenuti precedentemente, sono stati applicati due algoritmi di community detection forniti dalla suddetta libreria: modularity [9] e Louvain [7].

4.1 Rappresentazione delle comunità coordinate

Una volta individuate le comunità con i diversi algoritmi, i due grafi ottenuti sono stati importati su Gephi, dove - per entrambi gli algoritmi applicati - vengono utilizzati gli stessi algoritmi di layout: per la prima iterazione viene utilizzato ***Radial Axis*** e per la seconda ***ForceAtlas 2***.

I parametri di Radial Axis sono:

- **Group Nodes by:** Modularity Class;
- **Node Layout Direction:** Counter Clockwise;
- **Order Nodes in Spar/Axis by:** Modularity Class;
- **Ascending Order of Spar/Axis:** true;
- **Draw Spar/Axis as Spiral:** true.

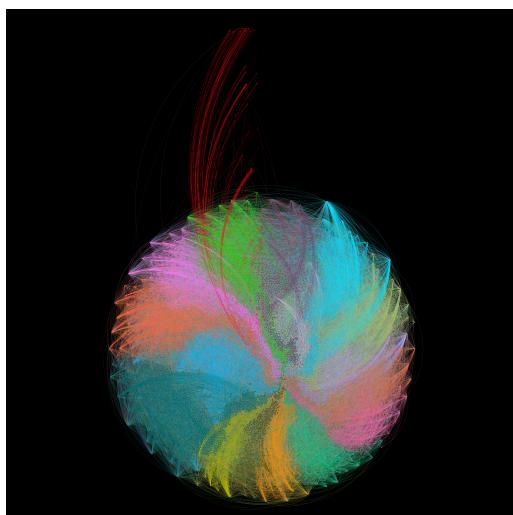
Per l'applicazione di ForceAtlas 2 sono stati usati i seguenti parametri:

- **Scaling:** 4.0;
- **Stronger Gravity:** true;

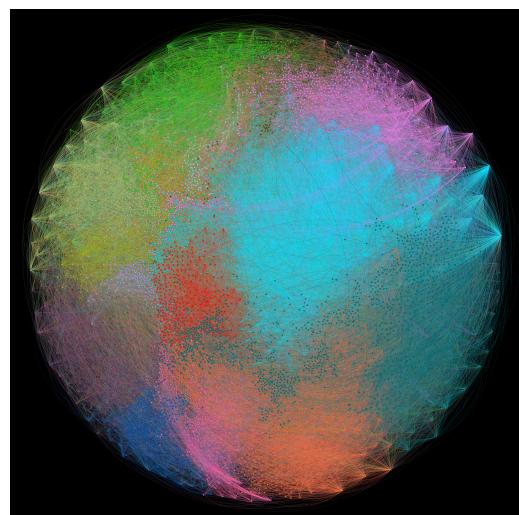
- **Gravity:** 0.03;
- **Dissuade Hubs:** true;
- **LinLog mode:** true;
- **Prevent Overlap:** true;

4.1.1 Modularity

L'algoritmo modularity viene calcolato soltanto con risoluzione 1, sia per il grafo relativo alle comunità prima dell'intervento, sia per quello dopo l'intervento.



(a) Grafo pre-quarantena.



(b) Grafo post-quarantena.

Figura 4.1: Grafo applicando modularità con risoluzione 1

Dal confronto tra i due grafici emerge come l'intervento di moderazione abbia influenzato la dinamica di coordinazione tra le diverse comunità.

Prima della quarantena, le comunità apparivano più isolate, senza particolari interazioni, mentre dopo questa fase si nota un aumento della loro coesione.

4.1.2 Louvain

Al momento di analizzare le comunità con l'algoritmo di Louvain, è stato evidenziato come la variazione del parametro di risoluzione influensi il risultato. La teoria suggerisce che diminuendo la risoluzione, il numero di comunità dovrebbe diminuire mentre il numero di membri per comunità dovrebbe aumentare. Tuttavia, l'implementazione dell'algoritmo di Louvain nella libreria '*python-louvain*' ha prodotto risultati inattesi: al di fuori di una risoluzione pari a 1, sembra che il numero di comunità aumenti invece di diminuire.

Per rendere più comprensibile la visualizzazione delle comunità, abbiamo scelto di mostrare solo le 16 comunità più grandi nei grafici successivi, mantenendo costanti i parametri di layout descritti sopra. Questa selezione ha reso i grafici a bassa risoluzione molto meno densi rispetto a quelli ad alta risoluzione, poiché le 16 comunità principali contenevano un numero significativamente inferiore di membri.

Per migliorare la chiarezza nella rappresentazione delle comunità all'interno dei grafici successivi, sono state rappresentate esclusivamente le 16 comunità più numerose, mantenendo costanti i parametri di layout precedentemente specificati. Questa scelta ha prodotto un effetto notevole sui grafici a risoluzione ridotta, che sono risultati molto meno densi in confronto a quelli ad alta risoluzione. Ciò è dovuto al fatto che le 16 comunità principali nei grafici ad alta risoluzione, avevano un numero considerevolmente maggiore di membri.

Inoltre, analizzando l'applicazione dell'algoritmo di Louvain prima e dopo un evento come la quarantena, possiamo notare delle variazioni nelle interazioni tra gli utenti all'interno delle diverse comunità virtuali.

Questi cambiamenti indicano che le misure di moderazione possono avere un impatto tangibile sulle dinamiche di interazione delle comunità virtuali, influenzando la natura e l'intensità delle connessioni tra gli utenti che vi appartengono. Questa osservazione sottolinea l'importanza di comprendere come la gestione delle comunità online possa modificare le interazioni tra gli utenti e la struttura delle stesse comunità.

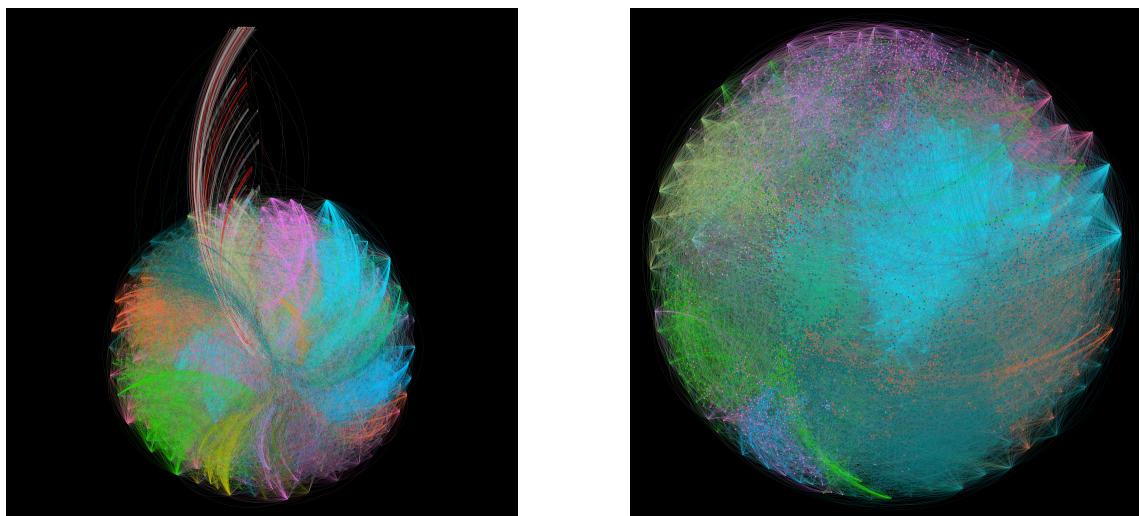
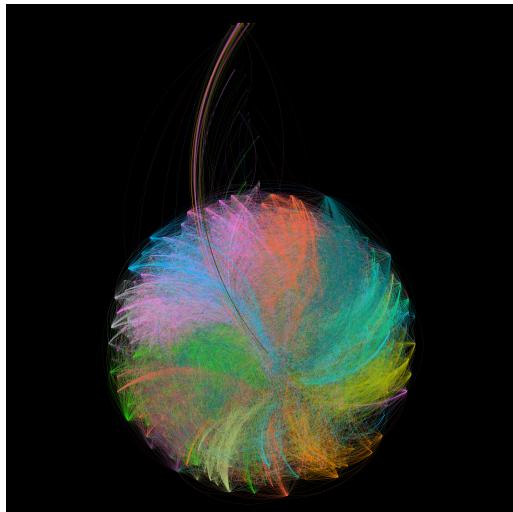
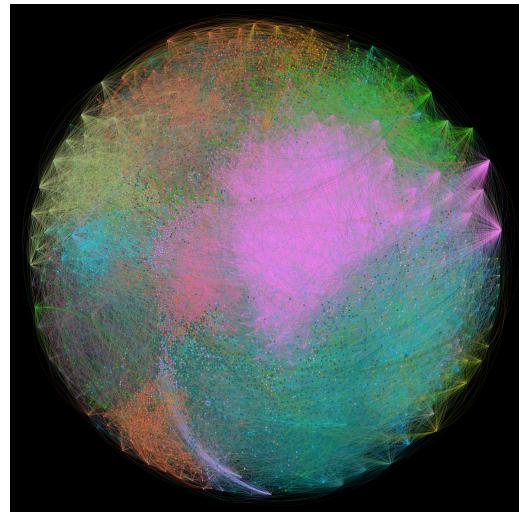


Figura 4.2: Grafo applicando Louvain con risoluzione 1

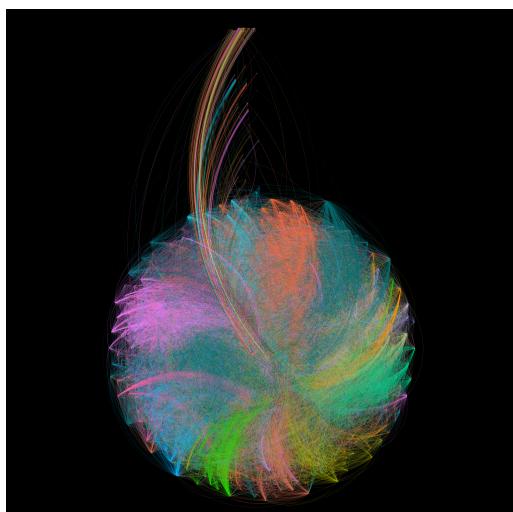


(a) Grafo pre-quarantena applicando Louvain.

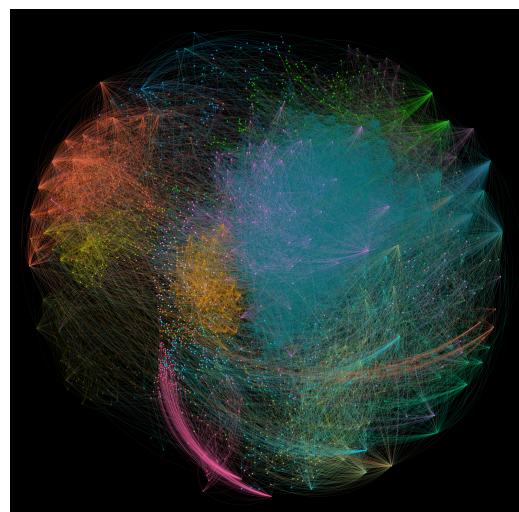


(b) Grafo post-quarantena applicando Louvain.

Figura 4.3: Grafo applicando Louvain con risoluzione 0.75

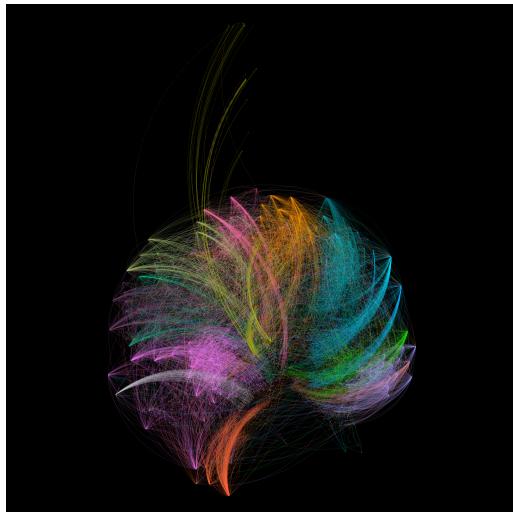


(a) Grafo prequarantena applicando Louvain.

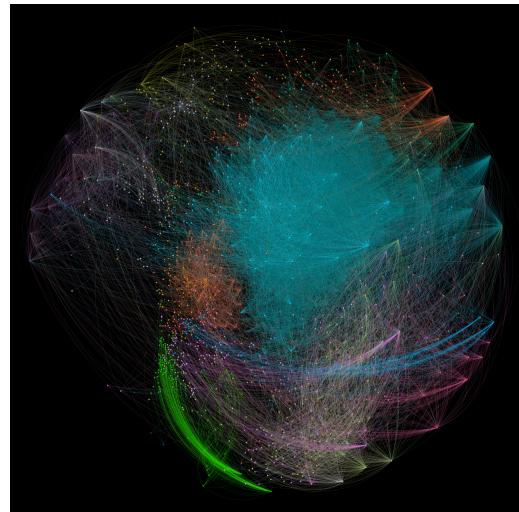


(b) Grafo postquarantena applicando Louvain.

Figura 4.4: Grafo applicando Louvain con risoluzione 0.5

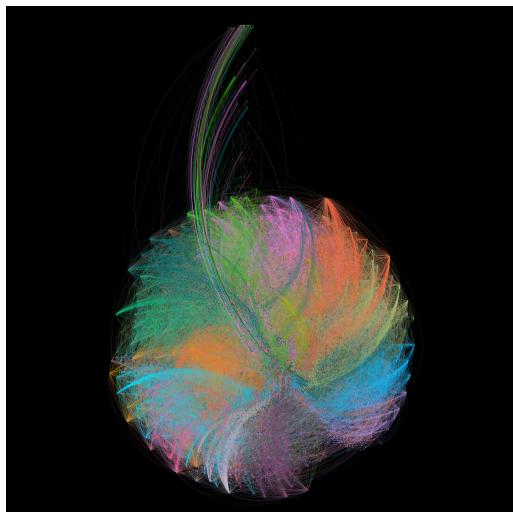


(a) Grafo pre-quarantena applicando Louvain.

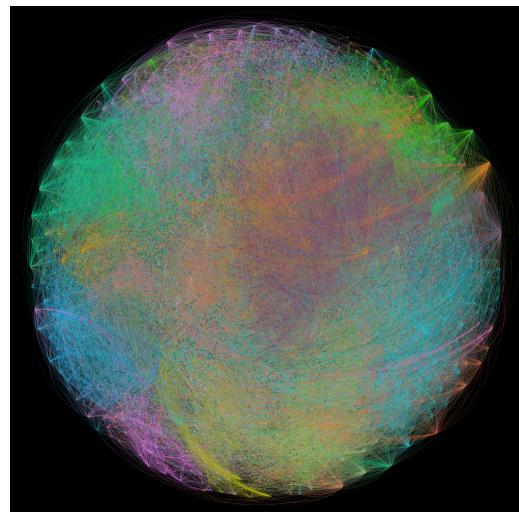


(b) Grafo post-quarantena applicando Louvain.

Figura 4.5: Grafo applicando Louvain con risoluzione 0.2

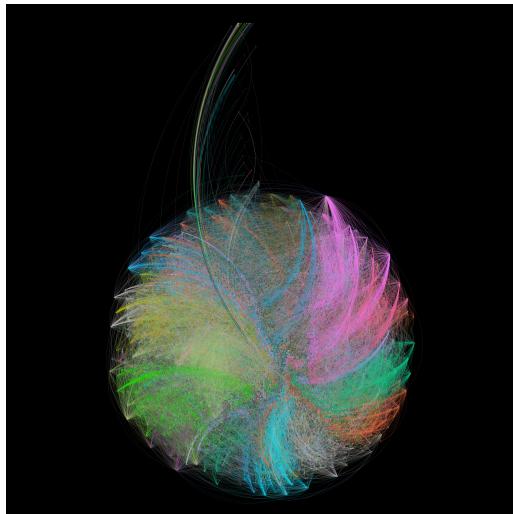


(a) Grafo pre-quarantena applicando Louvain.

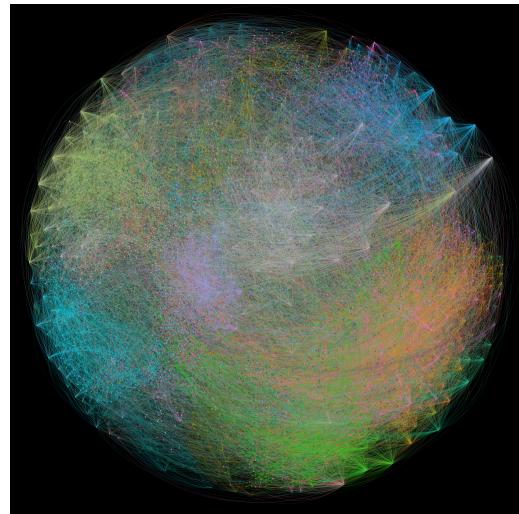


(b) Grafo post-quarantena applicando Louvain.

Figura 4.6: Grafo applicando Louvain con risoluzione 1.5

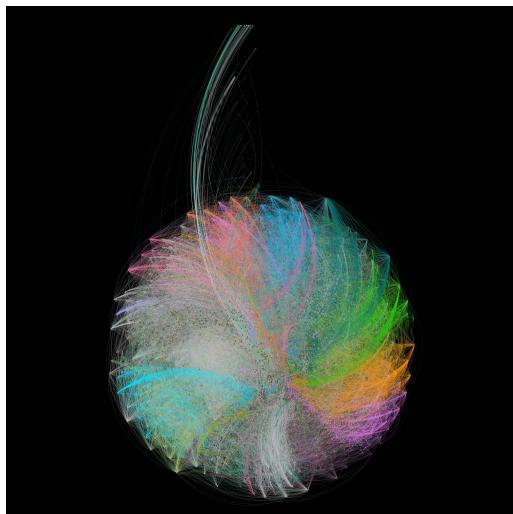


(a) Grafo pre-quarantena applicando Louvain.

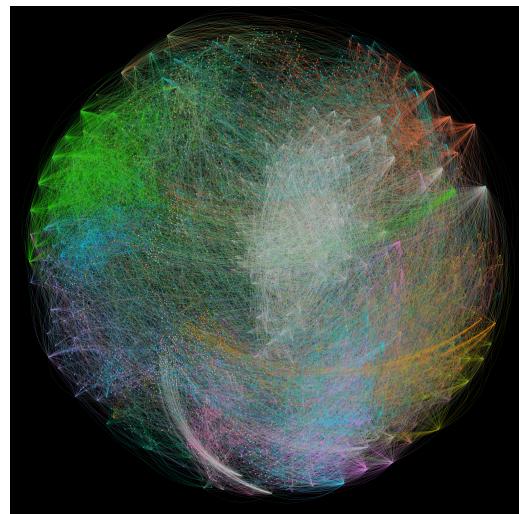


(b) Grafo post-quarantena applicando Louvain.

Figura 4.7: Grafo applicando Louvain con risoluzione 2



(a) Grafo pre-quarantena applicando Louvain.



(b) Grafo post-quarantena applicando Louvain.

Figura 4.8: Grafo applicando Louvain con risoluzione 3

Capitolo 5

Analisi numerica delle comunità

La libreria NetworkX [11] mette a disposizione delle funzionalità molto utili per la *network analysis*, ovvero ci consente di calcolare, per ciascuna comunità del grafo, delle metriche che la descrivono.

Le metriche utilizzate per la descrizione delle comunità individuate applicando gli algoritmi precedenti sono:

- **Nodes;**
- **Edges;**
- **Density:** la densità di un grafo è una misura che rappresenta la quantità di archi del grafo rispetto al numero massimo di archi possibili;
- **Assortativity:** ovvero la tendenza dei nodi di un grafo a collegarsi con nodi simili per quanto riguarda un determinato attributo;
- **Average clustering coefficient:** misura quanto i vicini di un nodo in una rete siano connessi tra loro;
- **Average Degree:** ovvero la media aritmetica dei gradi di ciascun nodo nella comunità;

Dato il gran numero di comunità individuate, al fine di migliorare la chiarezza dell’analisi, sono state filtrate le comunità con meno di dieci nodi.

5.1 Analisi per modularity

Le metriche calcolate per le comunità individuate con l’algoritmo modularity prima e dopo la quarantena sono riportate nelle tabelle seguenti. Inoltre, i risultati raccolti sono stati riportati sui grafici mediante la libreria Matplotlib [8] di Python.

ID	Nodes	Edges	Density	Assortativity	Avg clustering coeff	Avg degree
5	1040	7891	0.003987	0.284858	0.359874	7.5875
2	771	5809	0.005293	0.200178	0.390612	7.5344
1	740	5118	0.004389	0.277534	0.255002	6.9162
0	675	6433	0.008331	0.442701	0.509134	9.5304
36	659	5363	0.005922	0.250293	0.351253	8.1381
29	650	4729	0.006069	0.362727	0.395056	7.2754
38	608	5984	0.008882	0.426757	0.447506	9.8421
20	538	2910	0.005192	0.249666	0.180055	5.4089
18	460	3530	0.008222	0.335797	0.301869	7.6739
4	454	6569	0.019352	0.090888	0.416168	14.4692
8	400	2253	0.007657	0.134296	0.36171	5.6325
14	399	2729	0.009055	0.185055	0.346343	6.8396
22	383	3531	0.014654	0.092704	0.343131	9.2193
37	342	2279	0.009501	0.299006	0.298088	6.6637
23	341	2946	0.012144	0.170470	0.321444	8.6393
12	324	2145	0.010454	0.413862	0.292131	6.6203
25	305	1959	0.011195	0.439955	0.301974	6.423
16	277	1830	0.013185	0.288333	0.411055	6.6065
15	262	1668	0.011904	0.297013	0.316984	6.3664
17	246	1663	0.015364	0.253433	0.363733	6.7602
13	232	1579	0.013621	0.297446	0.289155	6.806
7	203	1195	0.014291	0.261317	0.190675	5.8867
10	164	901	0.018106	0.177438	0.432543	5.4939
41	118	769	0.025496	0.187442	0.311502	6.5169
3	106	539	0.027673	0.370631	0.321246	5.0849
39	49	307	0.056973	0.411276	0.301864	6.2653
48	35	272	0.095798	-0.160196	0.526693	7.7714
6	18	76	0.150327	0.140187	0.330688	4.2222

Tabella 5.1: Metriche delle comunità pre-quarantena. La comunità evidenziata è quella che si è distinta tra le altre per il valore delle sue metriche. La 9 non è stata riportata nella tabella in quanto aveva un numero di partecipanti inferiore a 10.

ID	Nodes	Edges	Density	Assortativity	Avg clustering coeff	Avg degree
22	1246	13865	0.004729	0.649846	0.511843	11.1276
21	828	12298	0.00949	0.617376	0.598246	14.8527
27	619	7222	0.009259	0.645648	0.623784	11.6672
5	559	5976	0.007451	0.515471	0.315108	10.6905
25	535	24694	0.064496	0.733111	0.770092	46.157
6	412	6763	0.020185	0.641829	0.664137	16.415
11	401	4343	0.011259	0.566245	0.532994	10.8304
20	390	4058	0.015306	0.730452	0.697402	10.4051
13	312	3037	0.0139129	0.583464	0.516365	9.734
23	302	4070	0.024114	0.568523	0.508468	13.4768
7	202	2779	0.032117	0.576515	0.630928	13.7574
2	189	1495	0.015873	0.508823	0.315189	7.9101
29	166	1352	0.02322	0.572531	0.496325	8.1446
33	145	980	0.023084	0.630581	0.471806	6.7586
8	133	1279	0.030759	0.573232	0.630307	9.6166
32	120	1340	0.030812	0.456904	0.382849	11.1667
28	58	471	0.049607	0.464064	0.362247	8.1207
9	38	2886	0.634423	0.496502	0.805273	75.9474
18	34	566	0.13369	0.389773	0.368167	16.6471
14	11	79	0.290909	0.552597	0.387879	7.1818

Tabella 5.2: Metriche delle comunità post-quarantena. Le comunità evidenziate, la 25 e la 9, sono quelle che si sono distinte particolarmente per il valore delle loro metriche.

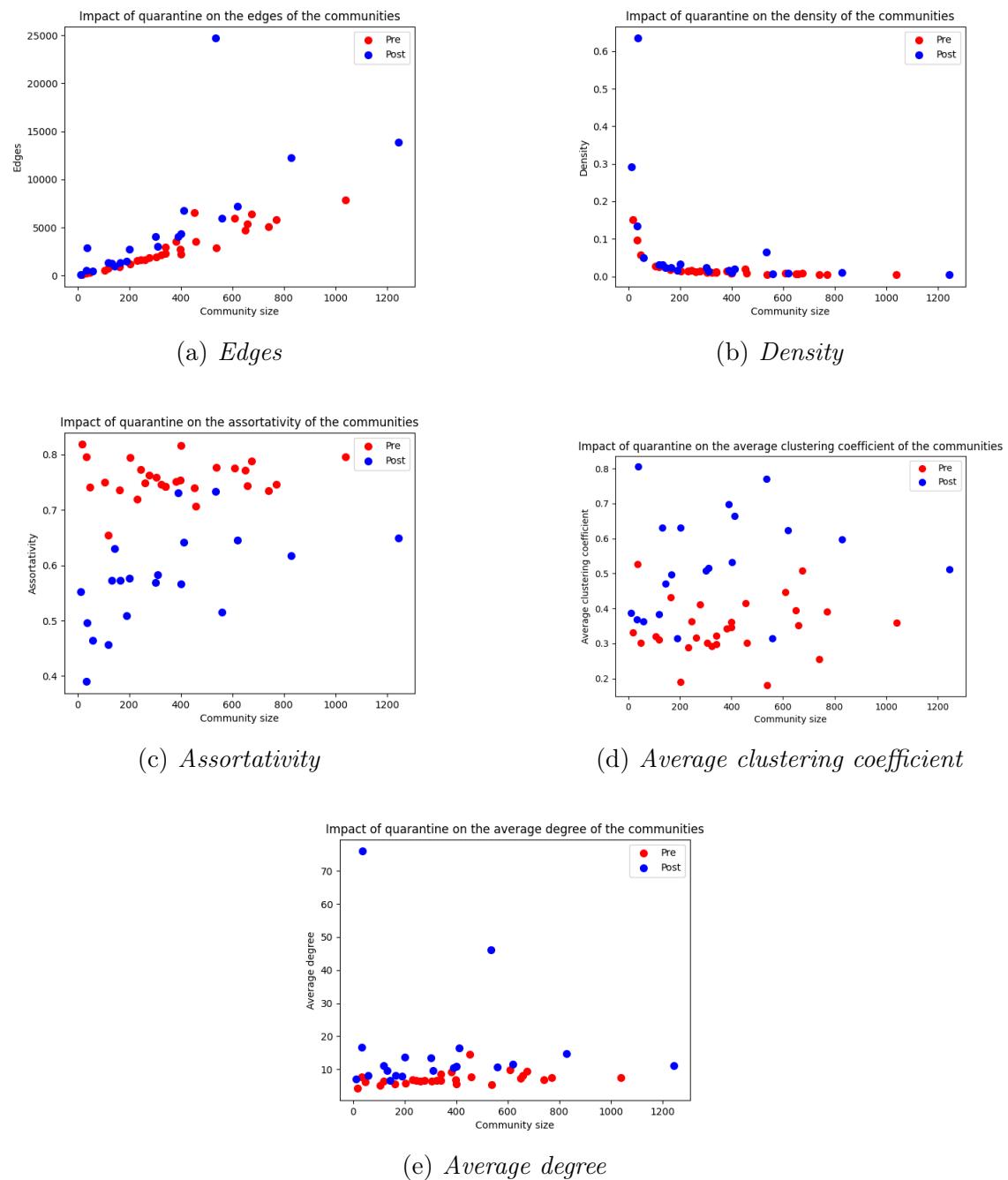


Figura 5.1: Metriche delle comunità trovate con modularità

Analizzando i valori delle metriche, possiamo dedurre le seguenti informazioni:

- Tabella 5.1 e Tabella 5.2: il numero dei nodi è diminuito significativamente in 14 comunità, mentre per altre 11 è aumentato. Solo due comunità sono rimaste quasi invariate. Il numero totale dei nodi nel grafo dopo la quarantena risulta essere circa 3/4 del numero di nodi prima dell'intervento.
- 5.1(a): il numero di archi cresce più o meno linearmente con il numero di nodi appartenenti alla comunità, eccezion fatta per la comunità 25, nella quale il numero di archi dopo la quarantena è dieci volte maggiore rispetto a quello prima dell'intervento. Quest'ultima informazione indica un aumento significativo dell'interconnessione tra i membri della comunità. Anche la comunità 9 si dimostra molto interconnessa, anche se potrebbe essere dovuto al numero limitato di partecipanti all'interno della stessa.
- 5.1(b): la quarantena non ha avuto impatto sulla densità delle comunità, ovvero le interazioni tra i membri delle comunità sono limitate e poco frequenti sia prima che dopo l'intervento. Anche in questo caso si distingue la comunità 25, che si dimostra fortemente interconnessa.
- 5.1(c): mentre, prima dell'intervento, i membri delle diverse comunità sembravano interagire prevalentemente tra di loro - alta assortatività -, dopo possiamo vedere che questa tendenza diminuisce significativamente; quindi, tranne per una lieve preferenza nello stabilire connessioni con membri appartenenti alla propria comunità, l'assortatività dimostra un andamento abbastanza neutrale. Si distinguono la comunità 20 e 25 che dopo l'intervento continuano ad avere una forte tendenza ad interagire principalmente con i membri appartenenti alla loro comunità.
- 5.1(d): il coefficiente di clustering medio è globalmente aumentato rispetto al suo valore prima della quarantena. Ciò può significare che l'intervento ha portato ad una maggiore tendenza a formare gruppi interconnessi da parte dei nodi.
- 5.1(e): il grado medio delle comunità ha subito un lieve aumento dopo l'intervento, ovvero sono leggermente aumentate le interazioni tra i membri di queste ultime. Inoltre, la quarantena non ha avuto effetto in termini di diminuzione o aumento di connessioni tra le diverse comunità. Anche per questa metrica, si distinguono le comunità 9 e 25 per la loro cooperazione intensiva. L'aumento del grado medio della seconda comunità potrebbe essere dovuto, non solo all'aumento delle interazioni nella comunità, ma anche all'aumento significativo del numero dei membri.

In conclusione, la quarantena ha portato ad una diminuzione degli utenti, ad un forte aumento della coesione tra i membri delle comunità 9 e 25 e un leggero aumento per quanto riguarda la coesione nelle altre comunità.

5.2 Analisi per Louvain

Per rendere più chiara la visualizzazione dei dati, sono stati riportati due grafici per ogni metrica: uno che riguarda i dati raccolti prima della quarantena e uno per quelli dopo l'intervento.

I valori visti in Tabella 5.1 sono rappresentati in questi grafici con delle "x" per poterli distinguere dai dati ottenuti applicando l'algoritmo di Louvain.

Analizzando le metriche delle comunità identificate tramite l'algoritmo di Louvain prima e dopo l'intervento di moderazione, possiamo rilevare le stesse informazioni esposte nella sezione 5.1.

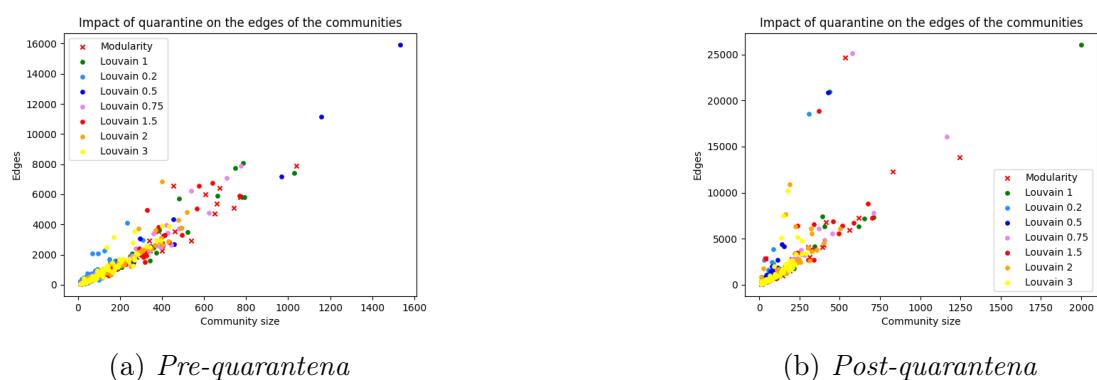


Figura 5.2: Numero degli archi delle comunità trovate con Louvain

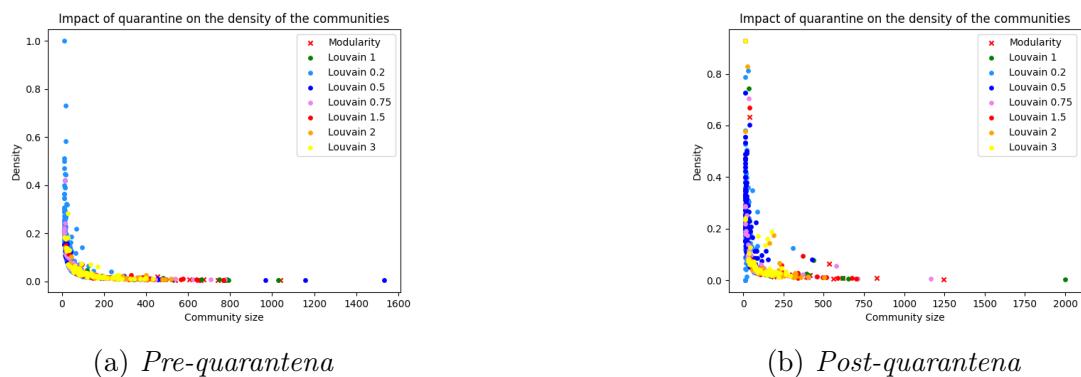


Figura 5.3: Densità delle comunità trovate con Louvain

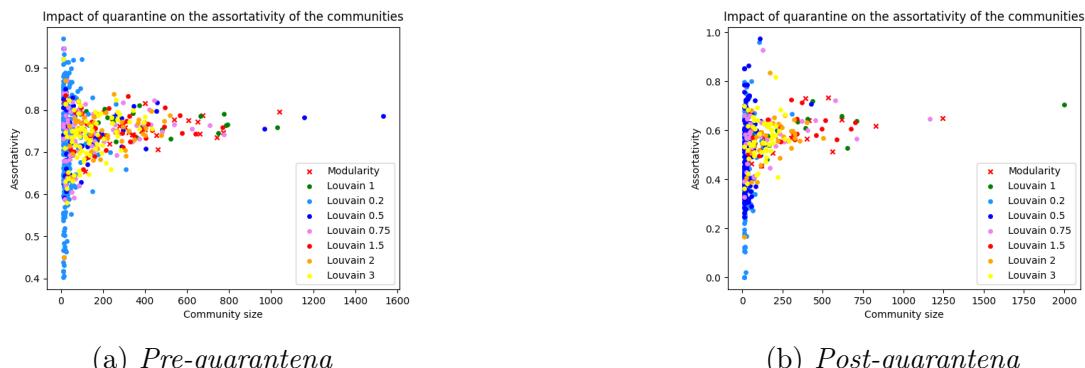


Figura 5.4: Assortatività delle comunità trovate con Louvain

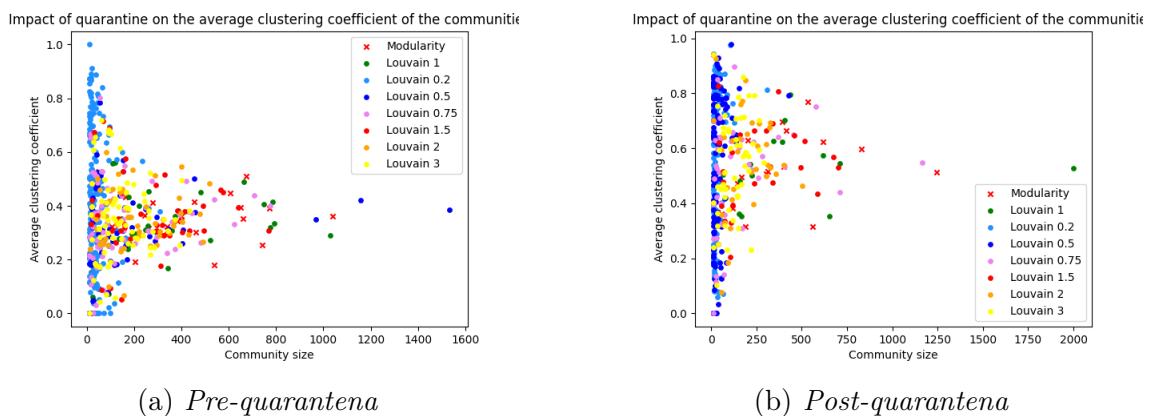


Figura 5.5: Coefficiente di clustering medio delle comunità trovate con Louvain

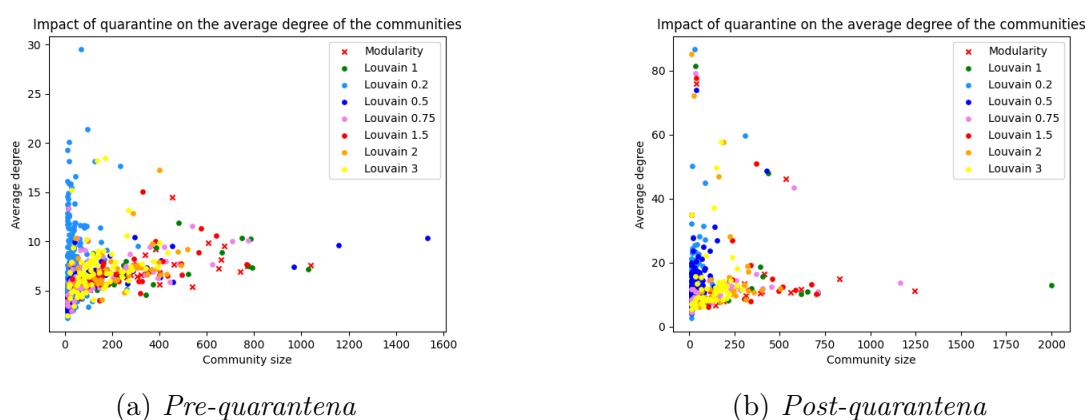


Figura 5.6: Grado medio delle comunità trovate con Louvain

Capitolo 6

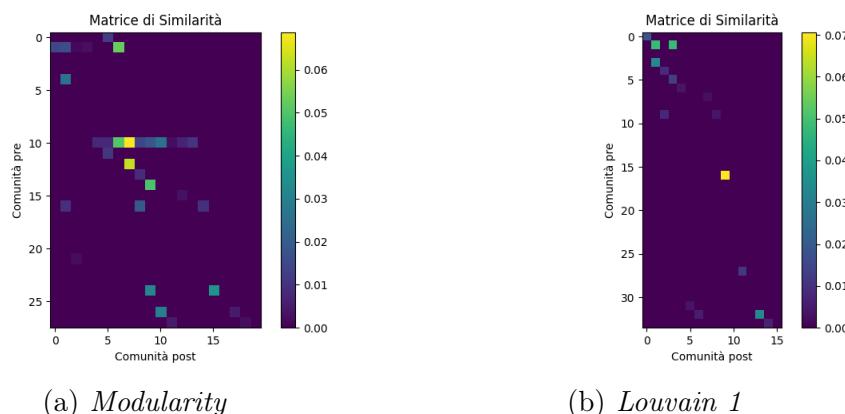
Calcolo della similarità

In questa sezione dello studio, vengono analizzate delle matrici di similarità al fine di esaminare l'effetto della quarantena sui membri delle comunità. Nello specifico, viene utilizzato come criterio di similarità il numero di nodi che appartengono alla stessa comunità sia prima che dopo l'attuazione della moderazione. In altre parole, l'obiettivo è quello di misurare quanto le comunità abbiano mantenuto o modificato la loro composizione di membri a seguito della quarantena.

6.1 Matrici di similarità normalizzate

Per mettere maggiormente in risalto le distinzioni tra le varie celle della matrice, queste ultime sono state normalizzate. In pratica, ogni cella è stata divisa per il numero totale di nodi presenti nella comunità della riga corrispondente.

Questo processo di normalizzazione è stato implementato per offrire una maggiore chiarezza sulle proporzioni e sulle relazioni intrinseche tra i dati presenti nella matrice.



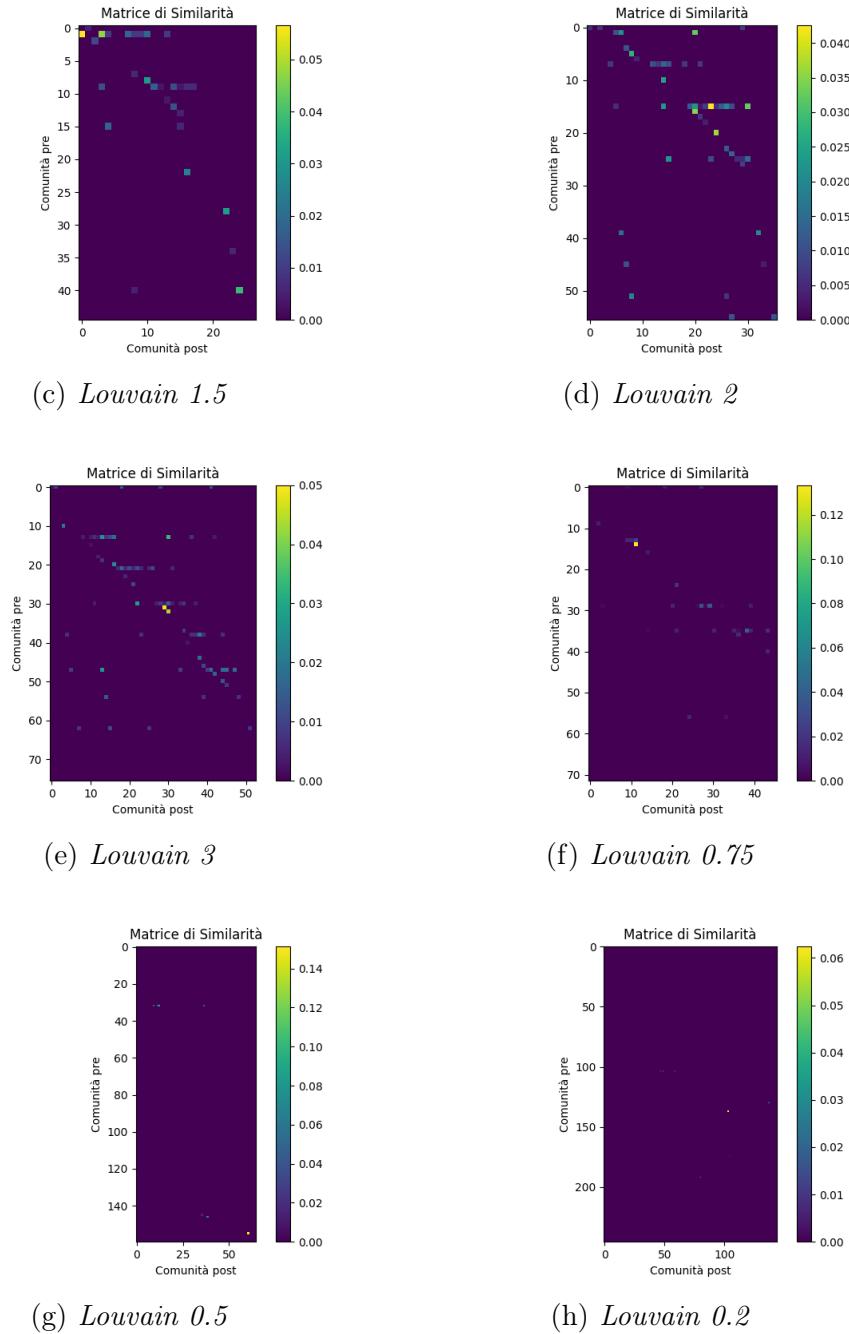


Figura 6.1: Matrici di similarità normalizzate

6.2 Matrici di similarità diagonalizzate

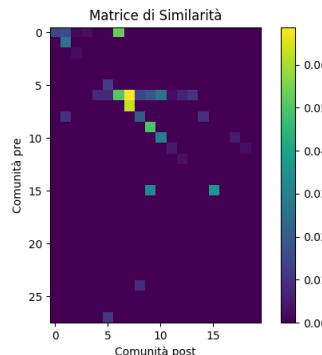
Dopo aver ottenuto la matrice normalizzata, si procede all'applicazione dell'*algoritmo ungherese* per ottenere una matrice diagonalizzata tramite la libreria Munkres [10] di Python.

Volendo applicare un algoritmo di minimizzazione, è necessario adottare una specifica procedura per preparare la matrice di input:

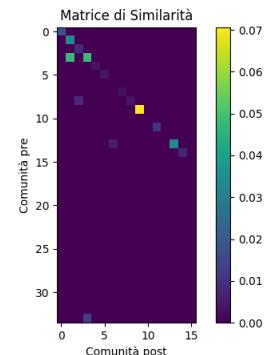
1. complementare i valori della matrice;
2. aggiungere un *padding* di celle che contengono '1' per rendere la matrice quadrata;

Una volta completata l'applicazione dell'algoritmo ungherese, è importante ripristinare la matrice alle sue condizioni originali. Questo processo avviene in due passaggi:

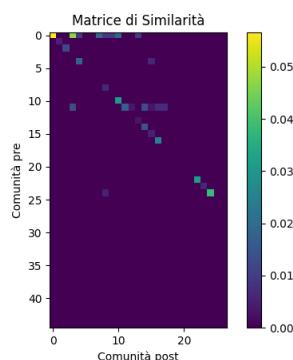
1. rimuovere il *padding* aggiunto in precedenza;
2. complementare nuovamente i valori all'interno della matrice.



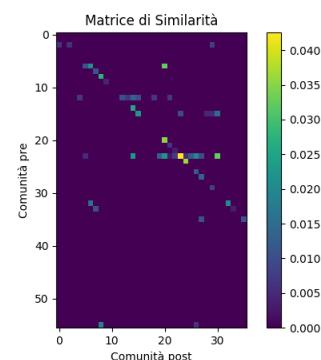
(a) *Modularity*



(b) *Louvain 1*



(c) *Louvain 1.5*



(d) *Louvain 2*

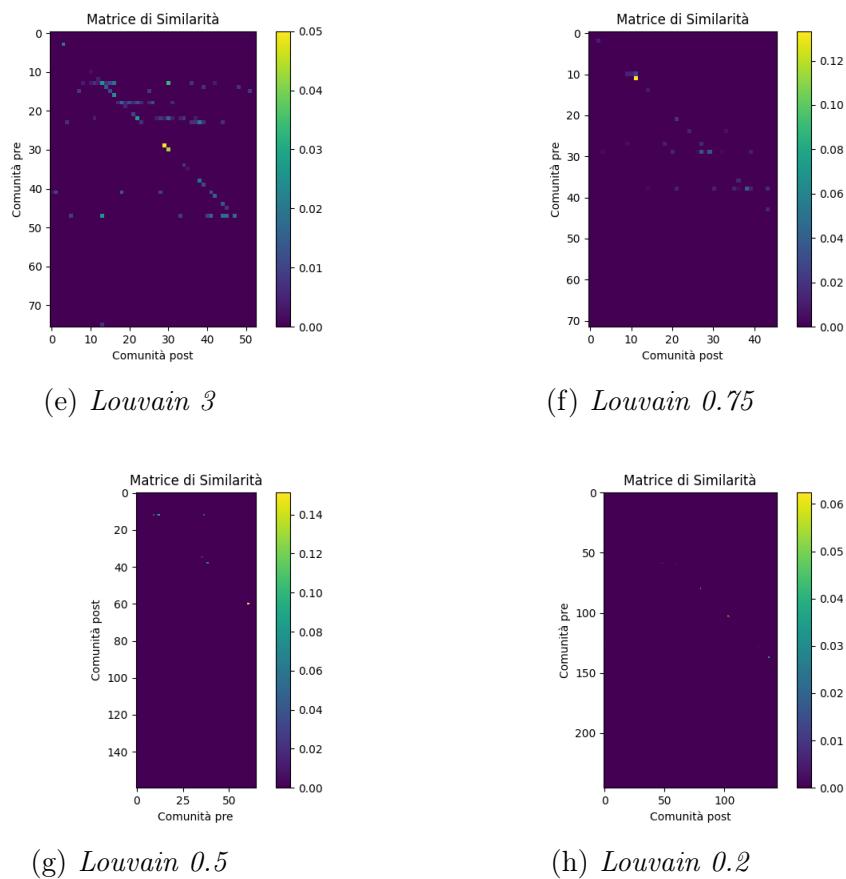


Figura 6.2: Matrici di similarità diagonalizzate

6.3 Coefficiente di similarità

Nel fase finale della nostra analisi, abbiamo cercato di misurare quanto due comunità fossero simili prima e dopo l'intervento, utilizzando un coefficiente di similarità come indicatore, per ogni algoritmo di individuazione delle comunità utilizzato precedentemente.

Questo coefficiente è stato ottenuto dal rapporto tra la somma delle celle sulla diagonale principale della matrice e la somma di tutte le celle presenti nella matrice. Quanto più questo valore è alto, tanto maggiore è la similarità tra le comunità individuate nelle due fasi dell'analisi.

In altre parole, il coefficiente di similarità ci aiuta a valutare quanto le comunità abbiano mantenuto la loro struttura interna prima e dopo l'intervento.

Algorithm	Resolution	Similarity
Modularity	1	0.496957
Louvain	1	0.765831
Louvain	2	0.472002
Louvain	3	0.440759
Louvain	1.5	0.598689
Louvain	0.2	0.788706
Louvain	0.5	0.700952
Louvain	0.75	0.613275

Tabella 6.1: Coefficienti di similarità

I coefficienti di similarità ottenuti per la matrice di similarità tra le comunità identificate tramite l'algoritmo di Louvain con risoluzioni di 0,2 e 0,5 potrebbero non essere significativi, poiché i valori sulla diagonale non differiscono in modo sostanziale da quelli nel resto della matrice. Questa situazione si verifica a causa dell'abbondanza di comunità identificate a queste risoluzioni, ciascuna composta da un numero limitato di partecipanti. Di conseguenza, è improbabile che, durante l'applicazione dell'algoritmo dopo la quarantena, le comunità identificate abbiano mantenuto gli stessi membri presenti prima dell'intervento.

In conclusione, oltre alla riduzione numerica, le comunità hanno anche subito cambiamenti nella loro composizione dopo la quarantena.

Capitolo 7

Conclusioni

Gli strumenti di moderazione sono fondamentali per mantenere un ambiente online sano e rispettoso, preservare la qualità dei contenuti e promuovere discussioni costruttive. Partendo da questa premessa, lo scopo principale di questo studio è stato condurre una valutazione approfondita degli eventuali impatti secondari che tali strumenti possono avere sulle dinamiche delle relazioni tra gli utenti e sul loro comportamento nell'utilizzo della piattaforma online.

Una volta estratti due grafi sulla base delle interazioni tra gli utenti appartenenti al subreddit *The_Donald* in una finestra temporale intorno all'intervento attuato il 26/06/2019, sono stati utilizzati due algoritmi di *community detection*: *Modularity* e *Louvain*. Per quest'ultimo è stata fatta un'analisi più approfondita al variare della risoluzione applicata. Per entrambi gli algoritmi sono stati utilizzati degli strumenti di *Network Analysis* per descrivere le proprietà delle diverse comunità individuate e l'effetto che la quarantena ha avuto su queste ultime. In particolare si è riscontrato come la quarantena abbia ottenuto effetti misti: se da un lato ha portato ad una riduzione degli utenti del subreddit, non ha intaccato la loro coesione.

Infine, abbiamo misurato la similarità tra le comunità prima e dopo l'intervento mediante l'uso di matrici di similarità e il calcolo dei *coefficienti di similarità*. Questi coefficienti ci forniscono un'indicazione chiara di come l'intervento di moderazione abbia influenzato la struttura interna delle comunità identificate attraverso i vari algoritmi.

Come da aspettative, non è stato possibile realizzare un'analisi multidimensionale sugli effetti dell'intervento all'interno del subreddit o sulla piattaforma in generale. È infatti ragionevole sostenere che questo lavoro presenti ancora ampio margine di miglioramento. A questo fine, si potrebbe utilizzare un criterio alternativo per definire l'interazione tra i diversi utenti o valutare gli effetti a lungo termine dell'intervento.

In futuro, sarà cruciale approfondire e migliorare gli strumenti di moderazione online in modo da poter raggiungere gli obiettivi desiderati con minimi effetti collaterali.

Bibliografia

- [1] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The pushshift reddit dataset. <https://files.pushshift.io/reddit/comments/>.
- [2] C. Cao and J. Caverlee. Organic or organized? exploring url sharing behavior. Reperibile presso: <https://dl.acm.org/doi/10.1145/2806416.2806572>.
- [3] E. Chandrasekharan, S. Jhaver, A. Bruckman, and E. Gilbert. Quarantined! examining the effects of a community-wide moderation intervention on reddit. Reperibile presso: <https://doi.org/10.48550/arXiv.2009.11483>.
- [4] E. Chandrasekharan, U. Pavalanathan, A. Srinivasan, A. Glynn, J. Eisenstein, and E. Gilbert. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. Reperibile presso: <https://dl.acm.org/doi/10.1145/3134666>.
- [5] Gephi. The leading visualization and exploration software for graphs and networks. <https://gephi.org/>.
- [6] S. Jhaver, C. Boylston, D. Yang, and A. Bruckman. Evaluating the effectiveness of deplatforming as a moderation strategy on twitter. Reperibile presso: <https://dl.acm.org/doi/10.1145/3479525>.
- [7] N. Louvain. https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.louvain.louvain_communities.html.
- [8] Matplotlib. <https://matplotlib.org/>.
- [9] N. Modularity. https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.modularity_max.greedy_modularity_communities.html.
- [10] Munkres. <https://pypi.org/project/munkres/>.
- [11] NetworkX. <https://networkx.org/>.
- [12] L. Nizzoli and S. Tardelli. Coordinated behavior on social media in 2019 uk general election. Reperibile presso: <https://ojs.aaai.org/index.php/ICWSM/article/view/18074>.

- [13] M. H. Ribeiro, S. Jhaver, S. Zannettou, J. Blackburn, G. Stringhini, E. D. Cristofaro, and R. West. Do platform migrations compromise content moderation? evidence from r/the_donald and r/incels. Reperibile presso: <https://doi.org/10.48550/arXiv.2010.10397>.
- [14] H. M. Saleem and D. Ruths. The aftermath of disbanding an online hateful community. Reperibile presso: <https://doi.org/10.48550/arXiv.1804.07354>.
- [15] A. Trujillo and S. Cresci. Make reddit great again- assessing community effects of moderation interventions on rthe_donald. <https://doi.org/10.48550/arXiv.2201.06455>.
- [16] A. Trujillo and S. Cresci. One of many- assessing user-level effects of moderation interventions on rthe_donald. <https://doi.org/10.48550/arXiv.2209.08809>.
- [17] D. Weber and F. Neumann. Amplifying influence through coordinated behaviour in social networks. Reperibile presso: <https://link.springer.com/article/10.1007/s13278-021-00815-2>.