

Università Politecnica delle Marche  
Dipartimento di Ingegneria dell'Informazione

Facoltà di Ingegneria Informatica e dell'Automazione



**L'analisi della rete costituita dai personaggi della serie "Breaking Bad" tramite la libreria NetworkX**

**Docenti:**

Prof. Domenico Ursino  
Michele Marchetti

**Studenti:**

Silvia Ciuffreda  
Luca Liberatore  
Beatrice Moliterno

**ANNO ACCADEMICO 2022/2023**

# Indice

<b>1</b>	<b>Introduzione</b>	<b>3</b>
1.1	SNA . . . . .	3
1.2	NetworkX . . . . .	3
<b>2</b>	<b>Dataset</b>	<b>4</b>
2.1	Presentazione del dataset . . . . .	4
2.2	Descrizione del grafo . . . . .	7
<b>3</b>	<b>Le misure di Centralità</b>	<b>11</b>
3.1	Degree centrality . . . . .	11
3.2	Closeness centrality . . . . .	12
3.3	Betweenness centrality . . . . .	14
3.4	Eigenvector Centrality . . . . .	16
<b>4</b>	<b>La visualizzazione e l'analisi delle strutture</b>	<b>19</b>
4.1	Le triadi . . . . .	19
4.1.1	Clique . . . . .	20
4.2	Il K-core . . . . .	21
4.3	Le ego network . . . . .	21

## Elenco delle figure

1	Il logo di NetworkX . . . . .	3
2	Walter White . . . . .	4
3	Jesse Pinkman . . . . .	4
4	Skyler White . . . . .	5
5	Hank Schrader . . . . .	5
6	Gustavo Fring . . . . .	5
7	Saul Goodman . . . . .	6
8	Mike Ehrmantraut . . . . .	6
9	Schema della storia della maggior parte dei personaggi . . . . .	7
10	I primi dieci archi di peso maggiore . . . . .	8
11	Caratteristiche generali del grafo . . . . .	8
12	Spring layout . . . . .	9
13	Circular layout . . . . .	10
14	Conteggio peso degli archi . . . . .	10
15	Primi ed ultimi dieci nodi per degree centrality . . . . .	11
16	Distplot degree centrality . . . . .	12
17	Spring layout - Degree centrality . . . . .	12
18	Primi ed ultimi dieci nodi per closeness centrality . . . . .	13
19	Distplot Closeness centrality . . . . .	13
20	Spring layout - Closeness centrality . . . . .	14
21	Primi ed ultimi dieci nodi per Betweenness centrality . . . . .	15
22	Distplot Betweenness centrality . . . . .	15
23	Spring layout - Betweenness centrality . . . . .	16
24	Primi ed ultimi dieci nodi per Eigenvector centrality . . . . .	17
25	Distplot Eigenvector centrality . . . . .	17
26	Spring layout - Eigenvector centrality . . . . .	18
27	L'analisi delle triadi . . . . .	19
28	L'analisi delle clique . . . . .	20
29	L'analisi della clique massima . . . . .	21
30	I collegamenti di Gus Fring . . . . .	21
31	L'analisi della k-core . . . . .	22
32	L'analisi della ego network . . . . .	22

# 1 Introduzione

## 1.1 SNA

La Social Network Analysis (SNA), diversamente da quanto si crede, nasce prima dell'avvento dei più famosi social network come Twitter e Facebook. Questa metodologia consiste nello studio delle relazioni umane facendo uso della teoria dei grafi. Proprio per questa sua natura è stata utilizzata tantissimo nei social media ed è diventata estremamente famosa con il passare degli anni. La sociologia sta alla base di questa tecnica di Data Science. La SNA si fonda su alcuni principi fondamentali, il più importante - forse - può essere sintetizzato con la massima: "le nostre relazioni definiscono chi siamo e come agiamo". Quindi, lo studio delle relazioni porta a delle conclusioni che riguardano il comportamento sociale dell'individuo in esame. Le relazioni possono essere sia simmetriche che asimmetriche, sia di amicizia che di inimicizia, di somiglianza o di conflitto, ecc. Un concetto fondamentale che partecipa a queste analisi è quello di omofilia, la quale sostiene che un soggetto tende solitamente a formare legami sociali con chi è più simile a lui. In altre parole, il comportamento di un individuo tende a essere simile a quello dei suoi vicini. Tuttavia, per motivazioni biologiche, la specie umana ha un numero massimo di persone con cui può mantenere un contatto che risulta essere pari a circa 150 individui.

In sintesi, una rete sociale può essere rappresentata come una collezione di elementi che descrivono delle relazioni; nella teoria dei grafi ogni relazione è rappresentata da un arco ed ogni individuo coinvolto nella relazione viene descritto attraverso un nodo (o vertice).

## 1.2 NetworkX

In questo progetto si farà uso della libreria NetworkX per lo studio delle reti secondo la teoria dei grafi. NetworkX fornisce diversi strumenti, grafici e non, per lo studio della struttura delle reti sociali e permette di lavorare con grandi dataset, in modo rapido ed efficiente.



Figura 1: Il logo di NetworkX

## 2 Dataset

Il dataset scelto su cui svolgere l'analisi fa riferimento ai personaggi di una delle più famose serie tv americane: "*Breaking Bad*", scaricabile al seguente link: [https://github.com/jishnukoliyadan/the\\_breaking\\_bad\\_network/tree/master/data](https://github.com/jishnukoliyadan/the_breaking_bad_network/tree/master/data).

### 2.1 Presentazione del dataset

*Breaking Bad - Reazioni collaterali* è una serie televisiva creata da Vince Gilligan e andata in onda dal 2008 al 2013. La serie rientra nel genere drammatico/thriller e tratta sia temi familiari che inerenti alla malavita, in particolare, la produzione ed il commercio di metanfetamina. All'interno del dataset vengono considerati come nodi tutti i personaggi della serie, dalla prima all'ultima stagione, mentre come archi le relazioni che intercorrono tra di stessi. Tra i personaggi principali, che rivestono un ruolo fondamentale nella serie, è possibile trovare:

- *Walter White* (Figura 2), un professore di chimica di un liceo di Albuquerque, che in seguito alla scoperta di un cancro terminale ai polmoni, decide di usare il suo talento per produrre metanfetamine per provvedere alla sua famiglia dopo la morte. Entrerà nel traffico della droga e costruirà un impero, tanto da cancellare la sua umanità e da trasformare il mite chimico in uno spietato assassino e trafficante. Il suo nome in codice sarà "Heisenberg".



Figura 2: Walter White

- *Jesse Pinkman* (Figura 3), ex studente di Walter White, tossicodipendente e spacciatore di metanfetamine; sarà lui ad aiutare Walter nella produzione e distribuzione di droga, diventando così soci in affari. Al contrario di Walter che diverrà sempre più spietato, Jesse diverrà sempre più buono e pieno di umanità.



Figura 3: Jesse Pinkman

- *Skyler White* (Figura 4), la moglie di Walter White, che in seguito alla scoperta delle attività illegali del marito, deciderà di aiutarlo per il riciclaggio del denaro ottenuto dal traffico di droga, aprendo un autolavaggio. Alla fine il loro matrimonio diventerà sempre più teso portando i due alla separazione.

## 2.1 Presentazione del dataset



Figura 4: Skyler White

- *Hank Schrader* (Figura 5), cognato di Walter e Skyler White, nonché un ufficiale di alto rango presso l'ufficio di Albuquerque della "Drug Enforcement Administration" (DEA). Per tutta la serie, condurrà le indagini sul cuoco della metanfetamina "Heisenberg", ignaro che il misterioso boss della droga sia suo cognato, che appunto sarà chiamato con questo pseudonimo.



Figura 5: Hank Schrader

- *Gustavo "Gus" Fring* (Figura 6), un imprenditore cileno-americano di una catena di fast food chiamata "Los Pollos Hermanos", che usa come attività lavorativa di facciata per nascondere la sua attività di distribuzione di metanfetamina. Lui, secondo la storia, inizialmente sarà affiliato con il "Cartello", un'associazione criminale che gestisce il traffico di droga dell'intero New Mexico e delle zone limitrofe, per poi rendersi indipendente creando un proprio impero. Inoltre, Walter White diventerà suo dipendente, finché il protagonista non organizzerà il suo omicidio.



Figura 6: Gustavo Fring

- *Saul Goodman* (Figura 7), un pittoresco ed esilarante avvocato, specializzato nel difendere criminali ormai irrecuperabili, ma estremamente astuto e competente. Sarà assunto da Walter White e Jesse come consigliere.

## 2.1 Presentazione del dataset



Figura 7: Saul Goodman

- *Mike Ehrmantraut* (Figura 8), un criminale americano in carriera, veterano del Corpo dei Marines ed ex agente di polizia di Filadelfia. Lui sarà l'investigatore privato di Saul Goodman e l'uomo fidato al soldo di Gus Fring, infatti, lui sarà l'intermediario tra i due; grazie a questa figura, quindi, Heisenberg e il proprietario di Los Pollos Hermanos entreranno in collaborazione. In seguito, diventerà il socio in affari di Walter White e Jesse Pinkman.



Figura 8: Mike Ehrmantraut

In Figura 9 è riportato lo schema riassuntivo delle principali relazioni e vicende che intercorrono tra i diversi personaggi della serie.

## 2.2 Descrizione del grafo

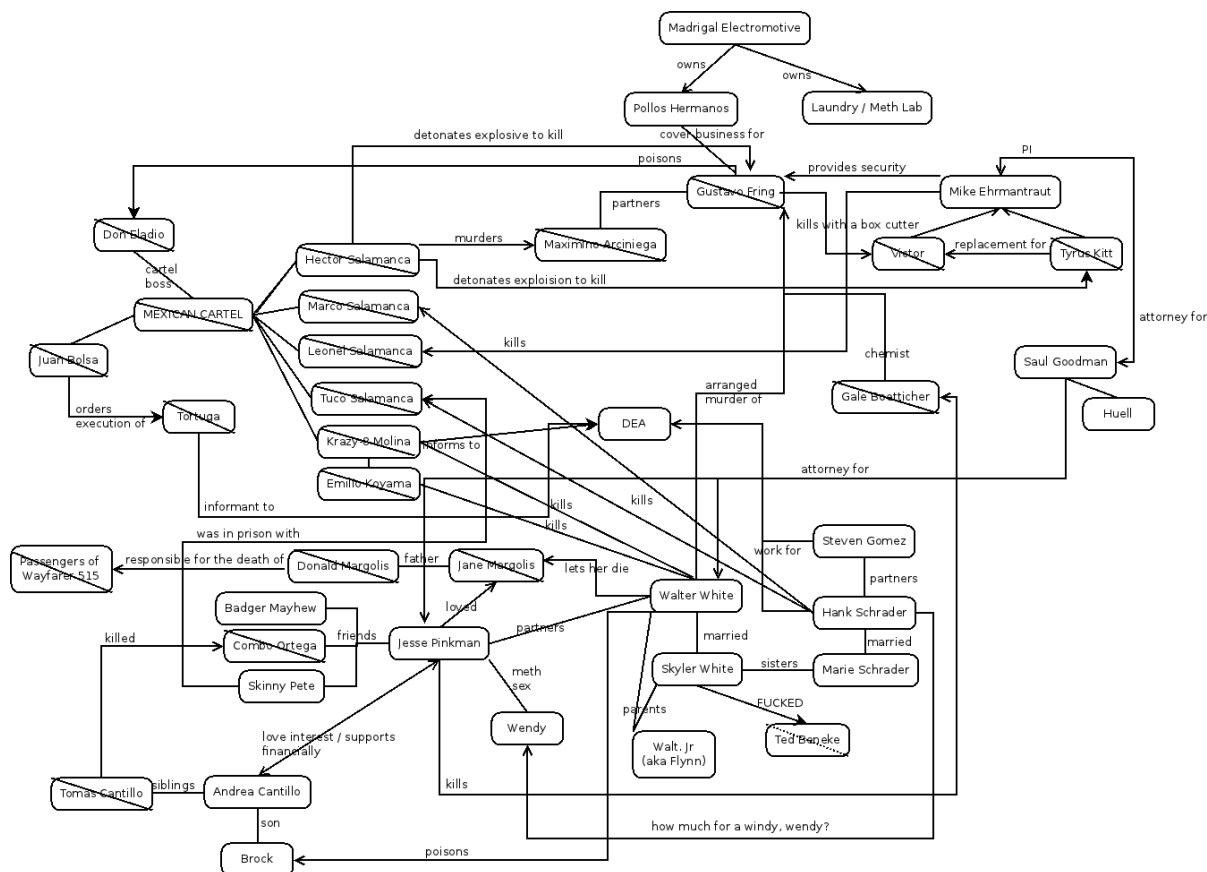


Figura 9: Schema della storia della maggior parte dei personaggi

## 2.2 Descrizione del grafo

La social network analysis è stata svolta con l'ausilio dell'ambiente di sviluppo *Jupyter Notebook* e delle librerie Python quali: *Pandas* per la manipolazione dei dati, *NetworkX* per l'implementazione dei grafi, *Seaborn* e *Matplotlib* per la visualizzazione.

Il dataset presenta 426 righe e 3 colonne; i campi *source* e *target* fanno riferimento ai due personaggi costituenti l'arco, il campo *weight*, invece, indica la "forza" della relazione tra i due personaggi, nonchè il peso dell'arco.

In Figura 10 sono riportati i primi 10 archi di peso maggiore del dataset; come ci si poteva aspettare, la maggior parte degli archi contiene Walter. Inoltre, sono presenti in questa tabella Marie e Tuco Salamanca; sebbene non siano stati citati tra i personaggi principali, queste figure hanno una loro importanza nella storia e questo giustifica il peso degli archi che li contengono. Infatti, Marie è la moglie di Hank e sorella di Skyler, che comparirà in tutte le stagioni, mentre, Tuco Salamanca è il capo del "Cartello" che, nonostante compaia solo alla fine della prima e all'inizio della seconda stagione, è fondamentale per Walter e Jesse affinché loro entrino nella rete di distribuzione della metanfetamina.



## 2.2 Descrizione del grafo

	source	target	weight
0	Walter	Jesse	2185
1	Walter	Skyler	847
2	Walter	Hank	708
3	Walter	Mike	394
4	Jesse	Mike	380
5	Hank	Marie	286
6	Walter	Saul Goodman	265
7	Hank	Jesse	249
8	Walter	Tuco Salamanca	198
9	Walter	Marie	198

Figura 10: I primi dieci archi di peso maggiore

A partire dal dataset sopracitato, è stato costruito il grafo; da questo possiamo andare ad estrarre le sue caratteristiche strutturali principali, quali, ad esempio, il numero di nodi, il numero di archi, la densità ed altri. I valori risultati sono espressi nella Figura 11.

```
Numero di Nodi: 104
Numero di Archi: 426
Densità: 0.08
Densità Pesata: 2.405
Raggio: 2
Diametro: 4
Periferia: ['ASAC George Merkert', 'Janice', 'Ortega', 'Don', 'Gang', 'Stephanie', 'Patrick', 'Soper', 'Getz', 'Picarus', 'Chow', 'Chris', 'Don Eladio Vuento']
Average Clustering Coeff: 0.652
Is connected: True
```

Figura 11: Caratteristiche generali del grafo

Il numero totale di nodi è 104, mentre il numero totale di archi è 426; questi ultimi sono bidirezionali. La densità rappresenta la probabilità che una coppia di nodi sia adiacente e il valore varia tra 0 e 1; in questo caso, si ha una probabilità bassa, di 0.08, e questo suggerisce che gli archi presenti nel grafo colleghino principalmente i protagonisti con i personaggi "secondari", i quali hanno pochi collegamenti tra loro. La densità pesata è definita come il rapporto della somma dei pesi degli archi esistenti rispetto al numero dei possibili archi; tenendo conto che il peso minimo di un arco è pari a 5, mentre il massimo è pari a 2185, una densità pesata di 2,405 è bassa ma sarebbe potuta essere inferiore se i primi tre valori riportati in Figura 10 non fossero stati così alti. Il diametro è il più grande valore di eccentricità possibile di tutti i nodi del grafo; per tale dataset, scelto un qualsiasi nodo e passando per al massimo 4 archi, è possibile raggiungere tutti gli altri nodi del grafo. Il raggio, invece, è il valore minimo di eccentricità di tutti i nodi del grafo, ed in questo caso tale valore è pari a 2. Per "Periferia" si indica il sotto-grafo costituito dai nodi che hanno l'eccentricità uguale al diametro; in Figura 11 possiamo osservare che gli elementi facenti parte di tale insieme sono personaggi secondari, ad esempio gli agenti "George Merkert", "Janice", "Picarus", "Getz" oppure "Chris", uno dei membri della sicurezza di Los Pollos Hermanos. Il parametro "Average Clustering Coeff", invece, restituisce l'informazione relativa al clustering medio all'interno del grafo, ovvero, dice quanto mediamente il grafo riesce ad essere unito e a formare triadi con i nodi e gli archi che ha; quindi, il grafico riesce mediamente a formare triadi, avendo un valore di 0.65. Il grafo è connesso, come si evince dall'ultimo parametro, pertanto nessun nodo è isolato, di conseguenza tutti i nodi sono raggiungibili.

Una prima rappresentazione del grafo è stata svolta per mezzo dell'algoritmo *spring\_layout*. Si può osservare in Figura 12 che, come descritto precedentemente, sono presenti molti nodi con pochi collegamenti, visualizzabili agli estremi del grafo, e un numero ristretto di nodi con molti collegamenti. Per motivi estetici sono state rimosse le label associate ai nodi.

In Figura 13 abbiamo riportato un altro esempio di rappresentazione del grafico, utilizzando l'algoritmo *circular\_layout* che, ci conferma quanto detto nella visualizzazione del grafo precedente, ma lo mostra in un'altra forma, ovvero disponendo i nodi in modo circolare. Tra le due rappresentazioni è preferibile

## 2.2 Descrizione del grafo

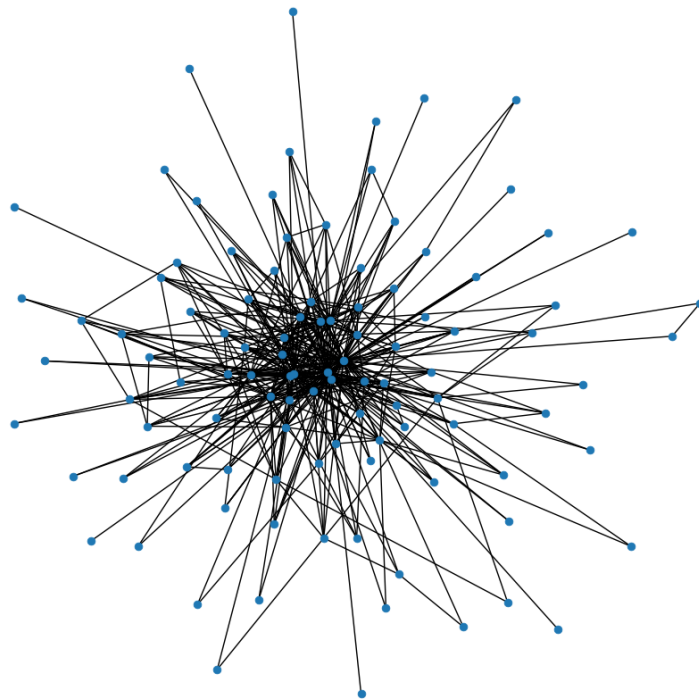


Figura 12: Spring layout

la prima poichè l'algoritmo euristico "Spring" tende ad avvicinare tra loro gruppi di nodi ben connessi, rendendo il risultato della visualizzazione più significativo ed esteticamente più gradevole.

Infine, si riporta un istogramma che mostra la distribuzione dei pesi degli archi nel presente dataset. Si noti che per questioni di scala il range dell'asse x è stato ridotto ad un limite di 300; tale operazione è stata svolta in quanto, come è visibile in figura 11, i pesi maggiori di 300 riguardano solo 5 archi. In Figura 14 si può osservare che la maggior parte degli archi ha peso inferiore a 20, pertanto è possibile dire che la maggior parte delle relazioni del dataset è debole.

## 2.2 Descrizione del grafo

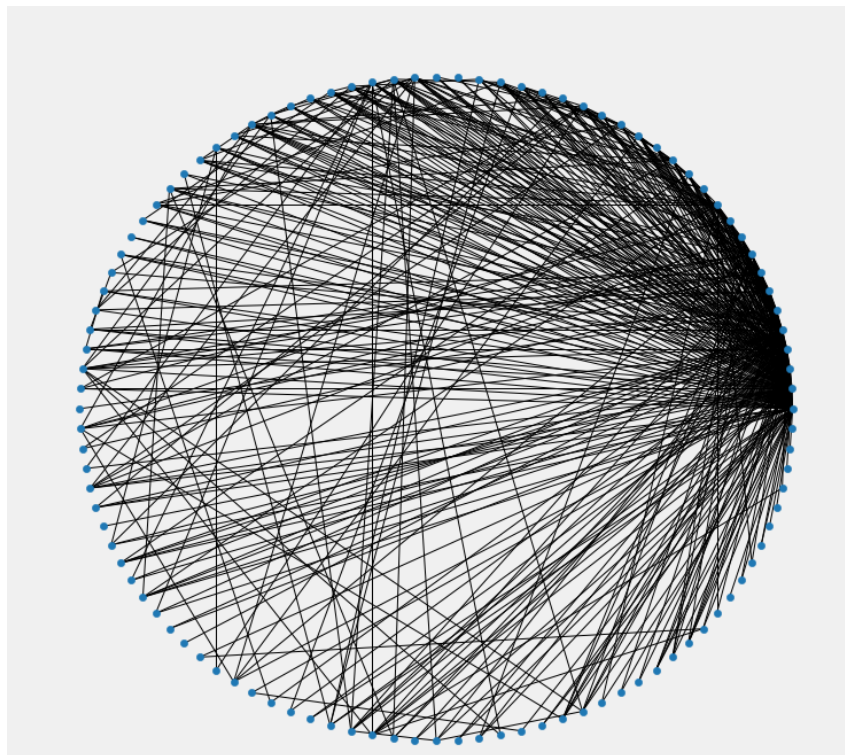


Figura 13: Circular layout

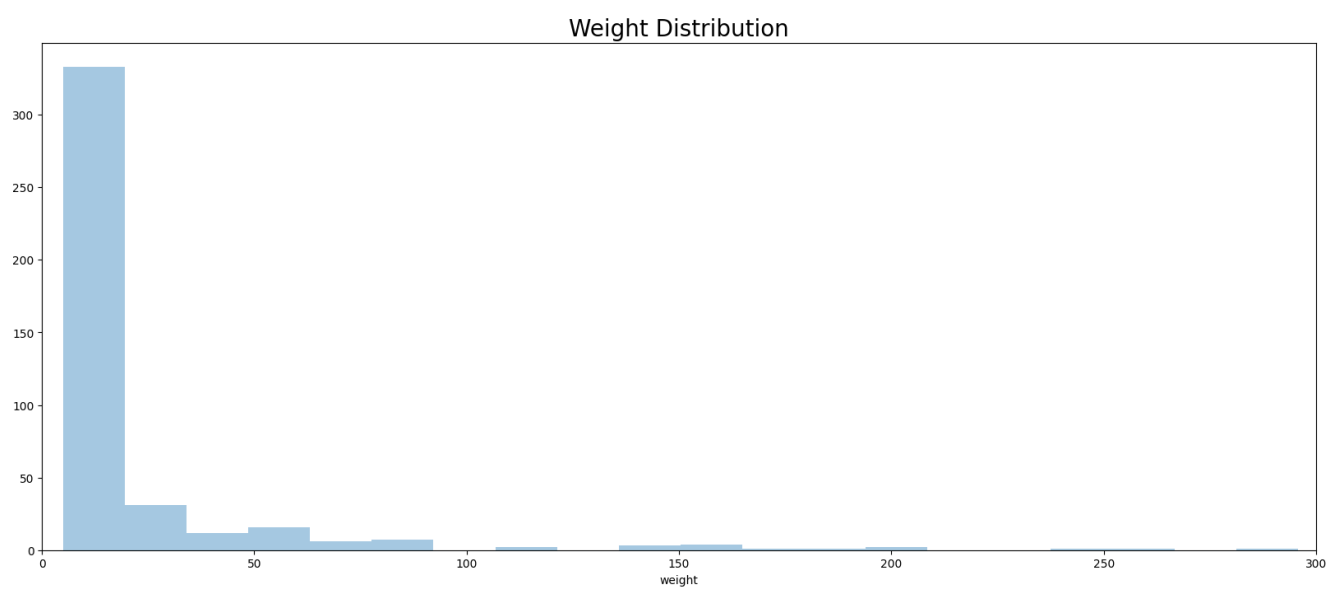


Figura 14: Conteggio peso degli archi

### 3 Le misure di Centralità

All'interno di una social network una delle prime analisi si focalizza sul quantificare il potere o l'influenza dei rispettivi componenti. Queste caratteristiche vengono valutate in base alle connessioni dei nodi, tuttavia non esiste un'unica misura che descriva esattamente il potere di un vertice rispetto agli altri all'interno della rete. Infatti, esistono diverse metriche, conosciute come "misure di centralità", che permettono di ricavare importanti informazioni sulla natura e la reciproca influenza dei nodi sotto diversi punti di vista.

#### 3.1 Degree centrality

La *degree centrality* parte dal presupposto che ciascuna comunità ha delle persone che sono più popolari di altre e, solitamente, queste persone rappresentano una minoranza dell'intero insieme. Questa metrica permette, infatti, di individuare quei nodi più connessi rispetto ad altri, cioè quei vertici che risultano incidenti da molti archi. Dunque, il grado di un nodo rappresenta il numero di connessioni che il nodo stesso possiede, ad esempio, su Twitter il numero che si ottiene è pari al numero dei follower. Tuttavia, come ricorda Scott Adamans *espero di social network*, il grado di un nodo non rappresenta necessariamente la sua forza all'interno di una organizzazione, anzi, secondo quest'ultimo, la forza di una persona è inversamente proporzionale al numero di archi incidenti.

Nel suddetto progetto è stata calcolata la *degree centrality* per tutti i nodi del grafo; in Figura 15 vengono riportati i primi 10 nodi con valore di centralità più alta e più bassa.

---

**Best 10 Degree Centrality:**

1. Walter: 0.7378640776699028
2. Jesse: 0.6407766990291262
3. Hank: 0.5242718446601942
4. Mike: 0.37864077669902907
5. Skyler: 0.3106796116504854
6. DEA: 0.24271844660194172
7. Saul Goodman: 0.2233009708737864
8. Gomez: 0.19417475728155337
9. Marie: 0.16504854368932037
10. Tuco Salamanca: 0.1553398058252427

**Worst 10 Degree Centrality:**

1. Chow: 0.009708737864077669
2. Picarus: 0.009708737864077669
3. Paul: 0.009708737864077669
4. Gardiner: 0.009708737864077669
5. Stephanie: 0.009708737864077669
6. Rebecca: 0.009708737864077669
7. Old Joe: 0.009708737864077669
8. Carmen: 0.009708737864077669
9. Carmen Molina: 0.009708737864077669
10. Don Eladio Vuento: 0.019417475728155338

Figura 15: Primi ed ultimi dieci nodi per degree centrality

Tra i primi 10 abbiamo ottenuto valori di centralità abbastanza elevati che, come prevedibile, coinvolgono i personaggi chiave della serie. Inoltre, è possibile notare che tra questi, sono presenti anche personaggi secondari che hanno una rilevanza nella storia come la DEA ("Drug Enforcement Administration"), l'agenzia delle forze dell'ordine incaricata di combattere il contrabbando e l'uso di droga negli Stati Uniti, che nel seguente dataset viene considerato come unico personaggio e Gomez, agente presso la DEA e partner di Hank. Tra i 10 peggiori, invece, troviamo alcuni dei personaggi che fanno parte della periferia del grafo.

In Figura 16 è riportato il grafico di distribuzione dei nodi all'interno della rete, secondo la metrica di degree centrality; è possibile notare che sono presenti pochi nodi con un valore di *degree centrality* alto, come del resto accade nella maggior parte delle social network. Tale risultato era prevedibile in quanto, dalle analisi precedenti, si è riscontrato che la rete ha una densità pari a 8%.

### 3.2 Closeness centrality

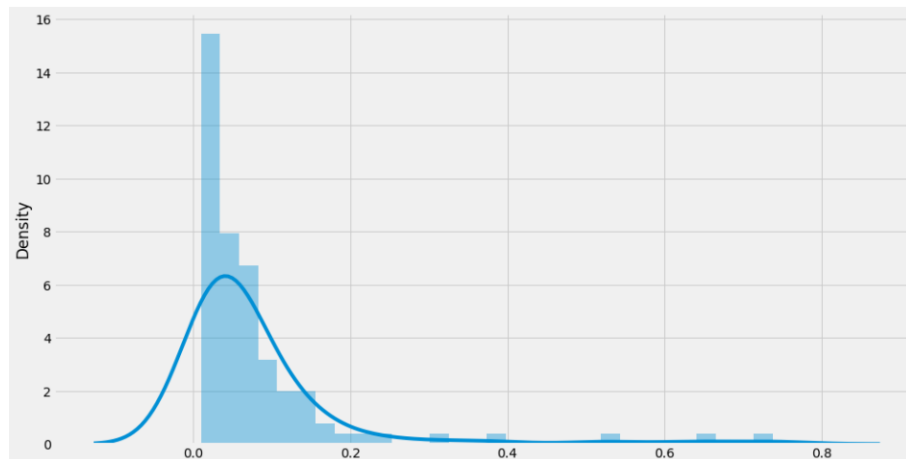


Figura 16: Distplot degree centrality

Si riporta anche la rappresentazione attraverso lo spring layout della *degree centrality* dei nodi della rete per avere un confronto con quanto anticipato precedentemente. I nodi centrali dello spring-layout hanno un valore di centralità alto, contraddistinti con una colorazione più chiara, in particolare, il nodo di colore giallo, sicuramente, corrisponderà a Walter. Andando verso l'esterno della rete, il valore di centralità diminuisce, ed il colore è più scuro ed è possibile osservare una predominanza di questi nodi, confermando le analisi svolte.

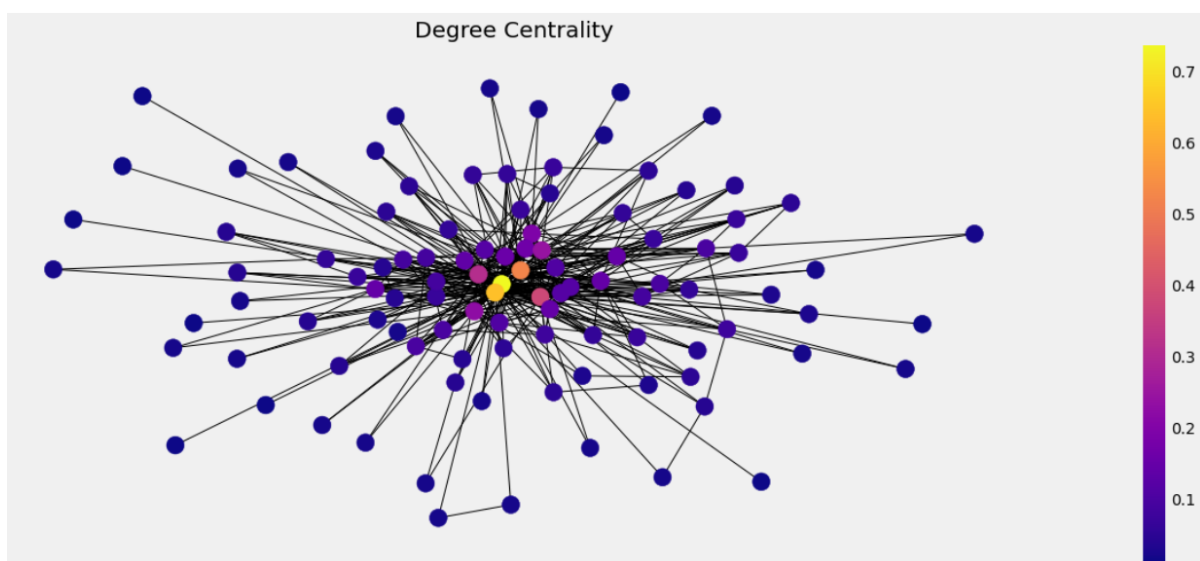


Figura 17: Spring layout - Degree centrality

### 3.2 Closeness centrality

La *Closeness centrality* misura la capacità di un nodo di ricevere e inviare informazione agli altri, cioè la facoltà di poter spostare l'informazione da un lato all'altro della rete. Questo tipo di centralità, definisce un nodo tanto più centrale quanto minore è la media delle sue distanze dagli altri nodi. In questo senso, la distanza o la vicinanza dagli altri può definire il ruolo di una entità all'interno di una rete. Il costo della *Closeness centrality* è piuttosto oneroso perché utilizza l'algoritmo di Dijkstra per calcolare il percorso più breve tra ciascuna coppia di nodi. Come nel caso precedente vengono riportati i primi dieci personaggi con *Closeness centrality* più alta ed i primi dieci con valore più basso.

### 3.2 Closeness centrality

```
Best 10 Closeness Centrality:

1. Walter: 0.7923076923076923
2. Jesse: 0.7357142857142858
3. Hank: 0.673202614379085
4. Mike: 0.6094674556213018
5. Skyler: 0.5919540229885057
6. DEA: 0.5659340659340659
7. Saul Goodman: 0.5597826086956522
8. Gomez: 0.544973544973545
9. Marie: 0.533678756476684
10. Tuco Salamanca: 0.5282051282051282

Worst 10 Closeness Centrality:

1. Picarus: 0.3366013071895425
2. Soper: 0.34915254237288135
3. Stephanie: 0.34915254237288135
4. Don Eladio Vuento: 0.3515358361774744
5. Getz: 0.37184115523465705
6. Chow: 0.3800738007380074
7. Patrick: 0.3828996282527881
8. Patrick Kuby: 0.38721804511278196
9. Chris: 0.3886792452830189
10. Don: 0.40551181102362205
```

Figura 18: Primi ed ultimi dieci nodi per closeness centrality

Nella prima lista, i personaggi con più alto valore sono gli stessi della Degree Centrality ed è abbastanza prevedibile in quanto questi, nella serie, sono in contatto con la maggior parte dei nodi appartenenti alle diverse realtà presenti nella serie. Questo permette di poter arrivare con facilità a tutti quei nodi per i quali non esiste un collegamento diretto.

Nella seconda lista, invece, sono riportati i personaggi con i valori più bassi di closeness e si può notare che quelli coinvolti sono i personaggi della periferia del grafo, nonché personaggi di contorno nella serie.

In Figura 19 è riportato il grafico di distribuzione dei nodi all'interno della rete secondo la metrica di *Closeness centrality*; è meno sbilanciata rispetto alla *Degree centrality*.

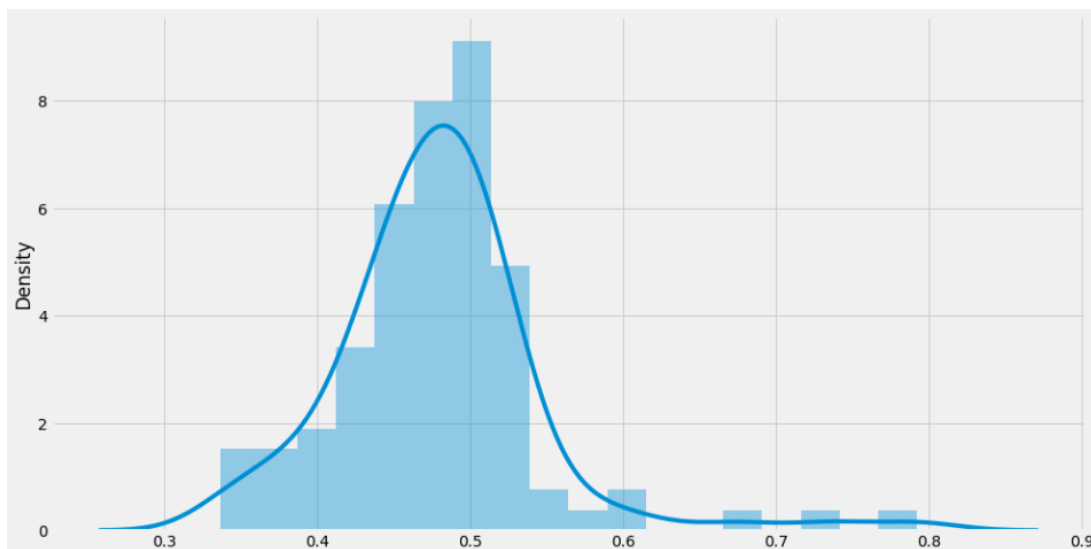


Figura 19: Distplot Closeness centrality

Si osserva che i valori sono compresi prevalentemente tra 0,35 e 0,65, valori piuttosto alti che ci permettono di dedurre una buona capacità, da parte dei nodi, di trasmettere l'informazione per tutta la rete. Quanto detto è facilmente riscontrabile dallo spring-layout di Figura 20, in cui si evidenzia una

### 3.3 Betweenness centrality

predominanza di nodi con una colorazione viola, contrariamente allo spring layout visto per la degree, in cui la predominanza dei nodi è di una colorazione più scura.

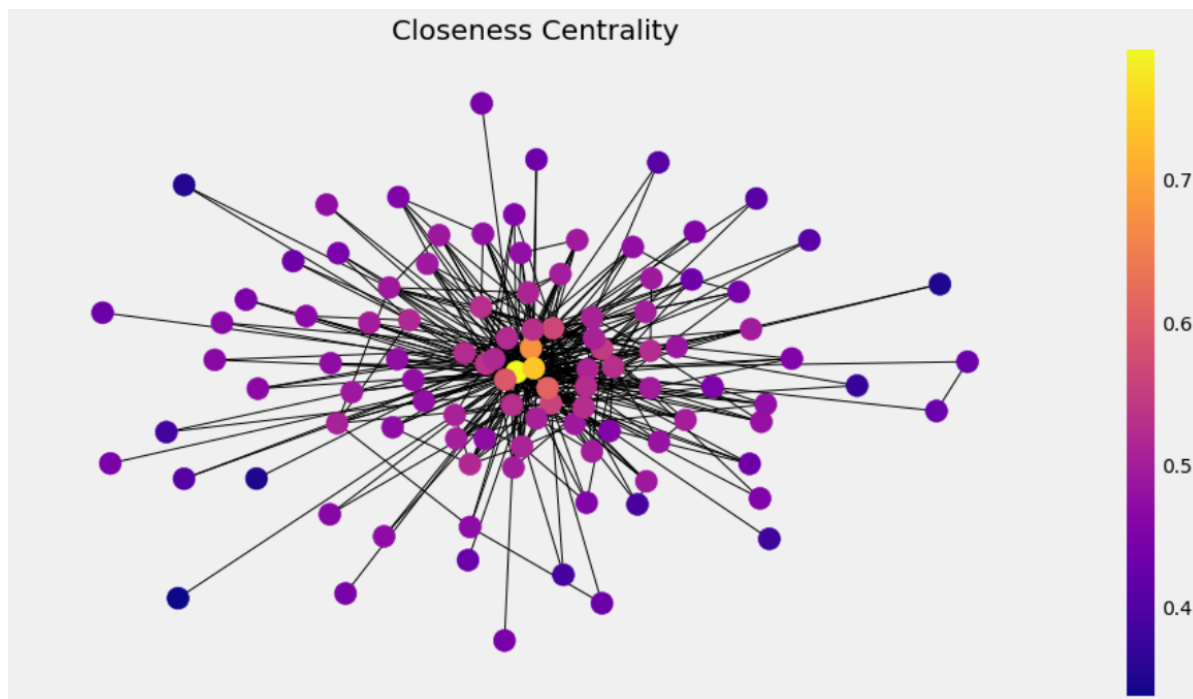


Figura 20: Spring layout - Closeness centrality

### 3.3 Betweenness centrality

La *Betweenness centrality* è una metrica che permette di misurare il potere o l'influenza di un individuo all'interno di una rete considerando un aspetto differente dalle misure viste fino ad ora. Infatti, la *Betweenness centrality* si fonda sull'assunzione che il potere di un nodo all'interno di una rete è tanto maggiore quanto questo assuma una posizione di ponte tra due o più gruppi di nodi. Dunque, fungere da collo di bottiglia per una comunicazione garantisce una posizione di potere, in quanto tutte le comunicazioni tra i vari i nodi devono necessariamente passare attraverso tale vertice strategico. Questa metrica è calcolata come il rapporto tra il numero di shortest path, calcolato tramite l'uso dell'algoritmo di Dijkstra, che passano per un nodo e il numero di shortest path totali nella rete. Il valore risultante sarà pari ad un numero compreso tra 0 ed 1.

Come fatto per le precedenti centralità, sono stati calcolati i primi dieci personaggi con *Betweenness centrality* più alta e gli ultimi dieci protagonisti con valore più basso.

Dalla Figura 21 è facile osservare che il 35% degli shortest path passa per il nodo corrispondente al personaggio Walter, il quale risulta il punto più strategico per far comunicare le parti diverse della rete che nasconde le molteplici realtà della serie, come i servizi antidroga, il "Cartello", "Los Pollos Hermanos", la banda di Jack Welker, che saranno protagonisti di diversi momenti caratterizzanti della storia. Nella stessa lista, a seguire, è possibile trovare alcuni dei personaggi chiave della serie che sono stati già incontrati nelle analisi precedenti, ad eccezione di Badger e Ted Beneke. Questi ultimi, sebbene siano secondari, risultano dei "colli di bottiglia" in quanto, attraverso i loro collegamenti, riescono ad arrivare a personaggi difficilmente raggiungibili nella rete.

Nella seconda lista, i personaggi con Betweenness centrality peggiore registrano il valore 0, in quanto il valore di centralità raggiunge il milionesimo, dunque, un numero così piccolo che l'algoritmo lo approssima direttamente a 0.

Tramite l'istogramma in figura 22, con un comportamento *power-low*, si evince, una predominanza di valori bassi che investono quasi la totalità dei nodi e questo permette di dedurre che i "colli di bottiglia" sono nel complesso pochi.

### 3.3 Betweenness centrality

#### Best 10 Betweenness Centrality:

```
1. Walter:0.35674512524399415
2. Jesse:0.26584441801605646
3. Hank:0.14937390274079695
4. Mike:0.08700902547573651
5. Skyler:0.04158049582470922
6. Saul Goodman:0.0327304736540751
7. Marie:0.025644558215789613
8. Badger:0.02160743967405011
9. Ted Beneke:0.019720844599485376
10. Gomez:0.019691787190923023
```

#### Worst 10 Betweenness Centrality:

```
1. Picarus: 0.0
2. Carmen Molina: 0.0
3. Paul: 0.0
4. Don: 0.0
5. Carmen: 0.0
6. Rebecca: 0.0
7. Peter Schuler: 0.0
8. Gang: 0.0
9. Carol: 0.0
10. Old Joe: 0.0
```

Figura 21: Primi ed ultimi dieci nodi per Betweenness centrality

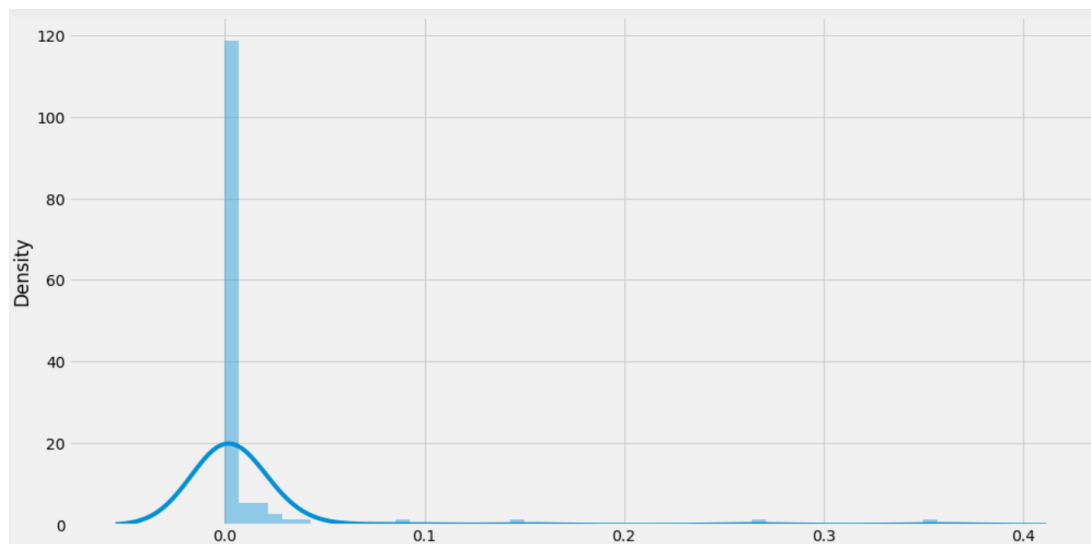


Figura 22: Distplot Betweenness centrality

Questo è ulteriormente dimostrato dalla rappresentazione attraverso lo *spring layout* in Figura 23, il quale evidenzia la preponderanza di nodi blu scuro; pochi nodi hanno una colorazione più chiara e sono disposti al centro della rete.



### 3.4 Eigenvector Centrality

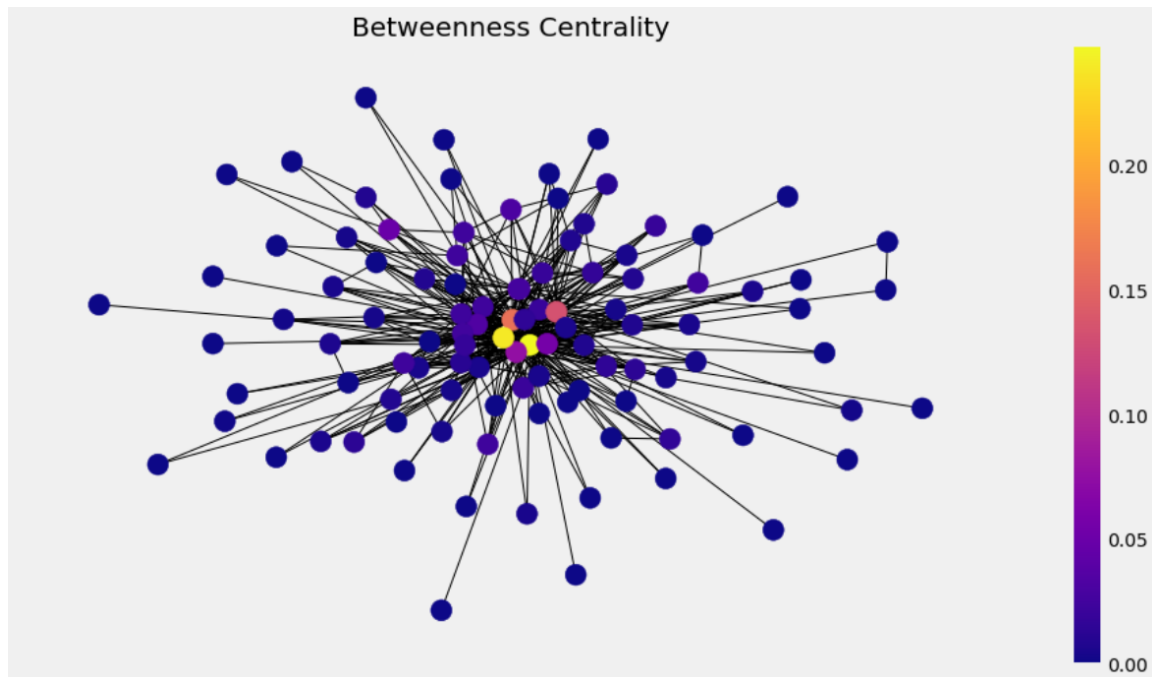


Figura 23: Spring layout - Betweenness centrality

### 3.4 Eigenvector Centrality

La *Eigenvector Centrality* nasce per scoprire quei nodi che rivestono il ruolo di una eminenza grigia. Una eminenza grigia è, per definizione, "un consigliere o un decisore potente che opera segretamente e in modo non ufficiale". Questi tipi di nodi, infatti, sono difficilmente individuabili nel grafo in quanto hanno poche connessioni con il resto della rete; tuttavia, tali legami esistono tra nodi altamente connessi e questo permette alle "eminenze grigie" di avere un'enorme influenza e potere. Dunque, questa forma di centralità non dipende solo dal numero di legami, ma da quanto è centrale ciascun legame. Per trovare questa tipologia di nodi nella rete, Philip Bonancich propose che, invece di contare il numero di link che connettono un nodo, sarebbe stato necessario effettuare una media ponderata, pesando ciascun link sulla base del grado del nodo adiacente. La *Eigenvector Centrality*, sfrutta un algoritmo ricorsivo e, per tale ragione, è facilmente calcolabile per grafi piccoli, come il caso del grafo prodotto dal dataset considerato in tale elaborato.

Sono state effettuate le stesse analisi svolte per le centralità calcolate nei paragrafi precedenti, con l'unica differenza che, in questo caso, sono stati presi in considerazione i pesi degli archi del grafo. In Figura 24 sono visualizzati i primi dieci personaggi con *Eigenvector centrality* più alta e gli ultimi dieci protagonisti con valore più basso. Dalla prima lista si evince che i personaggi con elevata centralità risultano gli stessi incontrati precedentemente ad esclusione di Jack Welker e Jane Margolis. Il primo, anche soprannominato "zio Jack", sebbene compaia solo nella quinta stagione della serie, riveste un ruolo di potere in quanto è il capo della banda criminale nota come "White Power Prison Gang", nel New Mexico, produttrice e spacciatrice di metanfetamina. Jack Welker diventerà socio di Walter White e sarà lui il responsabile di uno degli eventi più imprevedibili e sconvolgenti della serie, ovvero l'omicidio di Hank. Jane Margolis, invece, è stata la fidanzata di Jesse Pinkman per metà della seconda stagione ed è stata per lui una persona piuttosto influente, infatti, grazie a lei, lui ricominciò a drogarsi. Successivamente, a causa di un overdose, Jane morì sotto gli occhi di Walter White, che però non fece nulla per salvarla. La sua morte continuò ad avere ripercussioni significative su Jesse anche nelle stagioni successive.

Per quanto riguarda la seconda lista riportata nella stessa figura, i personaggi evidenziati sono quelli hanno un'importanza strategica più bassa.

Tramite l'istogramma in Figura 25 è possibile notare una predominanza di valori di centralità bassi. Questo è dovuto al fatto che per tale centralità è stato utilizzato il peso dell'arco, che permette di avere una maggiore consapevolezza della connessione tra i personaggi. I valori di centralità della rete, infatti, sarebbero decisamente più alti se gli archi fossero pesati allo stesso modo.

### 3.4 Eigenvector Centrality

#### Best 10 Eigenvector Centrality (Weight):

1. Walter:0.6614850489210113
2. Jesse:0.5942432468337115
3. Hank:0.255828902743377
4. Skyler:0.2411106136922531
5. Mike:0.1831926095944995
6. Saul Goodman:0.11380297323970585
7. Marie:0.09473257154677571
8. Tuco Salamanca:0.08649257036342074
9. Jack Welker:0.05713181843913594
10. Jane Margolis:0.05282049936343828

#### Worst 10 Eigenvector Centrality (Weight):

1. Don Eladio Vuento:3.6446105261828574e-05
2. Picarus:5.683327658777403e-05
3. Soper:5.836045568125407e-05
4. Getz:0.00032051750949792186
5. Patrick:0.00033913195589908493
6. Stephanie:0.00040258272758804923
7. Don:0.00045816011030401326
8. Gang:0.0005396881267551011
9. Ortega:0.0005857212254551449
10. Janice:0.000615836258909602

Figura 24: Primi ed ultimi dieci nodi per Eigenvector centrality

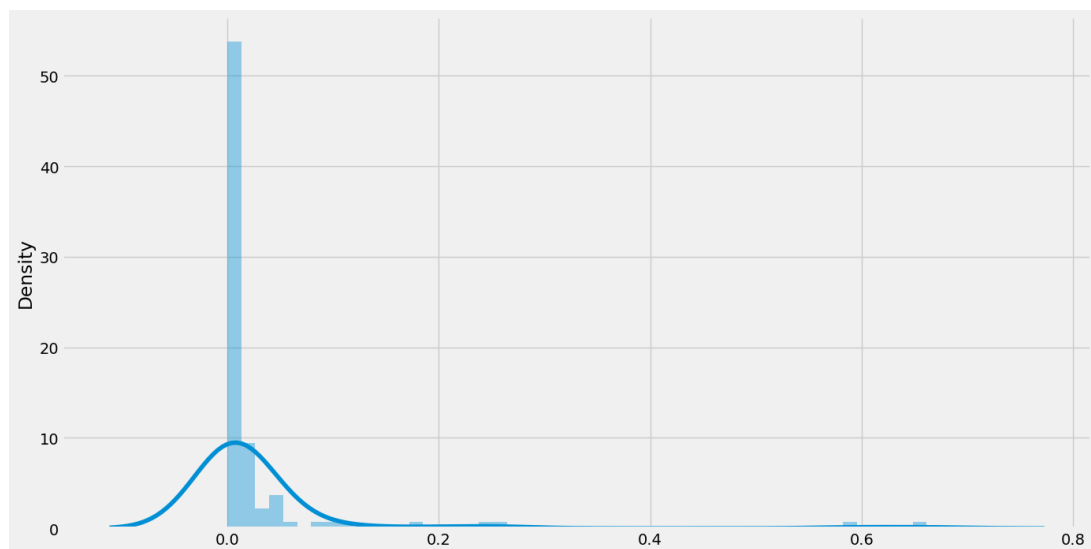


Figura 25: Distplot Eigenvector centrality

Per concludere l'analisi della *Eigenvector Centrality*, è stata riportata la rappresentazione attraverso lo *spring layout*, in Figura 26, in cui si evidenzia una concentrazione di nodi blu scuro dall'esterno all'interno della rete; i nodi tendono a schiarirsi verso il centro, ma il numero di nodi di colorazione più chiara evidenti sono pochi.

### 3.4 Eigenvector Centrality

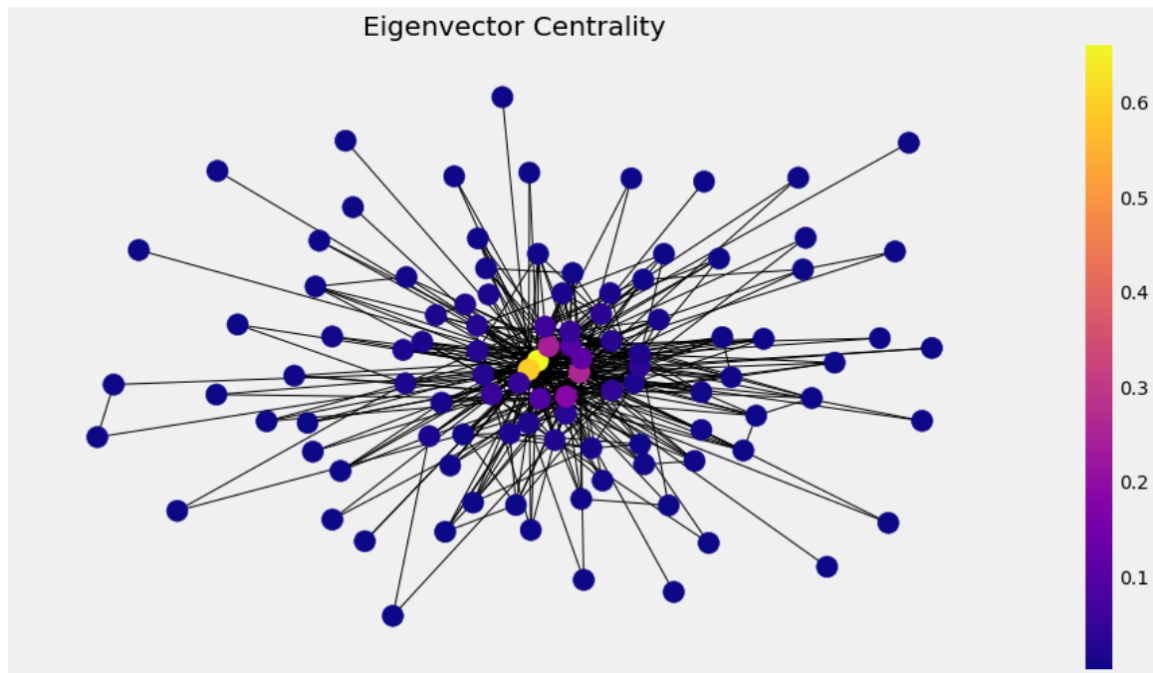


Figura 26: Spring layout - Eigenvector centrality

## 4 La visualizzazione e l'analisi delle strutture

In questo paragrafo sono stati analizzati i sottografi della rete, in particolare, le triadi, la clique massima, il K-core e l'ego network incentrata sul protagonista Walter White.

### 4.1 Le triadi

Una triade è un insieme di tre nodi; se il grafo è non diretto esistono quattro possibili combinazioni di triadi:

- triade chiusa, ovvero i tre nodi sono totalmente connessi tra di loro;
- triade aperta, ovvero due nodi non comunicano tra loro, però possono comunicare attraverso il terzo;
- una coppia di nodi è connessa mentre l'altro nodo non lo è;
- tutti i nodi sono sconnessi dagli altri.

I sociologi hanno dimostrato che la triade chiusa è la struttura più stabile che si possa immaginare dal punto di vista sociale, soprattutto per quel che riguarda la diffusione delle informazioni in quanto, se la connessione tra due qualsiasi nodi dovesse interrompersi, l'informazione comunque potrebbe arrivare a tutti e tre i nodi.

In tale progetto le triadi trovate dall'algoritmo elaborato sono 163 e tale informazione è riportata in Figura 27; nella stessa immagine è possibile visualizzare le prime 10 triadi con peso maggiore.

```
Triads:163
Top 10 Triads (weight based):

1. 'Walter', 'Hank', 'Jesse'(3142)
2. 'Skyler', 'Walter', 'Jesse'(3116)
3. 'Mike', 'Walter', 'Jesse'(2959)
4. 'Saul Goodman', 'Walter', 'Jesse'(2641)
5. 'Walter', 'Jesse', 'Tucó Salamanca'(2542)
6. 'Walter', 'Jack Welker', 'Jesse'(2427)
7. 'Marie', 'Walter', 'Jesse'(2424)
8. 'Walter', 'Jane Margolis', 'Jesse'(2416)
9. 'Walter', 'Todd Alquist', 'Jesse'(2387)
10. 'Gale Boetticher', 'Walter', 'Jesse'(2357)
```

Figura 27: L'analisi delle triadi

Tutte le triadi sono composte dai personaggi Walter e Jesse, e questo era piuttosto prevedibile in quanto risultano essere i personaggi principali, ed il peso del loro arco è 2'185, quindi permette di alzare di molto il peso complessivo della triade. In particolare, la prima triade trovata dall'algoritmo, nonché la più forte, è composta da Walter, Hank e Jesse, come ci si poteva aspettare dato che la storia ruota intorno a questi personaggi, i quali, per l'appunto, compaiono in tutte le stagioni. Tutti i personaggi presenti nelle triadi sono stati già incontrati nell'analisi delle centralità, gli unici che per ora non erano stati mai incontrati sono Todd Alquist e Gale Boetticher. Todd Alquist comparirà nella quinta stagione, in cui inizierà a lavorare per Walter, Mike e Jesse nella loro attività di produzione di metanfetamine. I collegamenti che lui ha sia con Jesse che con Walter hanno un peso non indifferente in quanto risulta coinvolto in diversi episodi, in particolare, presenterà Jack Welker, nonché suo zio, a Walter. Gale Boetticher, invece, è il chimico assunto da Gustavo Fring per allestire il laboratorio sotterraneo di metanfetamine ed aiutare Walter White. Questo personaggio compare nella terza e quarta stagione e gli archi che lo collegano a Walter e Jesse hanno un peso elevato in quanto saranno particolarmente in contatto nella produzione della droga, finché Jesse non sarà costretto ad ucciderlo.

## 4.1.1 Clique

Una clique è un insieme di nodi totalmente connessi tra di loro ed è possibile affermare che una triade totalmente connessa è una clique di dimensione 3. Essendo una triade totalmente connessa una struttura stabile, la clique è una struttura ancor più forte, infatti è piuttosto complesso trovarla.

Per il suddetto grafo, è stato elaborato un algoritmo per la ricerca del numero di clique con almeno 4 nodi e per la visualizzazione della clique massima. Come è riportato in Figura 28, il numero di clique con almeno 4 nodi sono 110; la clique massima ha 8 nodi. Nella stessa immagine è riportata la clique massima, in cui si evince una nuova figura: Lydia. Lei è la responsabile della logistica dell'azienda "Madrigal Electromotive", che aveva come filiale "Los Pollos Hermanos", dunque anche lei si trova coinvolta nelle attività illegali di Gus Fring. In seguito alla morte di Gus, alla fine della quarta stagione, e con il conseguente sgretolamento del suo impero, lei inizia a collaborare sia con Walter White che con Jack Welker.

---

Numero di clique con almeno 4 nodi: 110  
 Numero di nodi della clique massima: 8

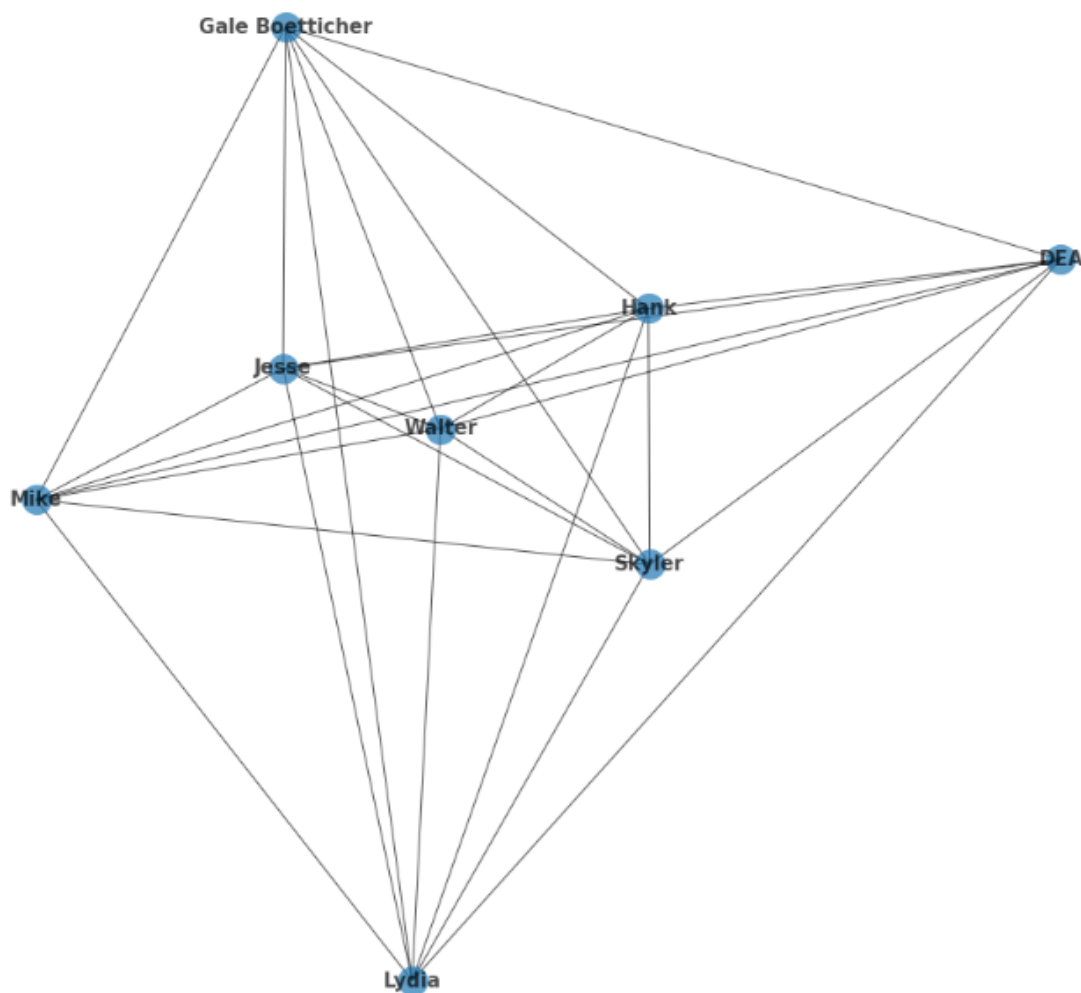


Figura 28: L'analisi delle clique

Tale clique riassume i collegamenti principali che poi porteranno Hank a sospettare dell'impero della droga costruito da Gus Fring che, in seguito alla sua morte, verrà piano piano smascherato, portando i personaggi davanti a dei vicoli ciechi. Per di più, mediante le indagini sempre più approfondite di Hank,

## 4.2 Il K-core

alla fine lui scoprirà che Walter White è l'uomo che, in verità, ha sempre cercato, in particolare, grazie al suo legame con Gale Boetticher.

Inoltre, è interessante sottolineare che Mike, Gale Boetticher e Lydia sono collegati a Walter White e Jesse grazie a Gus Fring, che, allo stesso tempo entrerà in contatto anche con Hank e la DEA, prima che venga ucciso e scoperte le sue attività. Ci si sarebbe aspettati, infatti, la presenza di un'altra clique massima con la stessa struttura, ma che al posto di Skyler ci fosse Gus Fring. Ma come si evince dalla Figura 29, l'output del codice precedente riporta che il numero di clique con almeno 8 nodi ne è solo una.

```
Numero di clique con almeno 8 nodi: 1
Numero di nodi della clique massima: 8
```

Figura 29: L'analisi della clique massima

Approfondendo si è scoperto che, per questa figura, il dataset risulta incompleto e lo si può verificare in Figura 30, dove sono presenti tutti gli archi in cui un nodo è Gus Fring. Si evince che alcuni collegamenti importanti della storia non sono presenti.

	source	target	weight
156	Walter	Gus Fring	10
163	Gus Fring	Mike	10
180	Jesse	Gus Fring	6
234	Gus Fring	Dan Wachsberger	6
276	Gomez	Gus Fring	6
408	Hank	Gus Fring	5

Figura 30: I collegamenti di Gus Fring

Ad ogni modo Gus Fring è un personaggio principale della serie e la sua influenza nella storia è fondamentale.

## 4.2 Il K-core

Un K-core è un sottografo massimale che contiene nodi di grado  $k$  o maggiore di  $k$ ; tale struttura è più debole rispetto alla clique in quanto, quest'ultima, possiede qualche arco in meno.

Per il suddetto grafo, è stato elaborato un algoritmo tale da visualizzare il sottografo in questione, in particolare, è stato calcolato il *main core* del grafo di partenza, ovvero il core con grado maggiore. In Figura 31 è visualizzata la struttura appena descritta; i nodi coinvolti sono 14 e il grado di ciascun nodo è pari almeno a 9. I 6 personaggi in più rispetto a quelli visualizzati nella clique descritta nel paragrafo precedente sono: Todd Alquist, Jack Welker, Saul Goodman, Gomez, Marie e Walter Jr (il figlio di Walter e Skyler). I nodi completamente connessi sono quelli che corrispondono alle figure di Walter, Jesse e Hank. Tra le vicende che riassumono i collegamenti presenti in questa struttura, ci sono episodi che riguardano la quinta stagione in quanto, solo in questa, compaiono i personaggi Lydia, Todd Alquist e Jack Welker. In particolare, in questa struttura si evincono i legami legati agli eventi che seguono la scoperta di Hank su Walter.

## 4.3 Le ego network

Una ego network è una rete incentrata su un personaggio; il nodo centrale è chiamato *ego*, mentre i suoi vicini sono chiamati *alters*. Gli archi collegano gli ego con gli alters, oltre a collegare gli alters tra di loro. Questa struttura si basa sul concetto di omofilia, ovvero la tendenza degli individui socialmente connessi a mostrare determinati tratti o somiglianze tra loro.

E' stato elaborato un algoritmo per la visualizzazione dell'ego network che avesse come centro Walter White, e un numero di archi con peso pari almeno a 45, come mostrato in Figura 32. Il peso è stato scelto in modo da visualizzare i legami più significativi oltre quelli già analizzati.

### 4.3 Le ego network

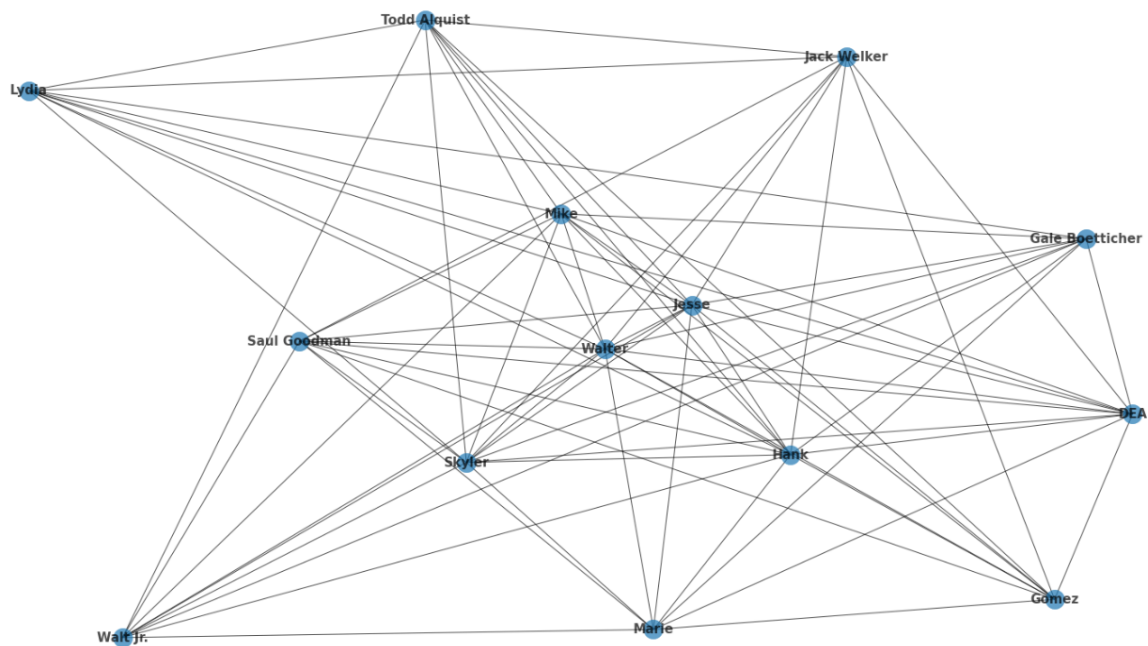


Figura 31: L'analisi della k-core

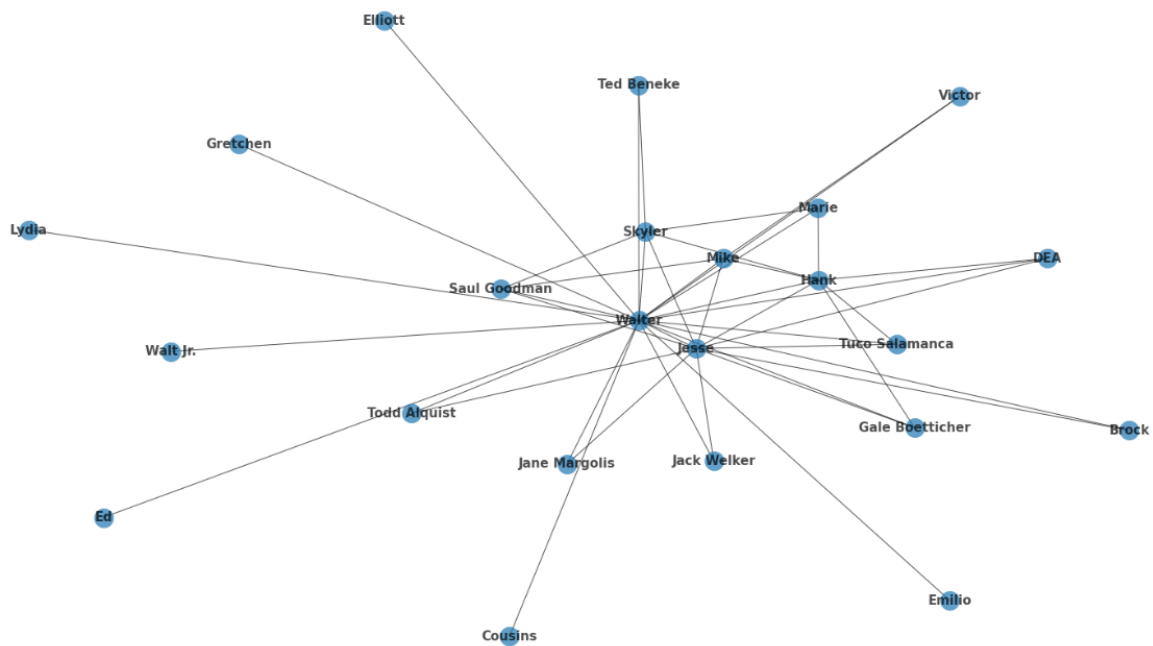


Figura 32: L'analisi della ego network

È possibile notare che il 30% della totalità dei nodi è collegato esclusivamente a Walter, il restante 70% è interconnesso anche con altri nodi del grafo rappresentato in figura. Inoltre, si nota che tutti i personaggi corrispondenti ai nodi presenti, sono stati significativi non solo per il susseguirsi degli eventi ma anche il cambiamento che ha contraddistinto Walter White nel corso della storia, portandolo ad essere uno spregiudicato criminale disposto a tutto pur di raggiungere il suo obiettivo.