



**Università Politecnica delle Marche**

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Ingegneria Informatica e dell'Automazione

---

## **Social Newtork Analysis**

---

Corso di Data Science

Professore

Prof. Domenico Ursino

Gruppo 1

Michele Pasqualini

Denil Nicolosi

Eris Prifti

Anno accademico 2021-2022

# Indice

<b>1</b>	<b>Dataset</b>	<b>2</b>
1.1	Struttura del dataset . . . . .	2
<b>2</b>	<b>Analisi preliminare e data visualization</b>	<b>2</b>
2.1	Canali partner di Twitch . . . . .	2
2.2	Canali più visti . . . . .	3
<b>3</b>	<b>Descrizione del grafo</b>	<b>4</b>
<b>4</b>	<b>Centralità</b>	<b>5</b>
4.1	Degree centrality . . . . .	5
4.2	Betweenness centrality . . . . .	8
4.3	Closeness Centrality . . . . .	11
4.4	Eigenvector Centrality . . . . .	14
4.5	Intersezione delle diverse centralità . . . . .	17
<b>5</b>	<b>Clique</b>	<b>18</b>
<b>6</b>	<b>Community</b>	<b>20</b>

# 1 Dataset

Il dataset utilizzato è realizzato a partire dai dati raccolti dalla piattaforma Twitch relativi al Portogallo. Twitch è una piattaforma di livestreaming di proprietà di Amazon.com, consentendo principalmente lo streaming di videogiochi. I nodi sono gli utenti stessi e i collegamenti sono amicizie reciproche tra di loro. Le funzionalità dei vertici vengono estratte in base ai giochi giocati e apprezzati, alla posizione e alle abitudini di streaming. Questi social network sono stati raccolti nel maggio 2018 e le relazioni tra i vari streamer e gli spettatori creano un social network, ideale per l'analisi basata su grafi.

## 1.1 Struttura del dataset

Il dataset preso in considerazione è suddiviso in due file, uno per i dati relativi ai nodi e uno per i dati degli archi del grafo. Tutti i file presenti nel dataset sono in formato CSV.

<b>Id</b>	Id del canale Twitch.
<b>Days</b>	Numero di giorni da quando il canale è attivo.
<b>Mature</b>	True or False.
<b>Views</b>	Numero di visualizzazioni ricevute dal canale.
<b>Partner</b>	True or False, indica se il canale è affiliato alla piattaforma, cioè se può monetizzare oppure no.
<b>New id</b>	Nuovo id del canale Twitch.

Tabella 1: Descrizione campi del dataset - nodi

<b>From</b>	Id del nodo di partenza.
<b>To</b>	Id del nodo di arrivo.

Tabella 2: Descrizione campi del dataset - archi

# 2 Analisi preliminare e data visualization

Prima di iniziare con l'analisi della rete, è stata svolta un'analisi preliminare dei dati utile ad approfondire la tematica. Nel dettaglio, vengono riportati alcuni grafici sui canali affiliati e su quelli con maggiori visualizzazioni.

## 2.1 Canali partner di Twitch

Nel grafico a torta in figura 1 viene mostrata la divisione tra i canali partner e non partner di Twitch Portogallo. Lo stato di partner Twitch è riservato a quegli autori di contenuti con performance elevate che prendono seriamente la trasmissione e hanno già creato un pubblico folto. Nel dataset preso in considerazione, solamente il 14.6% dei canali è partner, mentre l'85.4% risulta non essere partner. Questo ci dice che, la maggior parte dei canali non ottiene monetizzazioni dalla piattaforma. In questa porzione del grafico rientrano maggiormente quei canali che fungono da spettatori, cioè utenti non affiliati alla piattaforma che supportano content creator più affermati.

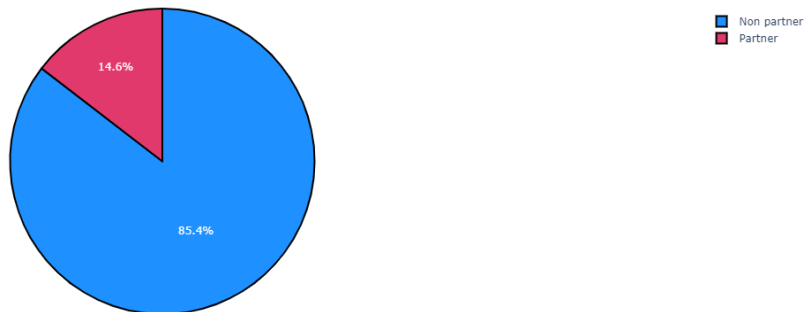


Figura 1: Divisione tra canali partner e non partner

## 2.2 Canali più visti

Nell'istogramma in figura 2 vengono mostrati i primi 20 canali di Twitch Portogallo con maggiori visualizzazioni. A causa del dataset, nell'asse delle ascisse vengono riportati solamente gli id dei rispettivi canali che corrispondono ai nodi del nostro grafo, senza l'ausilio di attributi descrittivi, come ad esempio il nome del canale, in quanto non presenti. E' possibile notare che il canale più visto riceve circa 120.000.000 milioni di visualizzazioni ed è altamente sopra la media. Infatti, il secondo e il terzo canale Twitch ricevono entrambi 40.000.000 visualizzazioni, un numero di gran lunga inferiori alle views ricevute dal primo. Questo indica che il nodo identificato dall'id 290 è abbastanza rilevante, in quanto potrebbe essere un content creator molto affermato nel Portogallo.

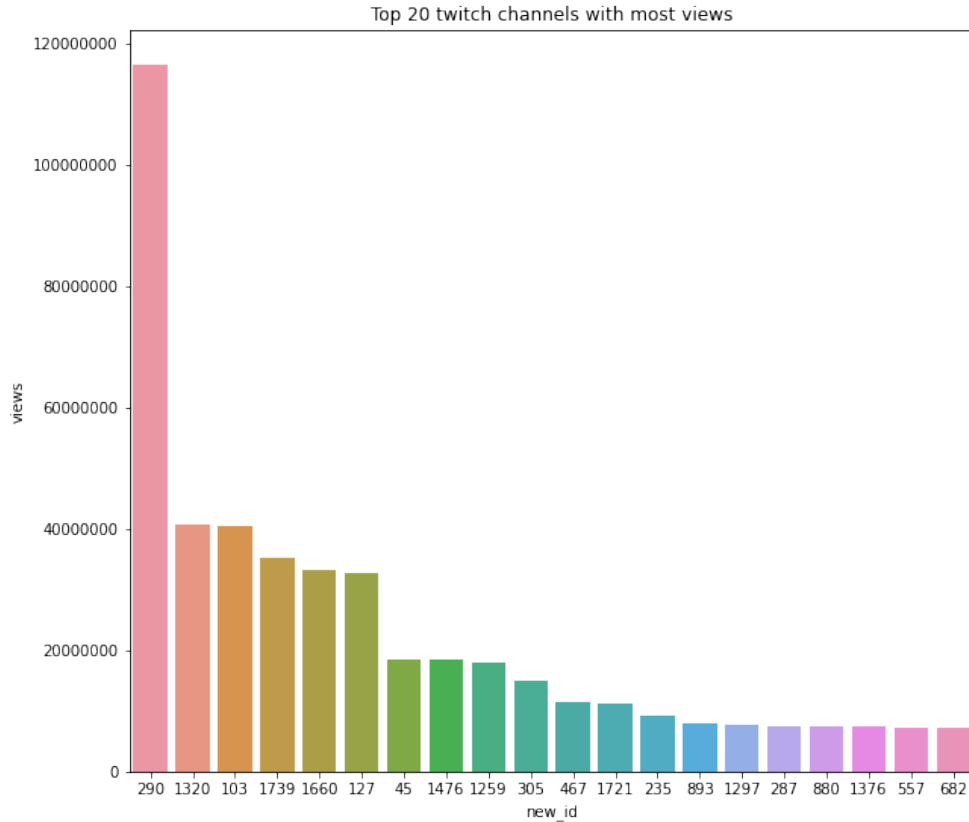


Figura 2: Canali Twitch con maggiori visualizzazioni

### 3 Descrizione del grafo

La figura 3 mostra il grafo preso in considerazione per il progetto di social network analysis. I nodi del grafo rappresentano vari canali Twitch. Gli archi indicano dei legami tra i diversi canali. Il grafo è caratterizzato da una serie di informazioni aggiuntive sotto forma di attributi sui nodi. In prima fase di analisi si sono ottenuti delle informazioni generali sul grafo, quali il numero dei nodi e degli archi, il raggio e il diametro. In particolare:

- I nodi del grafo sono in totale 1912;
- Gli archi del grafo sono in totale 31299;
- La densità del grafo, considerando unicamente la parte connessa, è pari a 0.0171.
- Il clustering del grafo è pari a 0.3198;
- Il raggio della parte connessa del grafo è pari a 4.
- Il diametro della parte connessa del grafo è pari a 7.
- La periferia della parte connessa del grafo, cioè l'insieme dei nodi con eccentricità massima, presenta 2 nodi, mentre il centro del grafo, cioè l'insieme dei nodi con eccentricità minima è composto da 285 nodi.

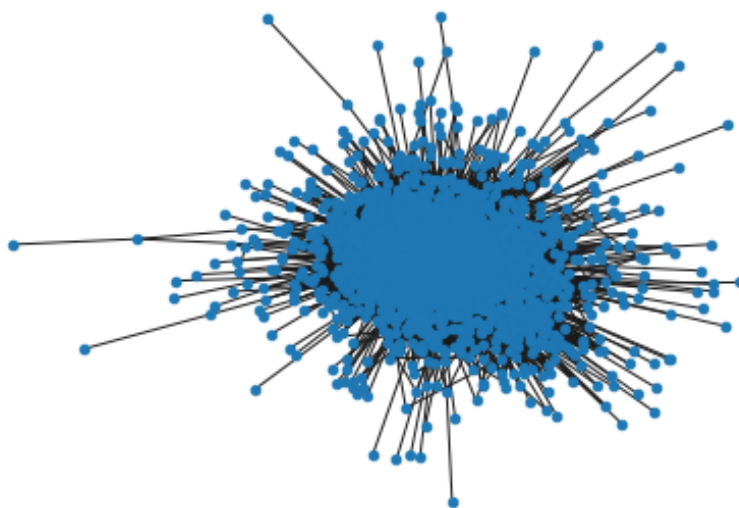


Figura 3: Grafo preso in considerazione per l'analisi

## 4 Centralità

Una delle prime analisi in una social network consiste nel misurare il potere o l'influenza dei componenti, sulla base delle loro connessioni. Per questo il concetto di centralità può essere importante in quanto consente di identificare le diverse aree di una social network e il loro grado di coesione, giudicare la rilevanza dei nodi o aree che appartengono ad una rete e attribuire una misura di distanza tra nodi o aree della rete. Esistono, pertanto, diverse soluzioni, conosciute come «misure di centralità».

### 4.1 Degree centrality

La prima metrica utilizzata è la degree centrality. Essa viene calcolata andando a contare il numero di legami che un nodo ha con gli altri nodi della rete. Tale metrica esprime il grado di connettività di un certo utente con gli altri individui, permettendo di capire se il nodo considerato ha una posizione strutturale rilevante nella rete di cui fa parte. Quindi, maggiore è il valore di degree centrality e maggiori saranno le connessioni di quel nodo con altri nodi. Questo significa che un nodo è importante quando ha più connessioni. Grazie alla funzione `nx.degree_centrality()` messa a disposizione dalla libreria NetworkX, è stato possibile calcolare la degree centrality per ciascun nodo della rete. Successivamente, i valori ottenuti sono stati ordinati in modo tale da avere i primi 10 nodi con valore di degree centrality più alto e i 10 con valore più basso. I risultati ottenuti vengono mostrati di seguito nella figura 4.

**Best 10 Degree Centrality:**

1. 127: 0.4013605442176871
2. 1476: 0.3129251700680272
3. 290: 0.3087388801674516
4. 1297: 0.30716902145473574
5. 467: 0.304552590266876
6. 1660: 0.24856096284667714
7. 67: 0.23757195185766616
8. 1320: 0.217687074829932
9. 1758: 0.20617477760334904
10. 1259: 0.20146520146520147

**Worst 10 Degree Centrality:**

1. 532: 0.0005232862375719519
2. 1828: 0.0005232862375719519
3. 1672: 0.0005232862375719519
4. 1468: 0.0005232862375719519
5. 1857: 0.0005232862375719519
6. 1116: 0.0005232862375719519
7. 1564: 0.0005232862375719519
8. 1519: 0.0005232862375719519
9. 1343: 0.0005232862375719519
10. 1658: 0.0005232862375719519

Figura 4: Tabella della classifica della degree centrality

Come è possibile vedere dalla figura 5, il canale che ha una maggiore degree centrality è il canale identificato dall'id 127, seguito subito dopo dai canali 1476 e 290. Questo fenomeno viene confermato da ciò che è stato osservato nell'analisi preliminare, in quanto i canali con maggior degree centrality sono gli stessi che hanno ricevuto un gran numero di visualizzazioni. Infatti, il canale con id 290 è il canale Twitch più visto, mentre il canale 127 riceve circa 40 milioni di views. Infine, gli ultimi 10 valori di degree centrality sono nulli e quindi potrebbero essere relativi a canali Twitch molto piccoli, non affermati e che non sono partner della piattaforma streaming.

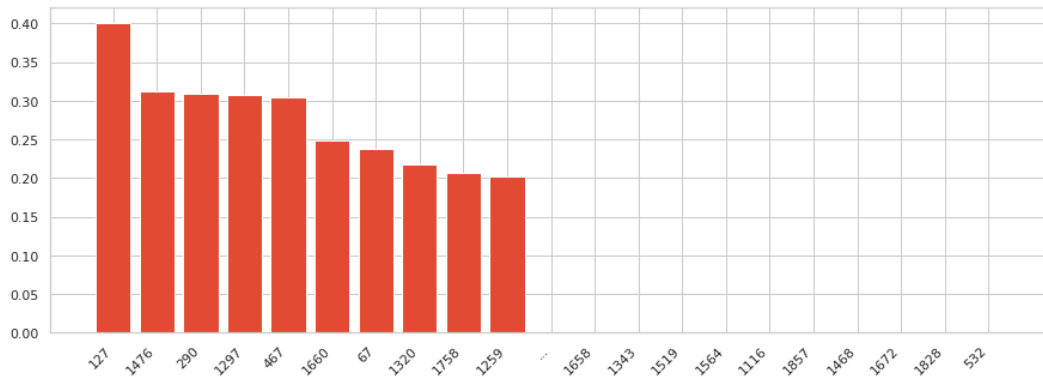


Figura 5: Grafico della classifica della degree centrality

Come mostrato in figura 6 sono molto pochi i nodi ad avere una elevata degree centrality, mentre sono molti i nodi con una degree centrality compresa tra lo 0 e 0.05.

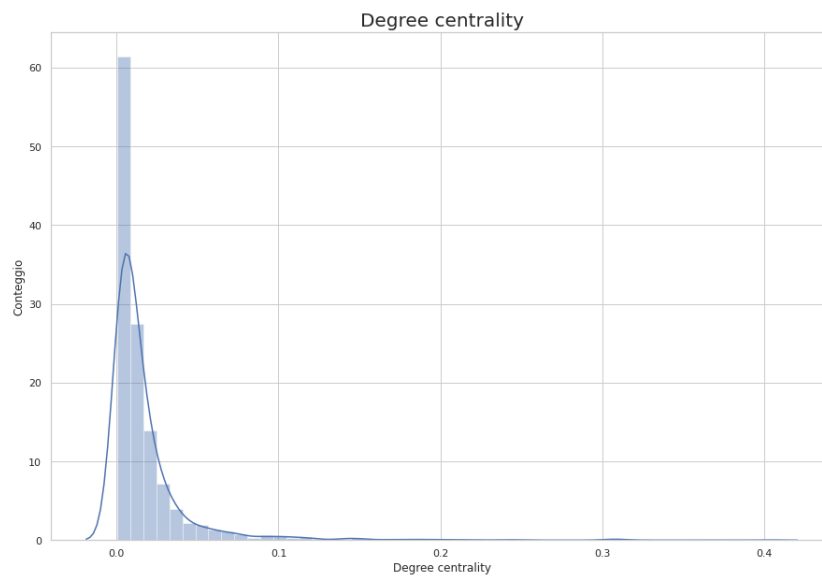


Figura 6: Grafico della distribuzione della degree centrality

Lo spring layout in figura 7 permette di osservare come vi sia una netta concentrazione di nodi nella parte centrale della rete, che rende difficilmente distinguibili i nodi al suo interno. Inoltre all'interno del grafo si nota una sfumatura più chiara che mostra in quale punto vi è un aumento della centrality.



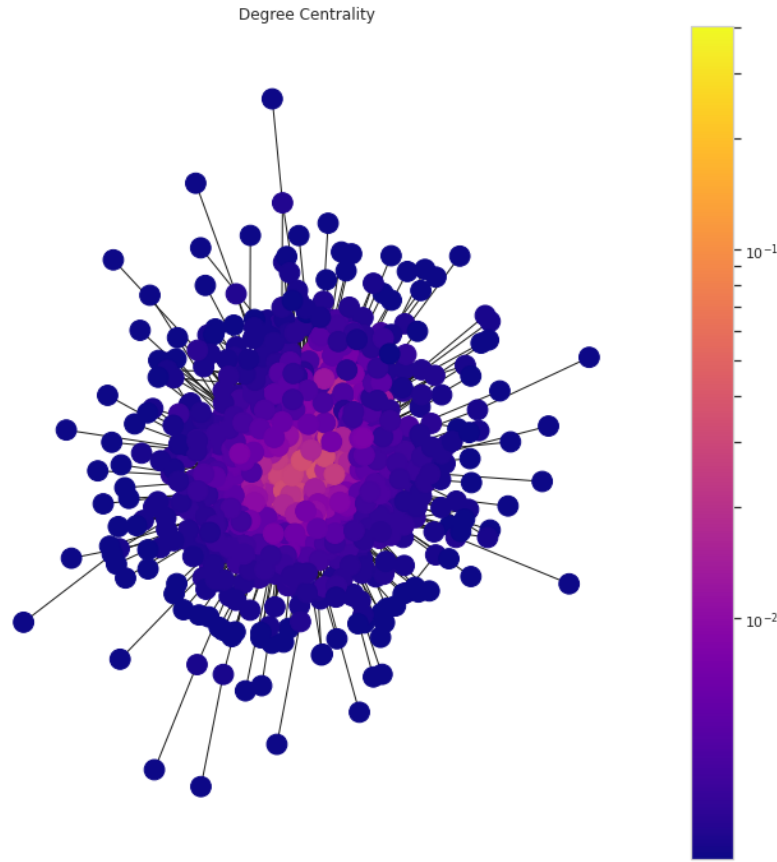


Figura 7: Spring layout del grafo della degree centrality

## 4.2 Betweenness centrality

La seconda metrica calcolata è la betweenness centrality. Essa è legata al numero di volte in cui un nodo si ritrova lungo lo shortest path tra le altre coppie di nodi della rete, descrivendo così la capacità di diffusione delle informazioni nella rete attraverso quel nodo. Un nodo con un'elevata betweenness centrality ha una grande influenza nel flusso di informazioni. La Betweenness centrality ha anche un altro ruolo: essa è in grado di identificare gli individui che agiscono da bridge tra due o più comunità che altrimenti non riuscirebbero a comunicare l'una con l'altra. L'algoritmo per il calcolo della betweenness centrality (piuttosto costoso) è il seguente:

- Si calcolano gli shortest path tra ciascuna coppia di nodi utilizzando l'Algoritmo di Dijkstra.
- Per ciascun nodo  $I$  della rete:
  - Si conta il numero di shortest path in cui  $I$  è coinvolto.
  - Si normalizza tale numero in modo da portare il risultato tra 0 e 1.

Quindi, ogni nodo del grafo assume dei valori di betweenness centrality compresi da 0 a 1: maggiore è tale valore e maggiore sarà la posizione di potere assunta dal nodo all'interno della rete. Questo significa che la maggior parte delle comunicazioni che avvengono tra elementi delle diverse sottoreti, dovranno passare proprio per il nodo in questione. Anche in questo caso, per calcolare la betweenness centrality è stata utilizzata la funzione `nx.betweenness centrality()` di NetworkX. I risultati ottenuti vengono riportati nella figura 8.

**Best 10 Betweenness Centrality:**

```
1. 127: 0.0992613670254655
2. 1476: 0.054894872658417665
3. 1297: 0.05042095543765276
4. 290: 0.05011201385679391
5. 467: 0.04382473622801392
6. 67: 0.034425706705748176
7. 1660: 0.02624664025473586
8. 1259: 0.023294665516722747
9. 287: 0.022120764244835684
10. 428: 0.021980008718637752
```

**Worst 10 Betweenness Centrality:**

```
1. 532: 0.0
2. 1828: 0.0
3. 1672: 0.0
4. 1468: 0.0
5. 1857: 0.0
6. 1116: 0.0
7. 1771: 0.0
8. 1564: 0.0
9. 1519: 0.0
10. 1343: 0.0
```

Figura 8: Tabella della classifica della betweenness centrality

Nel grafico a barre 9 vengono presentati gli id dei primi 10 nodi con valore di betweenness maggiori e gli ultimi 10 nodi con valore di betweenness peggiori. Si può notare che molti id sono in comune anche con la degree centrality.

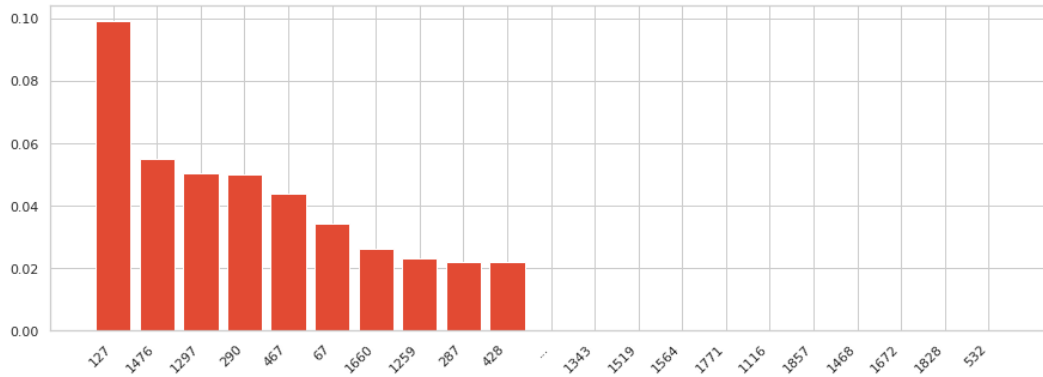


Figura 9: Grafico della classifica della betweenness centrality

Come è possibile vedere dalla figura 10, la distribuzione presenta un picco elevato su valori molto bassi, lasciando pochi altri elementi nel resto del grafico. Questo può essere visto anche dal grafico in figura 9, che riporta solo 4 canali con centralità superiore a 0.05, mentre le centralità rimanenti si attestano su valori più bassi di tale soglia. I canali Twitch con la maggior betweenness centrality sono quelli identificati dagli id 127, 1476, 1297 e 290.

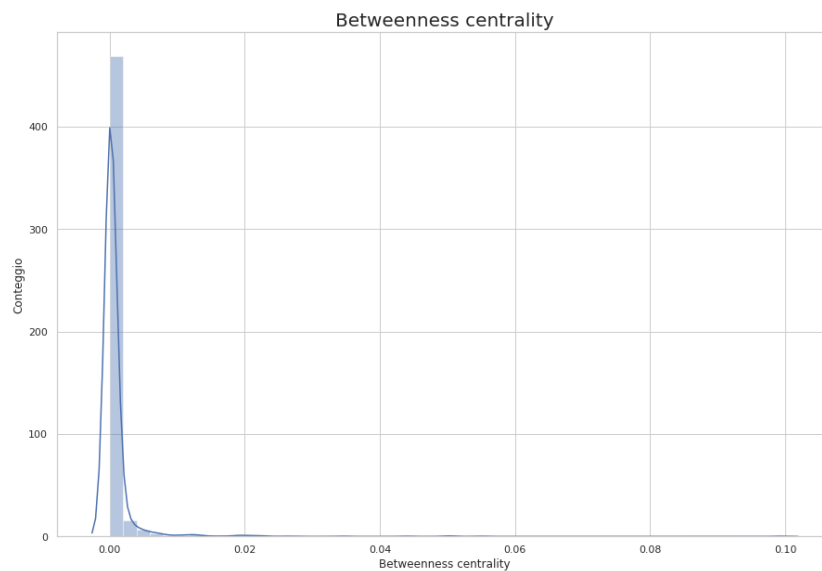


Figura 10: Grafico della distribuzione della betweenness centrality

In figura 11 è possibile vedere lo spring layout della betweenness centrality in cui si nota che, a differenza della degree centrality, nella parte centrale non vi è nessuna variazione di colore, segno che la variazione di centralità è molto meno significativa.

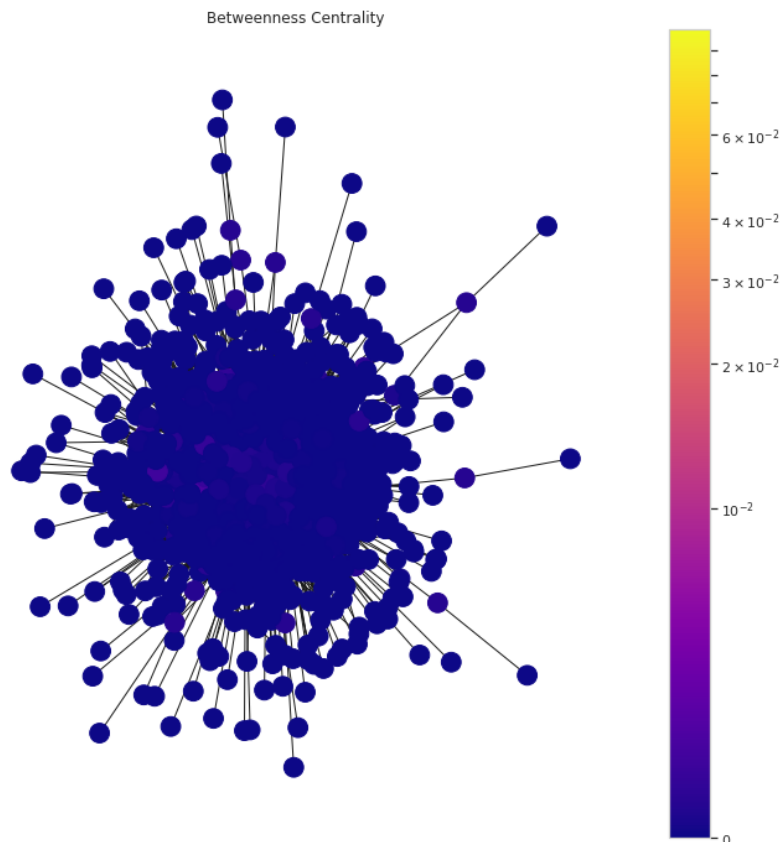


Figura 11: Spring layout del grafo della betweenness centrality

### 4.3 Closeness Centrality

La terza metrica appartenente alle “misure di centralità” è la closeness centrality. Tale metrica esprime il grado di prossimità di un nodo agli altri nodi della rete. La capacità di un nodo di ricevere e inviare informazioni agli altri dipende fortemente dalla distanza tra esso e il resto della rete. Ricordiamo che l’orizzonte di osservabilità di un nodo è di circa due livelli. La capacità di spostare informazioni da un lato all’altro della rete (cioè il gossip) è un passo importante nell’obiettivo di stabilire una percezione condivisa del mondo. Così, la distanza dagli altri, (o, al contrario, la vicinanza) può definire il ruolo di una persona in una rete. Il calcolo della closeness centrality è piuttosto costoso:

- Calcola il path più breve tra ciascuna coppia di nodi utilizzando l’algoritmo di Dijkstra, e memorizza queste distanze in una tabella.
- Per ciascun nodo:
  - Somma le distanze tra esso e tutti gli altri nodi.
  - La closeness distance è pari al reciproco di questa somma.

Il risultato è un numero tra 0 e 1; numeri più alti implicano maggiori vicinanze e quindi minori distanze. Maggiore è il valore della closeness centrality di un nodo e maggiore è la possibilità che quel nodo può trasportare informazioni in maniera efficiente nel grafo. Per calcolare questa metrica è stata utilizzata la funzione `nx.closeness centrality()` di NetworkX.

Così come è stato eseguito per le centralità precedenti, i risultati dei 10 canali più centrali e dei 10 canali meno centrali ottenuti dal calcolo della metrica sono riportati di seguito nella figura 12:

**Best 10 Closeness Centrality:**

```
1. 127: 0.6009433962264151
2. 1297: 0.5684116597263533
3. 467: 0.565050266114725
4. 290: 0.5595900439238654
5. 1476: 0.5594262295081968
6. 67: 0.5435153583617748
7. 1660: 0.5416666666666666
8. 1593: 0.5295095594347464
9. 1259: 0.5237051246916964
10. 287: 0.5227024070021882
```

**Worst 10 Closeness Centrality:**

```
1. 1832: 0.20484510665666203
2. 667: 0.23810117119362073
3. 770: 0.24043784599899345
4. 196: 0.24043784599899345
5. 1672: 0.24260505268503238
6. 1794: 0.24658064516129033
7. 1032: 0.25197784810126583
8. 1044: 0.2532467532467532
9. 1222: 0.2541899441340782
10. 1070: 0.2549019607843137
```

Figura 12: Tabella della classifica della closeness centrality

Anche in questo tipo di metrica i risultati ottenuti per i primi 10 canali risultano essere abbastanza simili a quelli ottenuti nelle metriche precedenti, mentre si ha qualche variazione per quanto riguarda i peggiori 10.

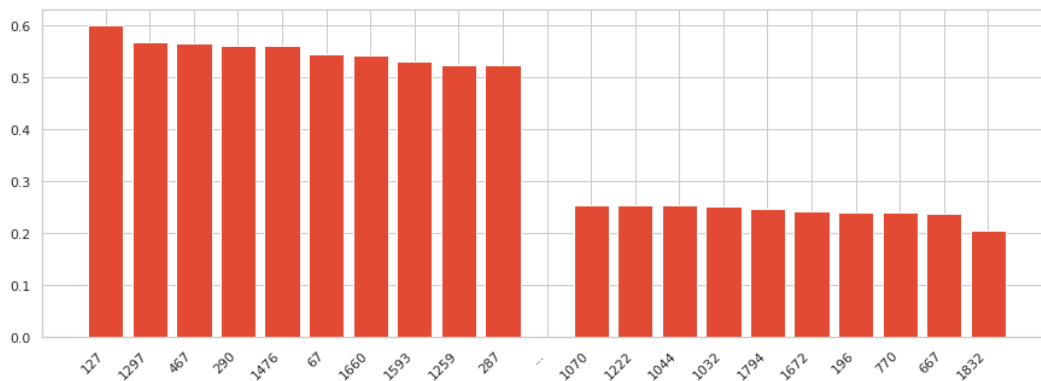


Figura 13: Grafico della classifica della closeness centrality

In figura 14 si può vedere come in questo caso la distribuzione della centralità sia molto più ampia. Infatti la maggior parte dei nodi si concentrano nella fascia di centralità compresa tra 0.3 e 0.5, mentre i nodi con centralità molto bassa partono da un minimo di 0.2.

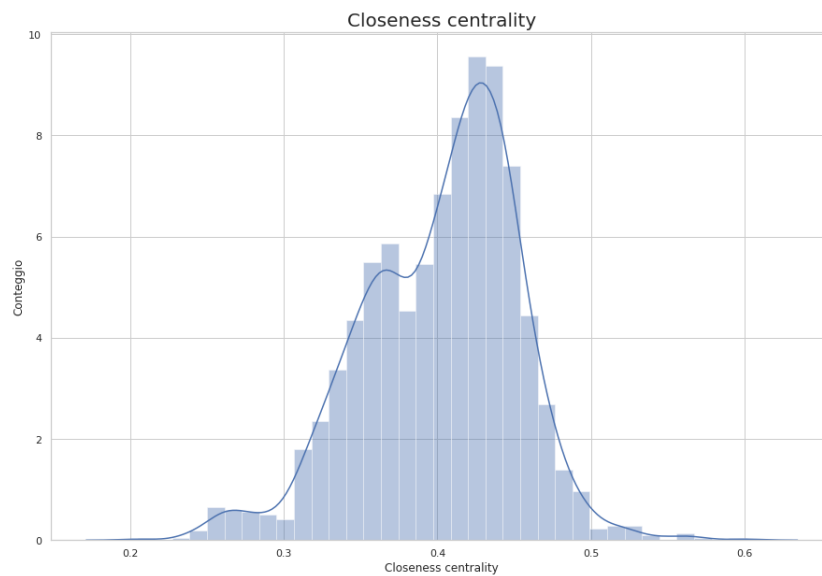


Figura 14: Grafico della distribuzione della closeness centrality

In figura 15 si può vedere lo spring layout abbia dei colori differenti rispetto ai layout delle centralità precedenti. Infatti i colori sono molto più chiari e si percepisce una sfumatura molto più progressiva.

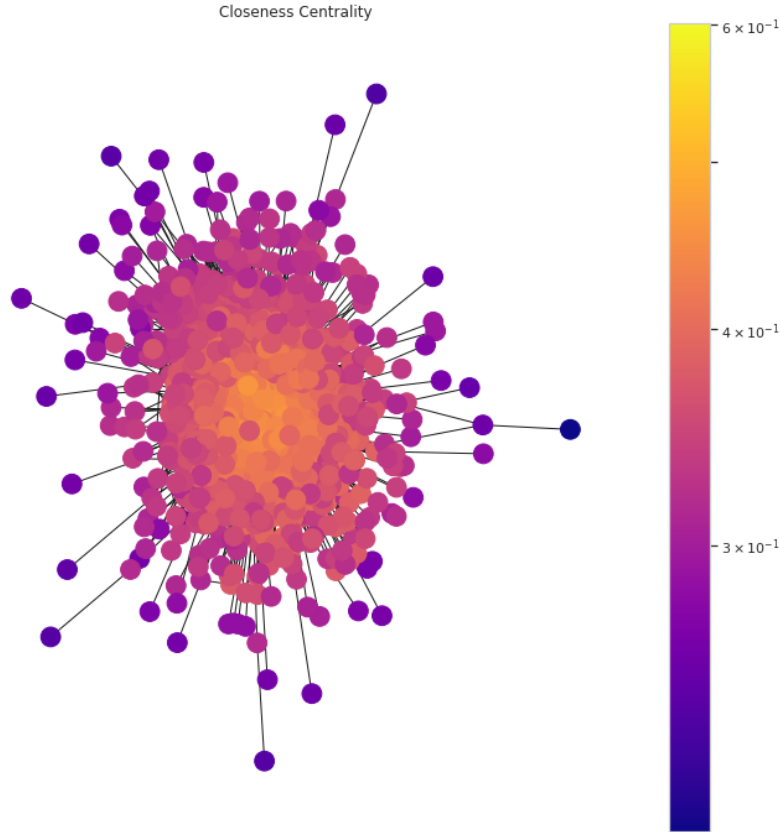


Figura 15: Spring layout del grafo della closeness centrality

#### 4.4 Eigenvector Centrality

L'ultima metrica appartenente alle “misure di centralità” è la eigenvector centrality. Questo particolare tipo di centralità misura quanto effettivamente sia autorevole ed influente un nodo all'interno della rete. Ciò, in pratica, misura la reputazione e l'importanza di un nodo rispetto a quelli che gli sono vicini. I punteggi più alti vengono assegnati a nodi che sono collegati ad altri nodi con punteggi elevati. L'eigenvector centrality si basa su tale principio. Con questa metrica un nodo è centrale se è connesso a molti nodi che sono, a loro volta, centrali. È una definizione ricorsiva. L'algoritmo per il calcolo di questa centralità è il seguente:

- Si inizia assegnando un grado di centralità pari ad 1 a tutti i nodi.
- Si ricalcolano i punteggi di ciascun nodo come una media pesata delle centralità di tutti i nodi appartenenti al suo vicinato:

$$\nu_i = \sum_{j \in N} x_{ij} * \nu_j \quad (1)$$

- Si normalizzano i numeri così ottenuti dividendo ciascun valore per quello massimo.
- Si ripetono i passi 2 e 3 fino a quando i valori non cambiano o cambiano di poco

Potrebbero essere necessarie centinaia di iterazioni prima di convergere; pertanto, tale metrica non è realistica per reti molto grandi. I 10 risultati migliori ed i 10 risultati peggiori ottenuti sono riportati di seguito in figura 16:

#### Best 10 Eigenvector Centrality:

```
1. 127: 0.17877831012617335
2. 1297: 0.15866803272155208
3. 467: 0.15723938915542482
4. 290: 0.15150156437553058
5. 1476: 0.1457607576802338
6. 1660: 0.1337570263037189
7. 67: 0.13118052973454408
8. 1593: 0.1258976204853149
9. 1320: 0.11858724919571513
10. 1758: 0.11540890407339238
```

#### Worst 10 Eigenvector Centrality:

```
1. 1832: 1.7436095815676368e-07
2. 667: 2.2159017286266766e-06
3. 770: 3.387873138125768e-06
4. 196: 3.387873138125768e-06
5. 1672: 3.941131608062454e-06
6. 1794: 8.248227713411063e-06
7. 1222: 8.384965671283339e-06
8. 1044: 9.353493634091062e-06
9. 1319: 1.1200682753021389e-05
10. 464: 1.2822153199565164e-05
```

Figura 16: Tabella della classifica della eigenvector centrality

Anche in questo tipo di metrica i risultati ottenuti per i primi 10 e gli ultimi 10 canali risultano essere abbastanza simili a quelli ottenuti nelle metriche precedenti. Infatti, come visibile dall'immagine 17, al primo posto abbiamo il canale Twitch identificato dall'id 127, seguito poi da quello con id 1297 e 467.

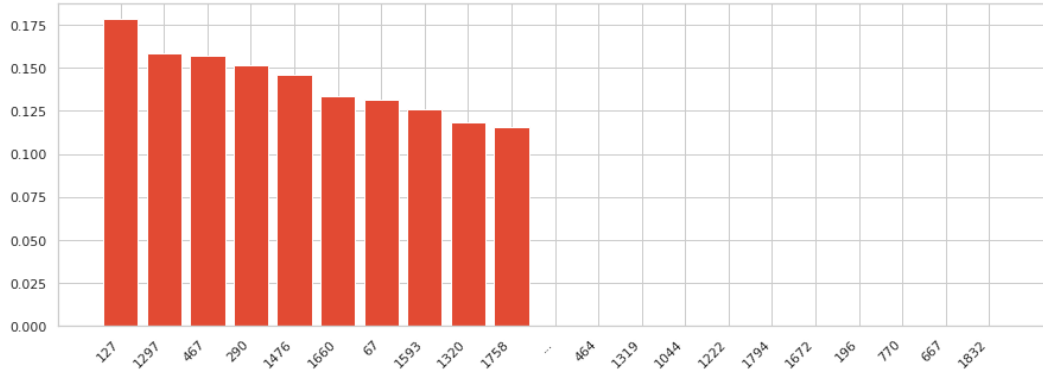


Figura 17: Grafico della classifica della eigenvector centrality

Anche in questo caso, come nella degree e la betweenness centrality, sono molto pochi i nodi ad avere una elevata eigenvector centrality, mentre sono molti i nodi con una centrality bassissima, compresa tra lo 0 e 0.001.



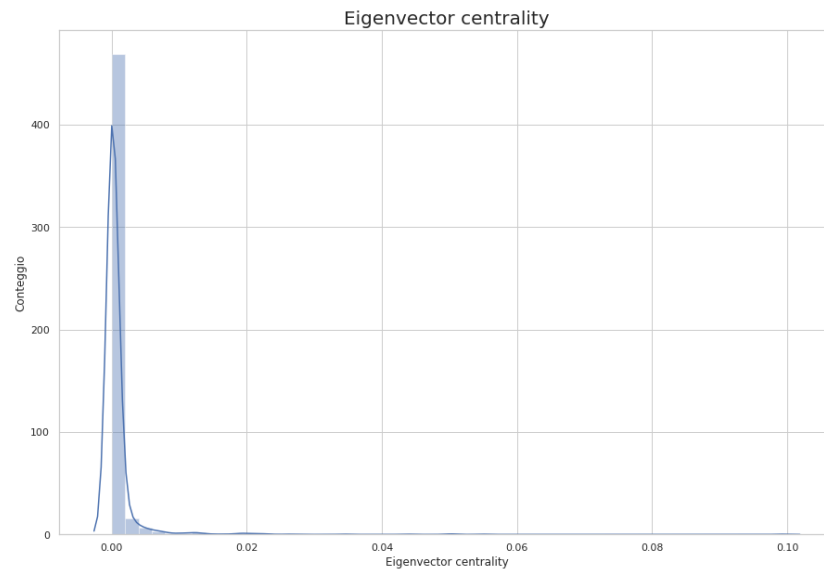


Figura 18: Grafico della distribuzione della eigenvector centrality

In figura 20 vediamo che in questo caso si nota un punto centrale in cui vi è una sfumatura più chiara, ovvero il punto in cui la centralità è più alta.

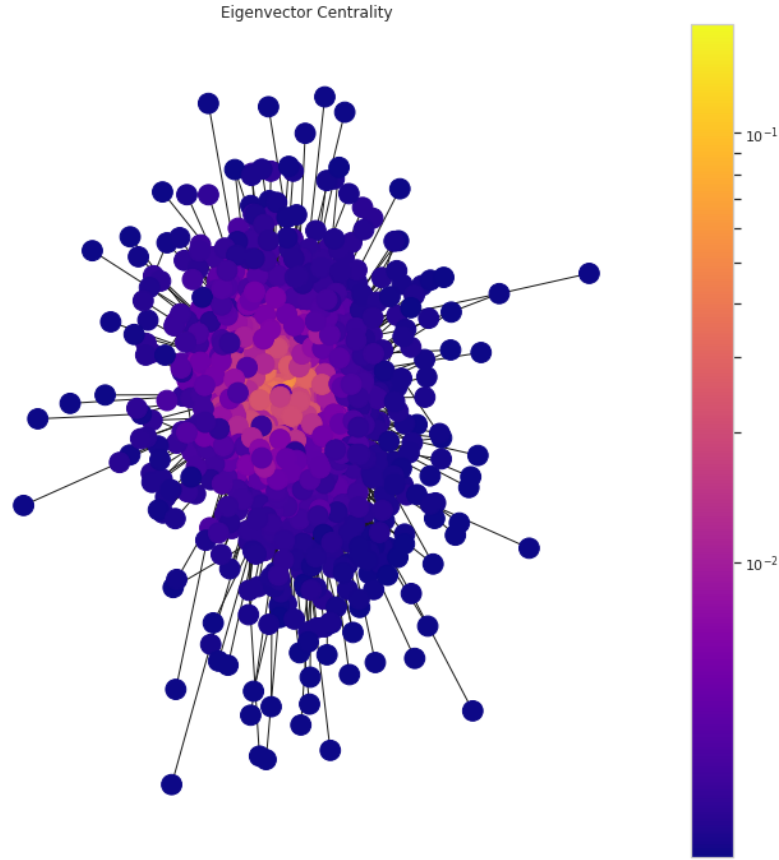


Figura 19: Spring layout del grafo della eigenvector centrality

#### 4.5 Intersezione delle diverse centralità

In questo paragrafo andiamo a confrontare i risultati ottenuti nelle diverse centralità intersecando i primi dieci nodi di ognuna. In particolare, nella tabella sottostante vengono inseriti solo i valori dei nodi presenti in tutte e quattro le centralità. Per normalizzare tale valore, è stata calcolata una media dei valori di ogni centralità. Come si può osservare dalla tabella sottostante, al primo posto è presente il nodo 127, infatti esso è il nodo che sta nella prima posizione di ogni centralità. Quindi ci permette di affermare che il canale Twitch con id 127 è un canale molto rilevante all'interno della rete. Le stesse considerazioni possono essere fatte per i canali identificati dagli id 1297, 1476, 467 e 290. E' interessante osservare che nelle quattro centralità, questi canali hanno valori di centralità abbastanza simili, che si distaccano dai valori di centralità del canale più importante.

Posizione	Nodo	Valore medio della centralità
1	127	0,3200
2	1297	0,2711
3	1476	0,2682
4	467	0,2676
5	290	0,2674
6	1660	0,2375
7	67	0,2366

Nel grafico sottostante viene riportato il valore medio risultante dal calcolo delle quattro centralità.

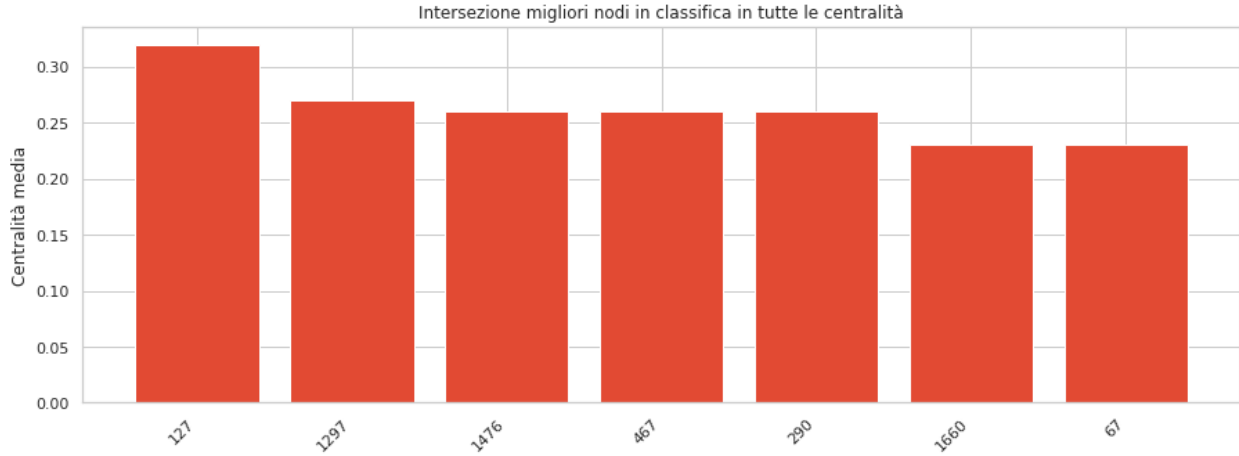


Figura 20: Grafico della classifica delle intersezioni delle diverse centralità

## 5 Clique

Nella teoria dei grafi una cricca (o clique) è un insieme  $V$  di vertici in un grafo non orientato  $G$ , tale che, per ogni coppia di vertici in  $V$ , esiste un arco che li collega. In modo equivalente, si potrebbe dire che il sottografo indotto da  $V$  è un grafo completo. La dimensione di una cricca è definita come il numero di vertici che contiene. Alcuni chiamano cricca ogni sottografo completo che sia di dimensione massima. Il problema di trovare, se esiste, una cricca di una dimensione fissata all'interno di un grafo è detto problema della cricca, ed è un problema NP-completo. Il concetto complementare a quello di cricca è l'insieme indipendente, nel senso che a ogni cricca corrisponde un insieme indipendente nel grafo complementare.

In questa sezione verranno analizzate alcune delle clique principali individuate nel grafo, al fine di individuare delle relazioni interessanti tra i suoi nodi. In particolare, si sono individuate le clique con il massimo numero di nodi. Nella figura 21 vengono riportate, attraverso una distribuzione, tutte le possibili clique con il loro rispettivo numero di nodi. E' interessante osservare come il numero massimo di clique si ha con 8 nodi, ovvero con 8 nodi possiamo avere più di 10000 sottografi completi. Anche con 7 nodi è possibile costruire circa 10000 clique.

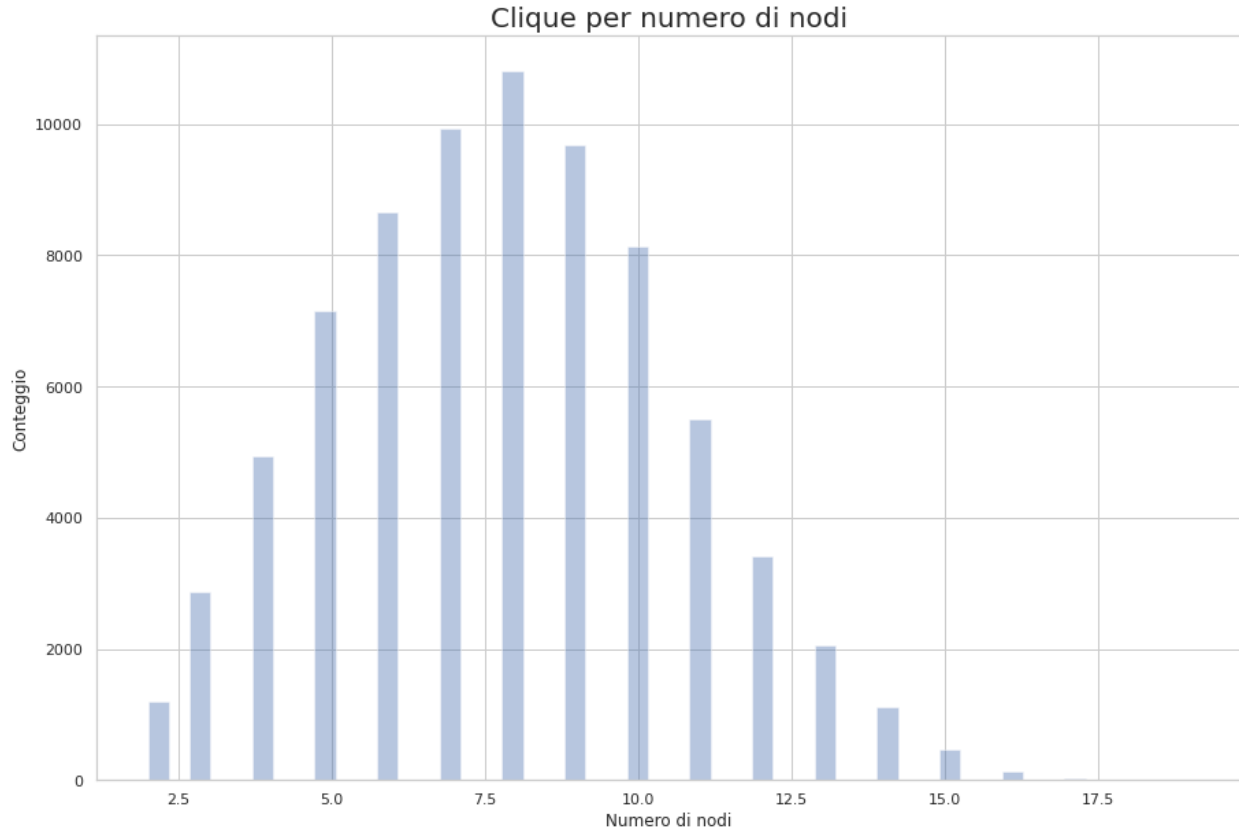


Figura 21: Distribuzione del numero di clique per numero di nodi

Nella figura 22 viene rappresentato graficamente attraverso lo spring layout la clique con il maggior numero di nodi. La clique massima si ha con 19 nodi ed essi corrispondono ai canali Twitch identificati dai seguenti id: [496, 158, 1732, 94, 224, 1199, 1100, 63, 1289, 1034, 1357, 309, 939, 1208, 237, 1803, 1506, 1107, 1347]. E' interessante osservare come il canale Twitch con id 127, identificato precedentemente come canale centrale della rete, non sia presente tra i nodi che formano la clique massima. A causa della scarsa presenza di informazioni sul nostro dataset, non è possibile risalire al nome del canale noto l'id dello stesso. Qualora questo fosse stato possibile, sicuramente avremmo avuto un'analisi più significativa e di maggiore livello.

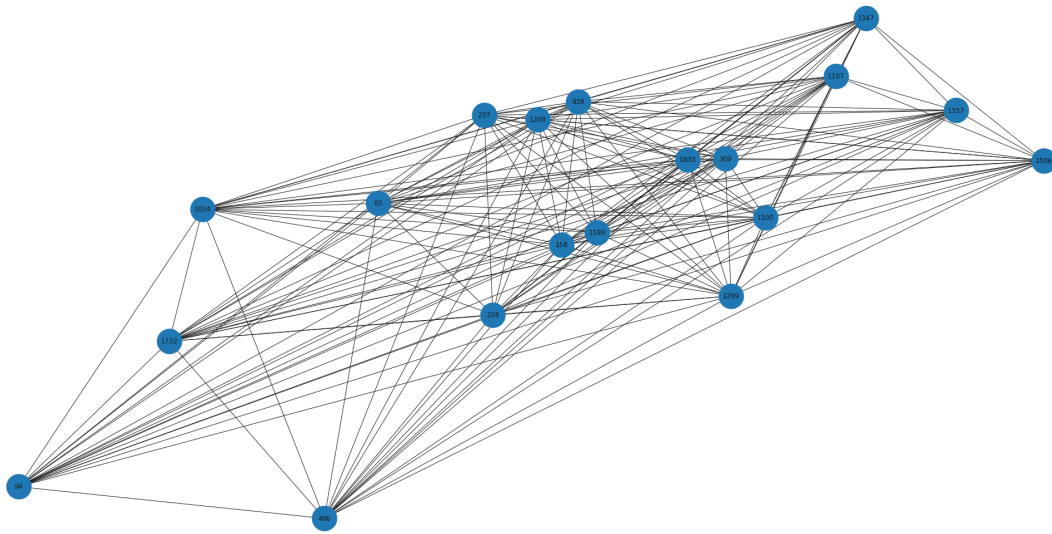


Figura 22: Grafo della clique massima con spring layout

## 6 Community

L'idea principale della Social Network Analysis è studiare, misurare e rappresentare le relazioni sociali tra gli individui o gruppi di individui. Per tale ragione in questa sezione è stata affrontata la problematica di community detection, ovvero cercare di classificare i nodi in dei sotto-insiemi chiamati comunità. I nodi appartenenti a quest'ultime rappresentano un insieme coeso che condividono alcune caratteristiche chiave, nel nostro caso dei collegamenti ad altri nodi. Questa non è l'unica feature che i nodi possono condividere in quanto l'operazione di community detection prende in considerazione tutte le caratteristiche associate al singolo nodo.

L'algoritmo utilizzato è di tipo greedy ovvero individua la soluzione ottima a livello globale andando, ad ogni iterazione, a compiere la "scelta" migliore per quel determinato passo. Nello specifico è stato utilizzato l'algoritmo proposto da Clauset-Newman-Moore il quale è già presente all'interno delle librerie NetworkX. Questo algoritmo utilizza la massima modularità per trovare la partizione di comunità con la maggiore modularità. Nella tabella sottostante vengono riportate tutte le community, contraddistinte dal numero di nodi, dal numero di archi e dalla densità e clustering. In particolare, sono state identificate 20 diverse community. E' interessante osservare che man mano che il numero di nodi e il numero di archi diminuiscono, la densità della community aumenta mentre la misura di clustering diminuisce. L'unica eccezione che possiamo avere è in corrispondenza della community numero 9, essendo costituita da tre nodi e tre archi. Questo ci fa già capire che si tratta di un sottografo completo in cui ogni coppia di nodi è collegata da un arco, infatti le sue misure di densità e clustering sono pari ad 1.

Community	Numero nodi	Numero archi	Densità	Clustering
1	932	11770	0.0271	0.4485
2	606	7933	0.0432	0.3140
3	261	1489	0.0438	0.3876
4	57	87	0.0545	0.2523
5	14	14	0.1538	0.15
6	6	5	0.3333	0.0
7	5	4	0.4	0.0
8	3	2	0.6666	0.0
9	3	3	1.0	1.0
10	3	2	0.6666	0.0
11	3	2	0.6666	0.0
12	3	2	0.6666	0.0
13	2	1	1.0	0.0
14	2	1	1.0	0.0
15	2	1	1.0	0.0
16	2	1	1.0	0.0
17	2	1	1.0	0.0
18	2	1	1.0	0.0
19	2	1	1.0	0.0
20	2	1	1.0	0.0

Nell'immagine 23 vengono mostrati i grafici delle 20 community presenti. Come si può notare, i primi cinque grafi mostrano le prime cinque community con più di sei nodi, mentre i restanti grafici identificano le successive community con un numero di nodi uguale o inferiore a sei fino ad arrivare alle ultime community composte solamente da due nodi.

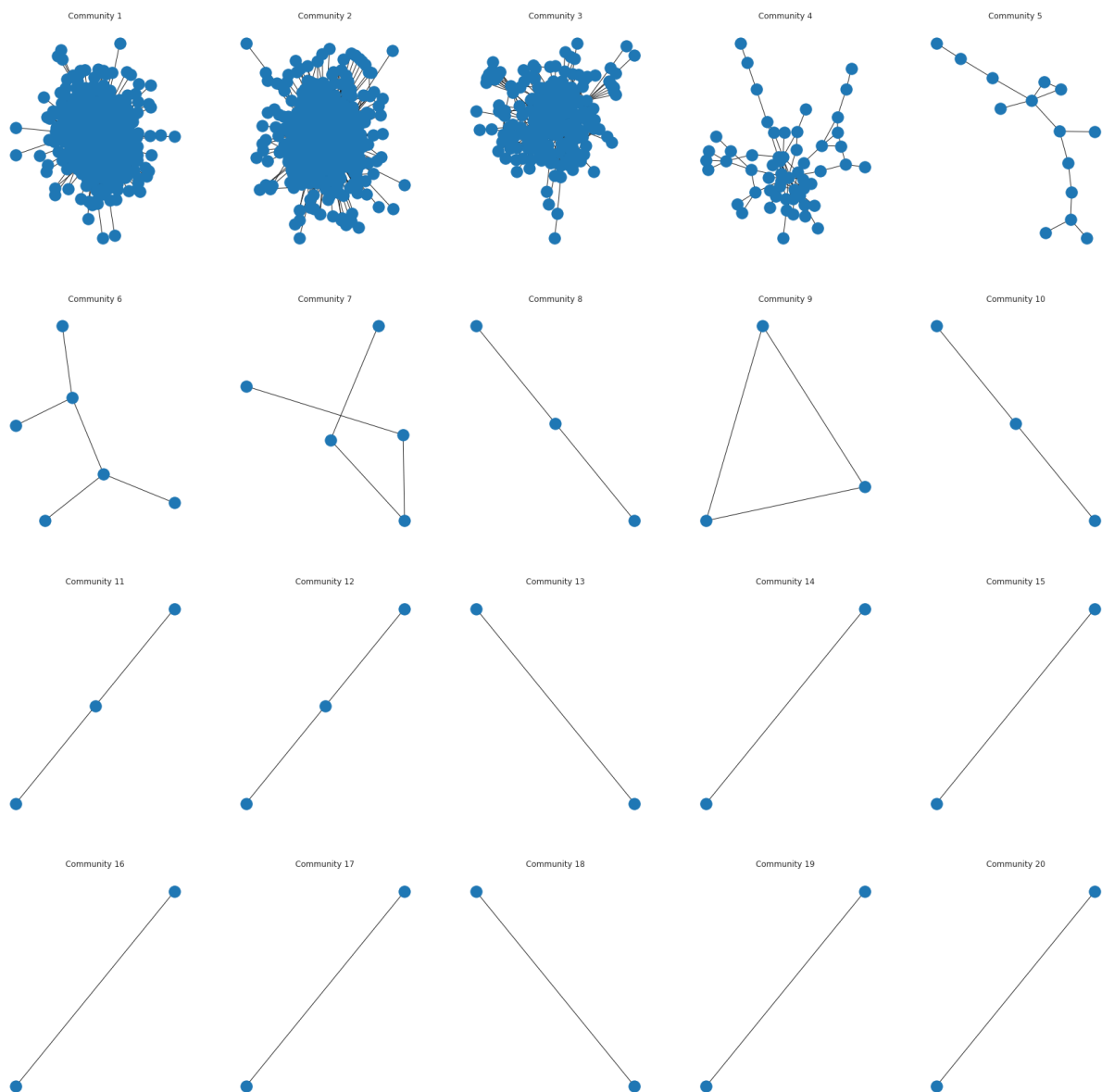


Figura 23: Grafi delle community

Nel grafico 24 viene rappresentato graficamente attraverso un spring layout la community più numerosa che nel nostro caso è la community numero 1.

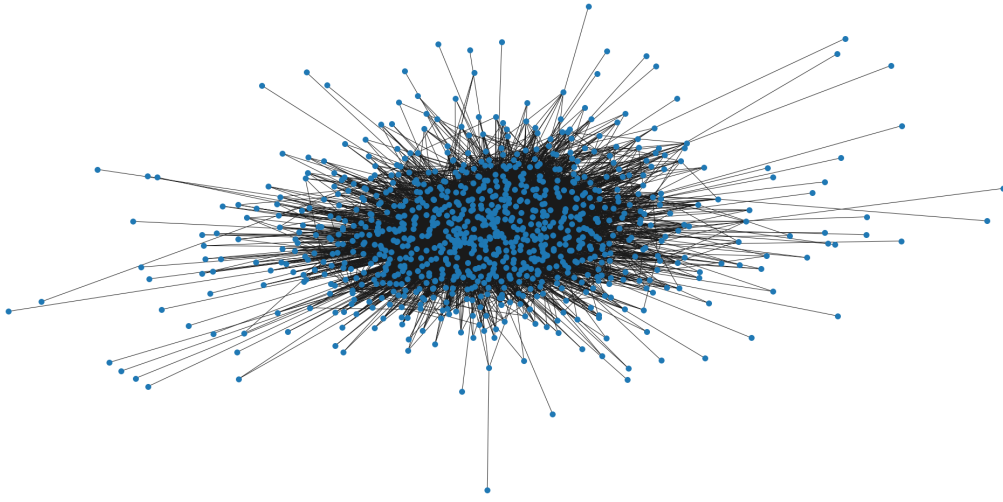


Figura 24: Grafi delle community più grande con spring layout