

Social Network Analysis della serie

RICK AND MORTY

Autori:

Chiara Amalia Caporusso
Margherita Galeazzi
Simone Scalella
Zhang Yihang



Sommario

1	Introduzione	2
1.1	La Social Network Analysis (SNA)	2
1.2	Rick And Morty Social Network.....	2
1.2.1	Creazione del dataset	2
1.3	Strumenti utilizzati	4
1.3.1	Beautiful Soup	4
1.3.2	NetworkX.....	4
1.3.3	Gravis	4
2	Analisi descrittiva.....	5
2.1	Metriche generali	6
2.2	Visualizzazioni e layout.....	6
3	Analisi delle centralità.....	9
3.1	Degree Centrality	9
3.2	Descrizione dei personaggi.....	11
3.2.1	Rick Sanchez	11
3.2.2	Morty Smith.....	12
3.2.3	Summer Smith.....	12
3.3	Closeness Centrality	13
3.4	Eigenvector Centrality	15
3.5	Betweenness Centrality.....	16
4	Analisi delle strutture	19
4.1	Triadi.....	19
4.2	Clique	19
4.3	K-core.....	21
4.4	Ego Network.....	22
4.4.1	Rick.....	22
4.4.2	Morty	23
5	Group Centrality	24



1 Introduzione

Il seguente progetto ha come scopo l'analisi una rete sociale basata sui personaggi di una nota serie tv, con l'obiettivo di trarre conoscenza da un insieme di dati che a primo impatto non è di facile lettura, come è appunto una rete.

1.1 La Social Network Analysis (SNA)

La Social Network Analysis costituisce una parte di spicco della Data Analytics, essa studia il comportamento degli individui a livello macroscopico e la struttura delle relazioni (che possono essere di diverso tipo) che sussistono tra gli individui e la correlazione che esiste tra questi due aspetti. Si può quindi dire che sostanzialmente la SNA sia: "lo studio delle relazioni umane per mezzo della teoria dei grafi".

La Social Network Analysis nasce molto prima dell'avvento dei Social Network, ma deve ad essi il suo sviluppo in quanto questi danno accesso ad un'enorme mole di dati, ed è qui che entrano in gioco le potenzialità della Data Science.

La SNA come detto in precedenza è basata sulla teoria dei grafi; perciò, gli elementi da studiare (potrebbero essere persone, dispositivi o altro) sono rappresentati sotto forma di nodi e le relazioni tra essi sono rappresentate da archi.

Questo mapping della rete su di un grafo permette di estrarre importanti proprietà dalla rete, quali il potere o l'influenza dei nodi, la coesione tra i componenti e la presenza di particolari pattern tra essi.

1.2 Rick And Morty Social Network

Il dataset su cui effettueremo la SNA contiene la rete di relazioni tra i personaggi della famosa serie "Rick And Morty".

Rick And Morty è una serie animata statunitense creata da Justin Roiland e Dan Harmon, di genere cosmic horror. La serie è composta da 6 stagioni (in Italia la 6° è in uscita in questo periodo), nelle quali si raccontano le avventure dello "scienziato" Rick e di suo nipote Morty.

Un dataset basato su questo mondo risulta interessante per l'analisi SNA, in quanto nella storia compaiono numerosi personaggi, provenienti da diverse dimensioni e di diverse specie che nei vari episodi sono portati a combattere insieme o a scontrarsi.

Nella Tabella 1.1 è riportato una parte del dataset, così da mostrare la sua struttura:

src	dest
Rick Sanchez	Beth Sanchez
Rick Sanchez	Diane Sanchez
...	...
Jessica (Cronenberged Dimension)	Jessica's Grandma

Tabella 1.1 - Dataset relazioni

dove src indica il nodo di partenza e dest il nodo di arrivo. Quindi ogni riga del dataset rappresenta un arco del grafo, ovvero una relazione che coinvolge due personaggi, i cui nomi appaiono negli attribuiti src e dest.

1.2.1 Creazione del dataset

Il dataset di cui si è parlato sopra non è stato scaricato da internet, in quanto non era presente un dataset con i dati che erano per noi di interesse; perciò, abbiamo deciso di "costruirlo" nella maniera che riportiamo di seguito.

I dati sono stati raccolti dal sito: [Rick And Morty Wiki](#), contenente svariate pagine relative ai vari personaggi della serie e contenenti informazioni su di essi, inoltre abbiamo appreso che tale sito consentiva la navigazione mediante un crawler (per fare ciò basta aggiungere a qualunque URL "robots.txt/").

Il primo step è stato quello di creare un dizionario contenente il titolo della pagina e il link ad essa (Figura 1.1).



SOCIAL NETWORK ANALYSIS RICK AND MORTY

```
{'Rickipedia': 'https://rickandmorty.fandom.com/wiki/Rickipedia',  
 'Rick_Sanchez': 'https://rickandmorty.fandom.com/wiki/Rick\_Sanchez',  
 'Morty_Smith': 'https://rickandmorty.fandom.com/wiki/Morty\_Smith',  
 ...  
 'Wedding_Bartender': 'https://rickandmorty.fandom.com/wiki/Wedding\_Bartender',  
 ...}
```

Figura 1.1 - Dizionario 'titolo':'url'

Ci siamo serviti della libreria Beautiful Soup che permette l'estrazione di dati da file HTML e XML e mediante un ciclo for, per ogni elemento del dizionario sopra illustrato abbiamo preso solamente le parti contenenti le relazioni e la famiglia e lo abbiamo trascritto in un file csv.

The screenshot shows the Fandom page for 'Jessica'. At the top, there's a navigation bar with links for 'Characters', 'Female Characters', 'Morty's Love Interests', and '3 more'. Below the title 'Jessica' is the subtitle 'NOME DEL PERSONAGGIO'. To the right are 'VIEW SOURCE' and a three-dot menu. The main content area features a portrait of Jessica with orange hair and a blue shirt. To the left is a sidebar with a 'Contents' table of contents and a 'FAMIGLIA' section. The main content area has sections for 'Biographical Information', 'RELAZIONI', and 'Meta Information'.

in: Characters, Female Characters, Morty's Love Interests, and 3 more English ▾

Jessica

NOME DEL PERSONAGGIO

[VIEW SOURCE](#) | [⋮](#)

Jessica W.^[2] is a student at Harry Herpson High School. She is a popular girl in Morty Smith's math class with whom he desires to be in a romantic relationship. Much to Morty's dismay, she is frequently in a relationship with Brad, though she experiences difficulties with him and the two have been shown to have an on-again-off-again relationship.

Contents [hide]

- 1. Appearance
- 2. Personality
- 3. Biography
- 4. Relationships
 - 4.1. Morty
 - 4.2. Brad
 - 4.3. Rick Sanchez
 - 4.4. Grace Smith
 - 4.5. Tricia Lange
- 5. Episode Appearances
 - 5.1. Episodes
 - 5.1.1. Season 1
 - 5.1.2. Season 2
 - 5.1.3. Season 3
 - 5.1.4. Season 4

FAMIGLIA

RELAZIONI

Biographical Information

SPECIES	Human
AGE	Teens
STATUS	Alive
AFFILIATION	Harry Herpson High School
FAMILY	Jessica's Grandma (grandmother; deceased)
RELATIONSHIP	Brad (ex-boyfriend) Morty Smith (brief love interest) ^[1]

Meta Information

FIRST SEEN IN	Pilot
VOICE ACTOR	Kari Wahlgren

Figura 1.2 - Elementi di interesse della pagina del personaggio



SOCIAL NETWORK ANALYSIS RICK AND MORTY

```
#tra i <div> risultanti va a prendere solo quelli relativi alla famiglia o alle relazioni
for elem in soup.find_all('div' ,class_='pi-item pi-data pi-item-spacing pi-border-color'):
    first_child=elem.contents
    if 'FAMILY' in str(first_child):
        famRel.append(first_child)
    elif 'RELATIONSHIPS' in str(first_child):
        famRel.append(first_child)
```

Figura 1.3 - Selezione degli elementi di interesse

Il dataset che ne è risultato è quello che abbiamo esplicato nel paragrafo precedente.

1.3 Strumenti utilizzati

1.3.1 Beautiful Soup

Abbiamo utilizzato la libreria Beautiful Soup, per poter estrarre i dati di interesse dalle pagine web dei vari personaggi.

A seguito dell'import di tale libreria, si hanno a disposizione classi e metodi per gestire lo scraping di dati da file di tipo HTML e XML, quindi permette di ricercare ed estrarre tag o il testo in essi presente, di effettuare il filtraggio in base al tipo di tag e così via...

1.3.2 NetworkX

Per poter realizzare il progetto utilizzando l'ambiente Python, si è deciso di utilizzare la libreria **NetworkX**, lo strumento ideale per la social network analysis.

Questa libreria mette a disposizione classi e metodi per gestire e manipolare grafi di diverso tipo, tra cui grafi diretti, indiretti, multigrafi e tanti altri. Tramite questa libreria è possibile anche convertire grafi tra vari formati, visualizzandoli in 2D o 3D, inoltre, possiamo esplorare proprietà dei grafi, quali grado, raggio, diametro e tante altre. Infine, possiamo calcolare sottografi e altre strutture particolari.

NetworkX ha una notevole flessibilità di rappresentazione, infatti permette di associare a nodi ed archi qualsiasi tipo di dato, come ad esempio testo, immagini, numeri, record XML e tanti altri.

In conclusione, si tratta di un framework efficiente, ampiamente scalabile, portatile e molto aggiornato. Verrà utilizzato per effettuare le varie operazioni di analisi e visualizzazione sul grafo d'interesse.

1.3.3 Gravis

Abbiamo utilizzato il package Gravis, per poter rappresentare il grafo risultante dalla nostra analisi in maniera dinamica per una più chiara comprensione di quest'ultimo.

Il nome di tale package sta per graph visualization e il suo scopo è quello di creare grafi (e reti) interattivi sia in 2D che in 3D.

Usa Python per preparare i dati dei grafici e tecnologie web (HTML/CSS/JS) per effettuare il rendering.

Abbiamo scelto questo package, in primo luogo, perché compatibile con NetworkX ed anche perché a differenza di altri package disponibili rende possibile incorporare i grafici risultanti in un Jupyter notebook oltre che come file HTML.

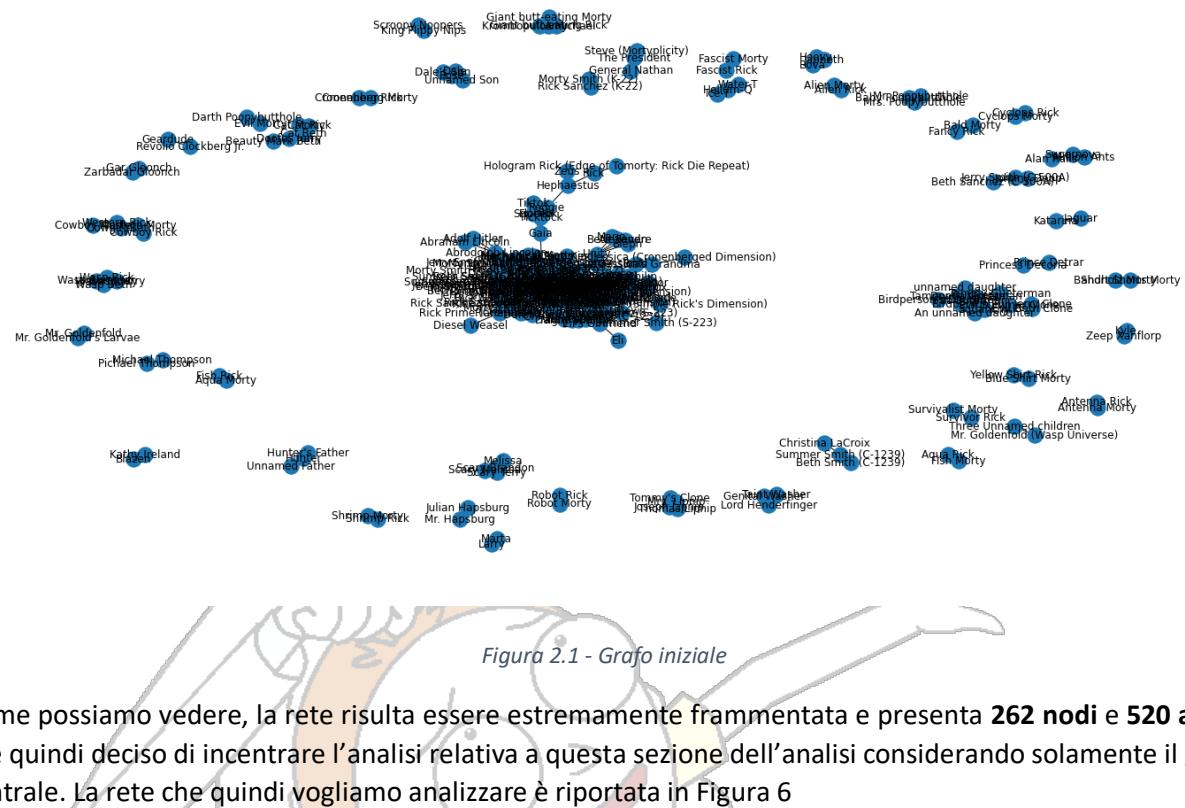


Figura 1.4 - Funzionamento del web scraping

2 Analisi descrittiva

In questo capitolo, verrà fornito un resoconto generale della rete in questione e delle sue caratteristiche sia attraverso la sua visualizzazione che attraverso delle prime “metriche”. Per questa parte del progetto viene preso in esame solamente il l’insieme più grande di nodi e archi connessi tra loro, in quanto il grafo iniziale risulta essere sconnesso, data la presenza di più dimensioni spaziali all’interno della serie tv, con l’impossibilità di calcolare le metriche generali relative alla rete.

La rete iniziale risulta quindi essere quella riportata in Figura 5.



Come possiamo vedere, la rete risulta essere estremamente frammentata e presenta **262 nodi** e **520 archi**. Si è quindi deciso di incentrare l'analisi relativa a questa sezione dell'analisi considerando solamente il grafo centrale. La rete che quindi vogliamo analizzare è riportata in Figura 6

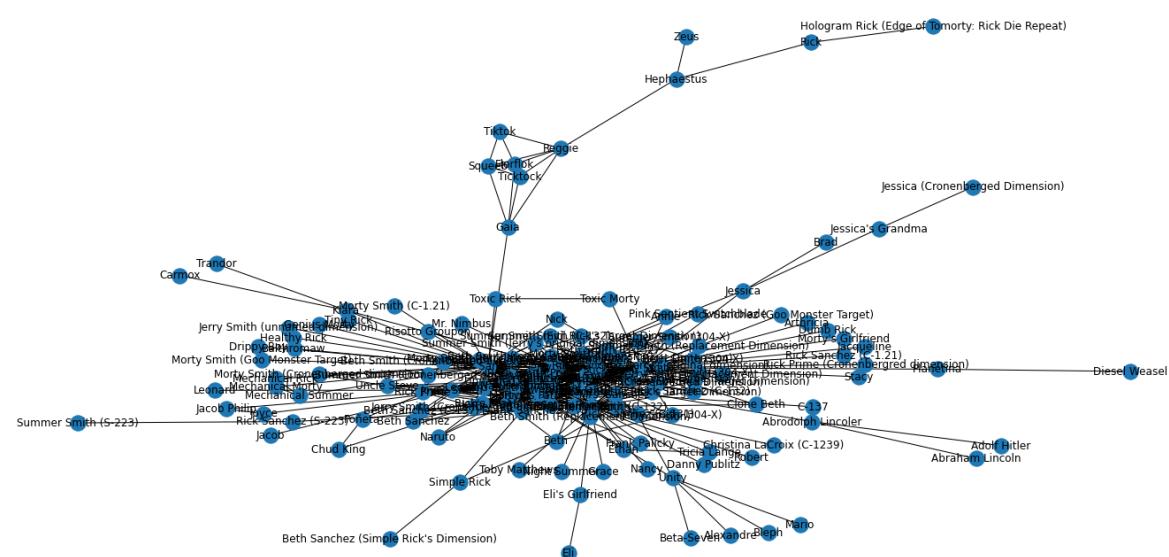


Figura 2.2 - Grafo in analisi



SOCIAL NETWORK ANALYSIS RICK AND MORTY

2.1 Metriche generali

Il primo passo è stato quello di esplorare alcune proprietà tipiche della struttura di una rete, per farsi un'idea in generale della social network in esame. A tal proposito abbiamo considerato:

- **Numero di nodi**
- **Numero di archi**
- **Densità**: è un attributo i cui possibili valori sono compresi tra 0 e 1; ci consente di misurare la probabilità che una qualsiasi coppia di nodi sia adiacente. Si tratta, dunque, di un indice che misura mediamente quanto sono legati, tra loro, i nodi del grafo.
- **Raggio**: corrisponde alla minima eccentricità dei nodi, dove l'eccentricità è definita come la lunghezza massima dello *shortest path* tra un nodo e i restanti nodi della rete.
- **Diametro**: corrisponde all'eccentricità massima tra i nodi
- **Periferia**: rappresenta l'insieme di nodi tali per cui l'eccentricità è pari al diametro
- **Average Clustering coefficient**: il *Clustering Coefficient* di un nodo fornisce una misura di quanti i suoi vicini tendano a formare una *clique*. Facendo una media riferita all'intero grafo, si ottiene una proporzione di quanti i nodi tendono ad essere connessi tra loro.
- **Connessione**: attributo booleano che restituirà "True" se il grafo è connesso, "False" se presenta dei nodi isolati.

I dati risultanti confermano la teoria di "*Six Degrees of Separation*", la quale afferma che ogni persona può essere collegata a qualunque altra persona attraverso una catena di conoscenze e relazioni con non più di 5 intermediari. Per il calcolo delle proprietà sono stati utilizzati opportuni metodi della libreria *NetworkX*.

- **NODI**: 138
- **ARCHI**: 420
- **DENSITÀ**: 0.0444
- **RAGGIO**: 5
- **DIAMETRO**: 9
- **PERIFERIA**: ['Hologram Rick (Edge of Tomorty: Rick Die Repeat)', 'Jessica (Cronenberged Dimension)']
- **CLUSTERING**: 0.377
- **IS CONNECTED**: True

Si nota una densità decisamente bassa, il che può lasciare presagire una disparità nella distribuzione dei gradi tra i vari nodi. Oltre a ciò, il basso rapporto tra il numero di archi del grafo e il numero di archi del medesimo grafo completo ($\frac{n(n-1)}{2} = 9453$) lascia intuire il perché di una densità così esigua.

Proseguendo, nella lista relativa alla *Periferia* sono presenti effettivamente personaggi minori, come ad esempio l'ologramma di Rick e Jessica proveniente dalla dimensione Cronenberged.

Infine, l'*Average Clustering Coefficient* pari a 0.377 indica che, mediamente, circa il 40% delle figure che sono in relazione con un certo personaggio sono anche in relazione tra loro.

2.2 Visualizzazioni e layout

In conclusione dell'analisi descrittiva si è provveduto a rappresentare graficamente la rete, attraverso le diverse possibilità di layout messe a disposizione dalla libreria *NetworkX*.



.social Network Analysis Rick and Morty

Inizialmente, il grafo è stato riorganizzato secondo lo *Spring Layout*, utilizzando poi la libreria *matplotlib* per la sua visualizzazione. Il risultato ottenuto viene riportato in Figura 7.

Come possiamo notare, il grafo risulta essere sparso ma, nell'insieme, una buona parte dei nodi è incentrata nella regione centrale della rete. Sono presenti un certo numero di nodi distanti dalla zona centrale della rete, a ragion del fatto che ogni episodio della serie presenta un certo numero di personaggi differenti o che comunque sono presenti in un numero di scene minori rispetto ai personaggi principali.

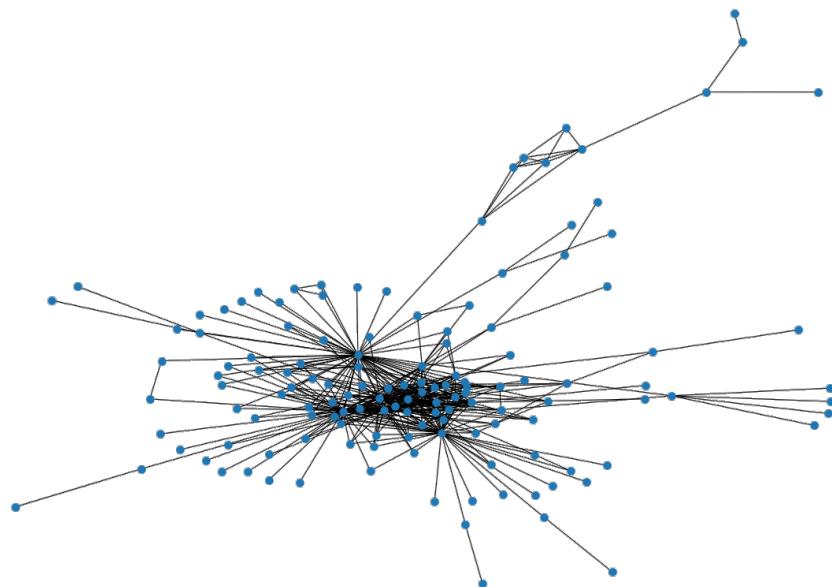
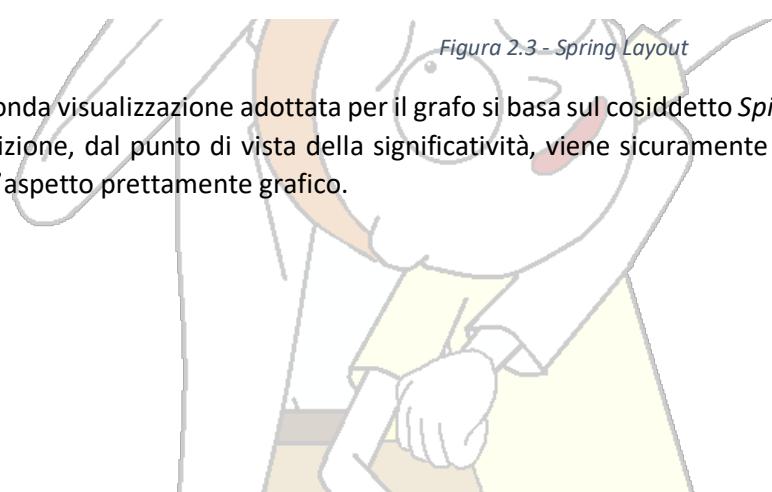


Figura 2.3 - *Spring Layout*

La seconda visualizzazione adottata per il grafo si basa sul cosiddetto *Spiral Layout*, mostrato in Figura 8. Tale disposizione, dal punto di vista della significatività, viene sicuramente a perdere valore ma è interessante sotto l'aspetto prettamente grafico.





SOCIAL NETWORK ANALYSIS RICK AND MORTY

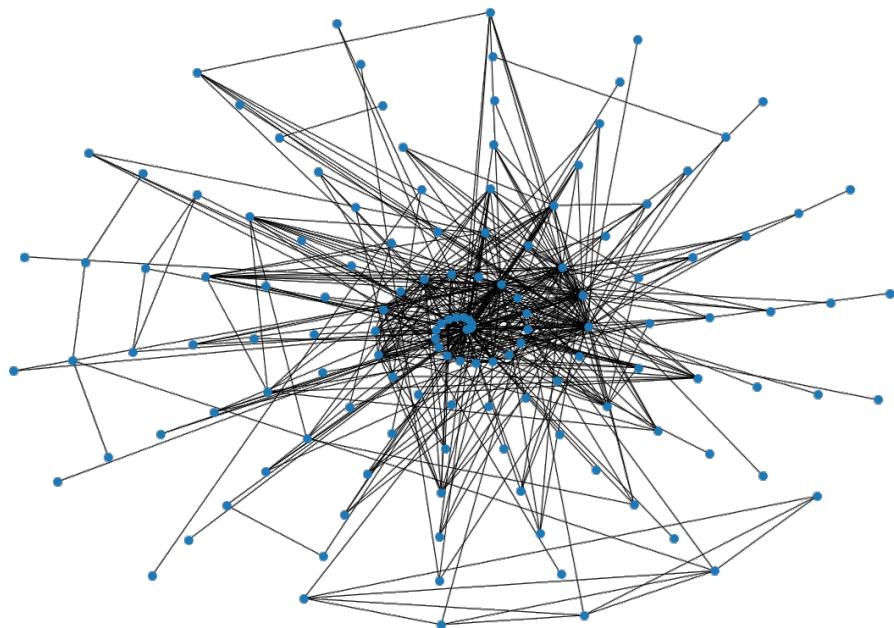


Figura 2.4 - Spiral Layout

Infine, è stato utilizzato il *Kamada Kaway Layout*, il quale dispone i nodi secondo l'omonima funzione-costo che calcola la lunghezza dei percorsi. Tale disposizione, riportata in Figura 9 è del tutto sommato simile allo Spring Layout.

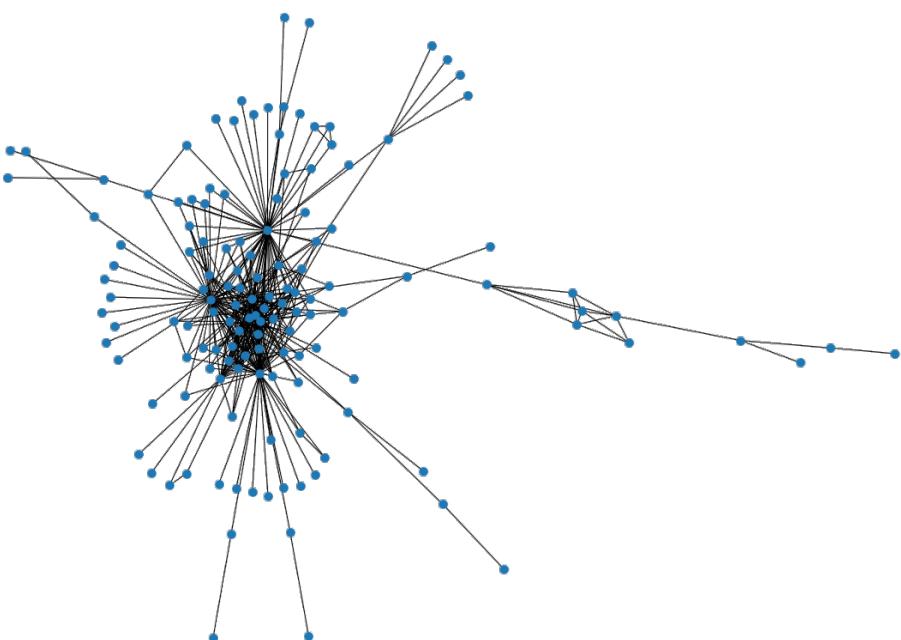


Figura 2.5 - Kamada Kaway Layout



3 Analisi delle centralità

La prima misura che si è deciso di effettuare una misura del potere, o dell'influenza, dei componenti di una rete, basandoci sulle connessioni presenti all'interno del grafo.

Non esistono metriche univoche per stabilire l'importanza di un nodo, per tale motivo si è deciso di realizzare diverse misure di centralità. Di seguito riportiamo i risultati ottenuti dalle analisi e dalle misure più utilizzate per la social network analysis.

3.1 Degree Centrality

La prima misura di centralità che si è deciso di calcolare è la **Degree Centrality** con la quale si calcolano il numero di legami che un certo nodo ha con gli altri della rete. Questa misura però non garantisce che il nodo con la degree centrality più alta sia il più importante, perché si potrebbe ricadere nel caso in cui si hanno tanti legami ma di poca importanza. Per realizzare tale misura abbiamo utilizzato una funzione, la `degree_centrality()`, fornita dalla libreria NetworkX. Di seguito riportiamo i risultati ottenuti.

1	Rick Sanchez:	0.22988505747126436
2	Morty Smith:	0.2222222222222222
3	Summer Smith:	0.20306513409961685
4	Jerry Smith:	0.15708812260536398
5	Beth Smith:	0.1532567049808429
6	Morty Jr.:	0.08045977011494253
7	Joyce Smith:	0.08045977011494253
8	Leonard Smith:	0.07279693486590037
9	Hemorrhage:	0.07279693486590037
10	Jerry Smith (C-131):	0.07279693486590037

Figura 3.1 – Risultati degree centrality

Questa è la lista dei dieci nodi con il valore di degree centrality più alto.

Tale risultato conferma le nostre aspettative, infatti, notiamo come i protagonisti della serie sono quelli con il valore più alto. Gli altri personaggi sono i componenti della famiglia di Rick e Morty, i quali partecipano solo in maniera ridotta alle loro avventure.



SOCIAL NETWORK ANALYSIS RICK AND MORTY

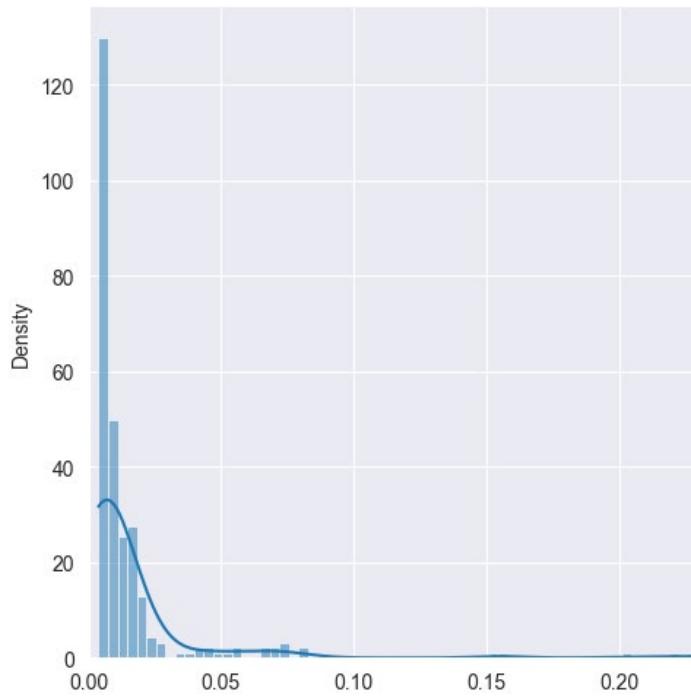


Figura 3.2 - Analisi della distribuzione della degree centrality

Arrivati a questo punto si è deciso di procedere con l'analisi della distribuzione di tale valore. Per procedere, si è preso in considerazione il numero di nodi del grafo, infatti, siccome nel grafo sono presenti molti nodi, riportarli in un bar plot con tutti i valori delle degree centrality avrebbe restituito un risultato illeggibile. Quindi si è deciso di procedere con l'analisi delle distribuzioni utilizzando il displot.

Come si può osservare abbiamo un elevato numero di nodi con un valore della degree centrality molto bassa, questo perché alcuni personaggi sono visibili solo in una o due puntate, mentre, come osservato precedentemente, solo i protagonisti e i loro familiari hanno un valore superiore a 0.10.

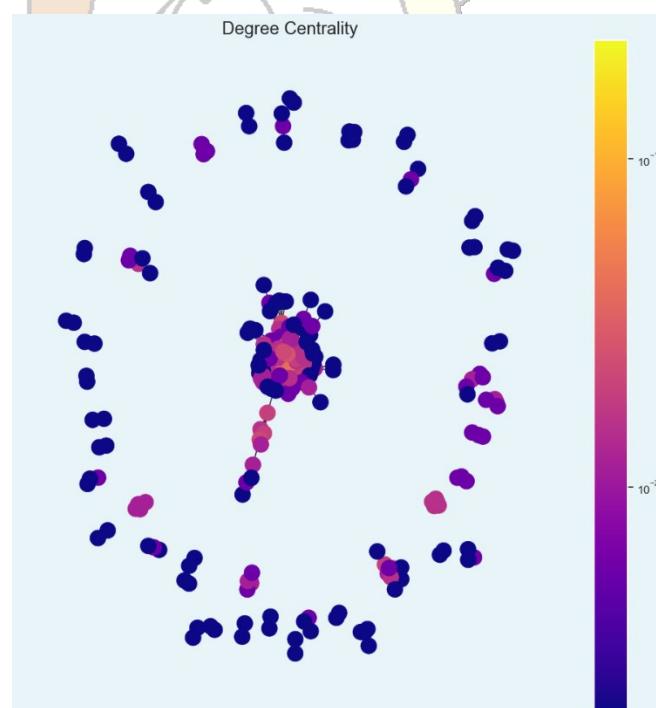


Figura 3.3 - Heatmap della degree centrality



Infine, si riporta la heatmap delle degree centrality in cui i nodi centrali, sono quelli interessati da un elevato numero di connessioni, quindi, hanno un valore molto alto di questa misura.

Viceversa, i nodi con i valori più bassi sono quelli periferici, situati più all'esterno, perché sono quelli con il minor numero di connessioni.

Di seguito procediamo con una breve descrizione dei personaggi principali, per contestualizzare meglio i risultati ottenuti. Quindi:

ALLERTA SPOILER !!!

3.2 Descrizione dei personaggi

3.2.1 Rick Sanchez



Rick Sanchez è uno dei protagonisti della serie animata Rick e Morty.

Rick è un grande scienziato, un vero genio che ha inventato un sacco di cose, tra cui il viaggio interdimensionale. Infatti, grazie ad uno strumento chiamato `sparaporte` è in grado di viaggiare tra diverse realtà. Queste realtà sono altre dimensioni, che possono differire dalla nostra per piccoli o per grandi dettagli. Ad esempio, abbiamo realtà dove le persone sono telefoni e chiamano con gli esseri umani, oppure, realtà identiche alla nostra, quindi stesse persone e stesse situazioni, solo che una parola viene pronunciata diversamente.

Rick è il nonno di Beth Sanchez, la quale ha sposato Jerry Smith dalla loro unione sono nati Summer Smith e suo fratello minore Morty Smith.

Nonostante la sua grande intelligenza, Rick soffre di una grande depressione, che cerca ripetutamente di affogare in alcool e droghe. Il dolore all'origine della sua depressione è dovuto ad un evento molto tragico, nel quale morirono sua figlia e sua moglie, quelle originali.

Durante questo evento, avvenuto nel passato, un Rick proveniente da un'altra realtà uccise la moglie e la figlia del nostro Rick, dopo che questi aveva rifiutato l'invito a realizzare una nuova tecnologia.

Nonostante i suoi sforzi e la sua infinita conoscenza Rick non riesce a trovare la sua versione che gli uccise moglie e figlia, e questo lo porta in uno stato di profonda depressione.

Per raggiungere i suoi scopi Rick è disposto a tutto, si preoccupa solo degli amici e dei parenti stretti. Inoltre, non si lascia mai sfuggire l'occasione di partire per un'avventura con il nipote.



SOCIAL NETWORK ANALYSIS RICK AND MORTY

3.2.2 Morty Smith



Morty Smith è il secondo protagonista della serie animata Rick e Morty.

Morty è il nipote di Rick, un ragazzino ingenuo, impressionabile e facilmente manipolabile, infatti, suo nonno lo sfrutta sempre come assistente nelle sue avventure. Morty viene considerato da tutti come un ragazzino stupido, anche se in più occasioni ha dimostrato di essere molto intelligente.

Ha problemi a relazionarsi con le ragazze, infatti, quando prova a parlarci è molto impacciato e insicuro. Questa condizione lo porta a praticare molto autoerotismo, e in episodio, diventa papà a seguito dell'utilizzo di un robot sessuale.

Morty vuole molto bene a suo nonno, anche se spesso è la causa dei suoi problemi. Vuole molto bene anche alla sua famiglia, anche se spesso litiga con la sorella Summer per ricevere le attenzioni del nonno.

Infine, anche se normalmente Morty è un ragazzo gentile e ingenuo, a volte diventa molto furbo e aggressivo per cercare di ottenere quello che vuole.

3.2.3 Summer Smith



Summer Smith è un personaggio della serie animata Rick e Morty.

Summer rappresenta la classica teenager americana, e tiene molto in considerazione l'opinione dei suoi coetanei. Passa molto del suo tempo al cellulare e sui social media, inoltre, si sente imbarazzata e frustrata



.social Network Analysis Rick and Morty

a causa della sua famiglia. Sebbene non sia intelligente quanto Rick, è molto sveglia, specialmente se paragonata al resto della sua famiglia.

Nonostante i frequenti litigi Summer vuole molto a suo fratello Morty, infatti, ci sono delle avventure dove i due collaborano per risolvere situazioni difficili, e per aiutare il nonno. Summer non ha un grande rapporto con il nonno all'inizio della serie, ma con il passare degli episodi legano sempre di più e alcune volte sostituirà Morty in alcune avventure. Ha un discreto rapporto con i genitori, in passato le avevano confidato che lei era il risultato di una gravidanza indesiderata, e questo l'aveva portata ad allontanarsi, anche se poi si è riavvicinata a entrambi. Summer è più legata alla madre che al padre, di quest'ultimo non sopporta la condizione di disoccupato.

3.3 Closeness Centrality

La Closeness Centrality è una misura che calcola la distanza di un nodo dagli altri nodi della rete. Un nodo con un valore alto di questa misura lo indica come un ottimo candidato per ricevere ed inviare informazioni agli altri componenti della rete. In quest'analisi, così come nella precedente, si è deciso di riportare i dieci personaggi con il valore più significativo. Per procedere si è utilizzata la funzione **closeness_centrality()**, messa a disposizione dalla libreria networkX. Di seguito riportiamo i risultati ottenuti.

1	Rick Sanchez:	0.30600798891334474
2	Morty Smith:	0.29232470485624396
3	Summer Smith:	0.28423666954401583
4	Beth Smith:	0.2683279007262538
5	Jerry Smith:	0.2605502804153479
6	Morty Jr.:	0.2505640327339234
7	Hemorrhage:	0.24797199101598624
8	Space Beth:	0.24711985358981448
9	Jerry Smith (C-131):	0.2462735527213562
10	Naruto Smith:	0.24459822243073476

Figura 3.4 - Risultati closeness centrality

Osserviamo come la maggior parte dei personaggi presenti in questa lista sono anche presenti in quella precedente, e questo conferma i nostri risultati. I nodi con il maggior numero di legami sono anche capaci di trasmettere informazioni velocemente all'interno della rete.

Successivamente si è realizzata l'analisi della distribuzione di tale misura. Di seguito riportiamo i risultati ottenuti.



SOCIAL NETWORK ANALYSIS RICK AND MORTY

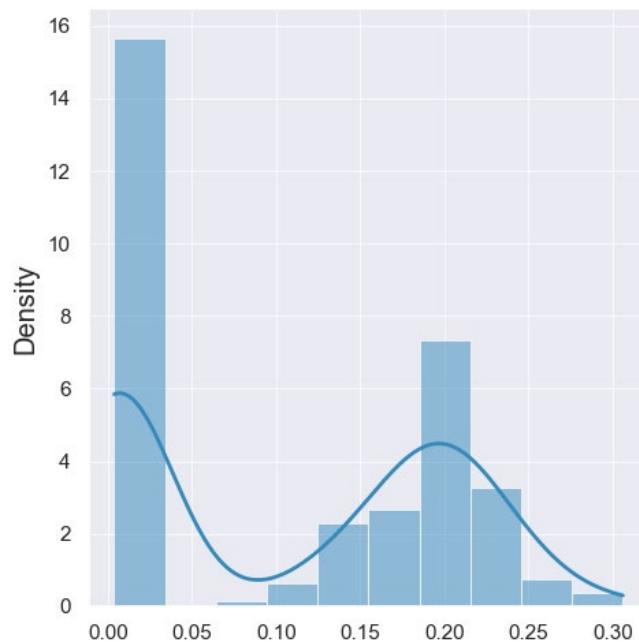


Figura 3.5 - Distribuzione closeness centrality

Da questo grafico è possibile osservare come molti nodi abbiano un valore della misura bassissimo. Questi sono tutti i nodi, che nella precedente analisi, avevano valori molto bassi della degree centrality perché erano i nodi periferici, sconnessi dagli altri nodi che formano il grafo centrale e con pochissimi nodi periferici a loro attaccati.

Vediamo, come un discreto numero di nodi abbia un valore della misura abbastanza alto, questo significa che questi nodi sono poco distanti da tutti gli altri.

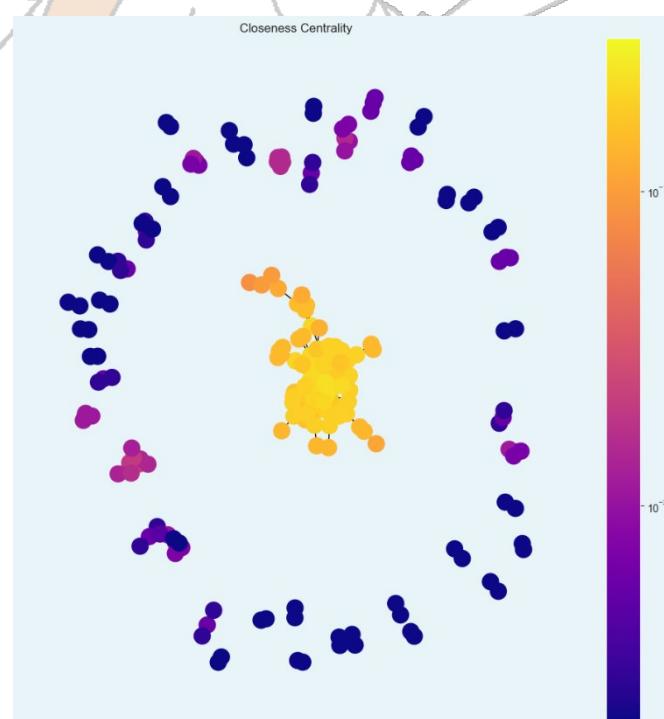


Figura 3.6 - Heatmap closeness centrality



.social Network Analysis Rick and Morty

Infine, realizziamo la heatmap della closeness centrality. Da questo grafico confermiamo le osservazioni precedenti, cioè, i nodi periferici sono caratterizzati da bassi valori, mentre i nodi presenti al centro della rete hanno un valore abbastanza alto di questa misura.

3.4 Eigenvector Centrality

La terza metrica che si è deciso di calcolare è la **Eigenvector Centrality**. Questa centralità ci permette di misurare quanto un nodo è autorevole e influente all'interno della rete. Ovvero, ci permette di dare un valore che quantifica la sua importanza rispetto ai nodi che gli sono vicini. L'analisi si è realizzata mediante l'uso della funzione `eigenvector_centrality()` messa a disposizione dalla libreria networkX. Riportiamo di seguito la lista dei dieci nodi con i valori più alti di questa misura.

```
1 Summer Smith: 0.3003362881137109
2 Morty Smith: 0.29578066660016405
3 Beth Smith: 0.28020034517224623
4 Rick Sanchez: 0.27187098468209375
5 Jerry Smith: 0.265334780729058
6 Morty Jr.: 0.21167111107564016
7 Hemorrhage: 0.1960894506499813
8 Rick's Father: 0.18210163696500467
9 Jerry Smith (C-131): 0.1819231445403477
10 Joyce Smith: 0.18166348200120544
```

Figura 3.7 - Risultati eigenvector centrality

In questo caso otteniamo dei risultati leggermente diversi rispetto alle analisi precedenti. Questo è dovuto al fatto che il nostro grafo non è pesato, per due motivi. Il primo fa riferimento all'origine stessa della rete, infatti, tutte le informazioni utilizzate per realizzarla sono state prese da internet, e questa non era presente, non avevamo dei pesi da assegnare agli archi che collegavano due nodi. Il secondo fa riferimento ai personaggi stessi della serie animata. Come abbiamo spiegato prima, nella prima analisi, Rick è uno scienziato che viaggia tra le realtà, quindi, non si capisce bene quanto sia effettivamente legato ai suoi "attuali" familiari, anche per Morty si fa fatica, in quanto sappiamo che i Morty sono solo degli scudi umani per i Rick. Inoltre, anche Morty, a causa di Rick, è costretto a cambiare realtà dopo pochi episodi.

Nel nostro caso la funzione calcola la misura assegnando a tutti gli archi un peso pari a 1.

A meno dell'ordine, i nodi con il valore maggiore sono gli stessi delle analisi precedenti, inoltre, la differenza di valore tra questi nodi è molto bassa.



SOCIAL NETWORK ANALYSIS RICK AND MORTY

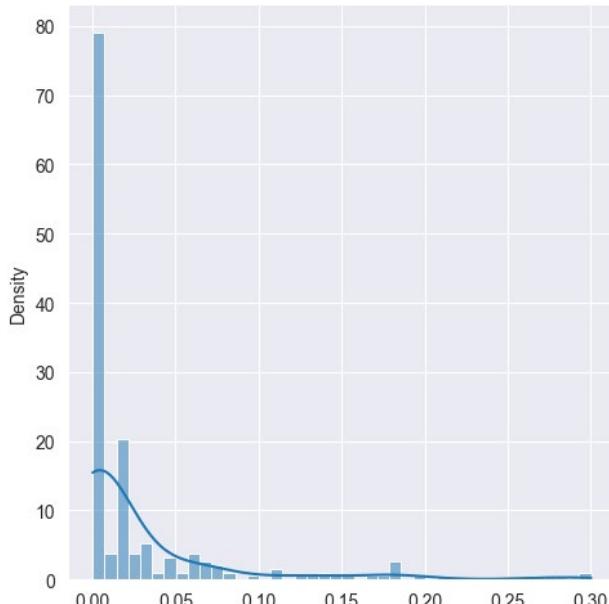


Figura 3.8 - Distribuzione della eigenvector centrality

Si è proseguita l'analisi con l'osservazione della distribuzione dei valori di tale misura. Anche in questo caso ci sono tanti nodi con un valore molto basso di questa centralità, le ipotesi che facciamo sono analoghe alle misure precedenti. Procediamo con la realizzazione della heatmap.

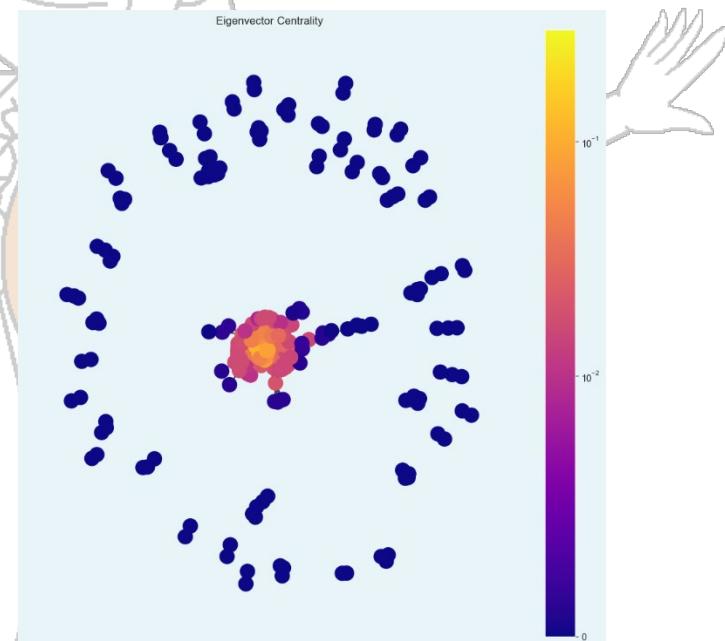


Figura 3.9 - Heatmap eigenvector centrality

La heatmap conferma le ipotesi precedenti, i nodi più esterni sono quelli con i valori più bassi di eigenvector centrality, mentre i nodi più interni sono quelli con i valori più alti.

3.5 Betweenness Centrality

A questo punto, si procede con l'analisi dell'ultima metrica, la **Betweenness Centrality**, che misura la strategicità di un nodo della rete tra due aree importanti della stessa. Un nodo con un valore di betweenness centrality alto è un nodo che funge da bridge tra due o più insiemi di nodi della rete. Per realizzare tale misura



.social Network Analysis Rick and Morty

si è fatto uso della funzione **betweenness_centrality()**, messa a disposizione dalla libreria networkX. Di seguito riportiamo i nomi dei dieci nodi con il valore più alto di questa misura.

1	Rick Sanchez:	0.12808106514057074
2	Morty Smith:	0.0741216988142787
3	Summer Smith:	0.06342072620731264
4	Gaia:	0.03395225464190981
5	Reggie:	0.016956479025444515
6	Jerry Smith:	0.01685046470239694
7	Unity:	0.01599441427027633
8	Beth Smith:	0.014966048835369831
9	Joyce Smith:	0.012846063221187325
10	Jessica:	0.01194583619870976

Figura 3.10 - Risultati betweenness centrality

Osservando i risultati si nota come i nodi con i valori più alti delle precedenti misure sono anche quelli con il valore della betweenness centrality più alto. Però, si nota come i valori di questa misura siano molto più bassi rispetto ai valori delle altre misure. Questo indica che nella rete non ci sono nodi che fungono da punto chiave di comunicazione tra diversi sottoinsiemi di nodi.

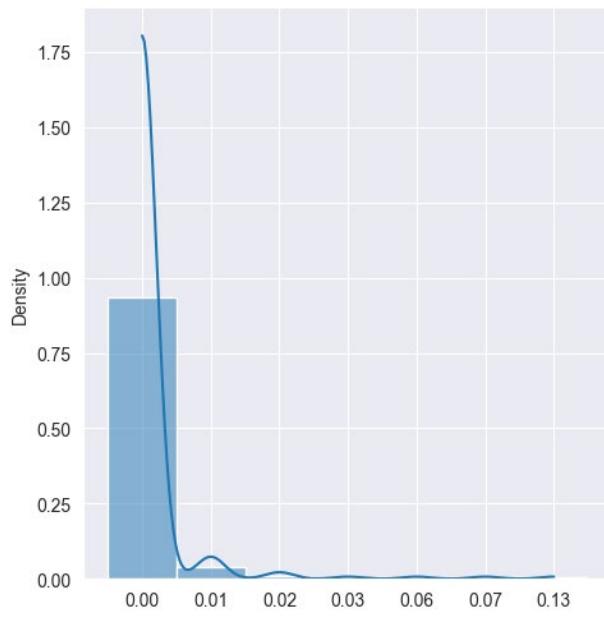


Figura 3.11 - Distribuzione betweenness centrality

Si è proceduto con l'analisi della distribuzione. Per realizzarla è stato necessario fare delle approssimazioni dei valori, in quanto molti nodi avevano dei valori di questa centralità talmente bassi da rendere il risultato illeggibile. Osserviamo come la maggior parte dei nodi abbia un valore tendente allo zero, con un solo nodo che ha un valore superiore a 0.1.



SOCIAL NETWORK ANALYSIS RICK AND MORTY

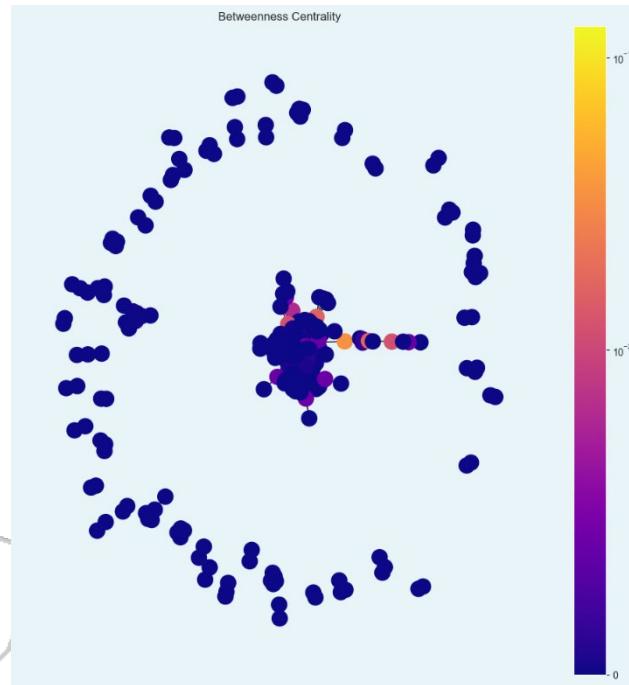
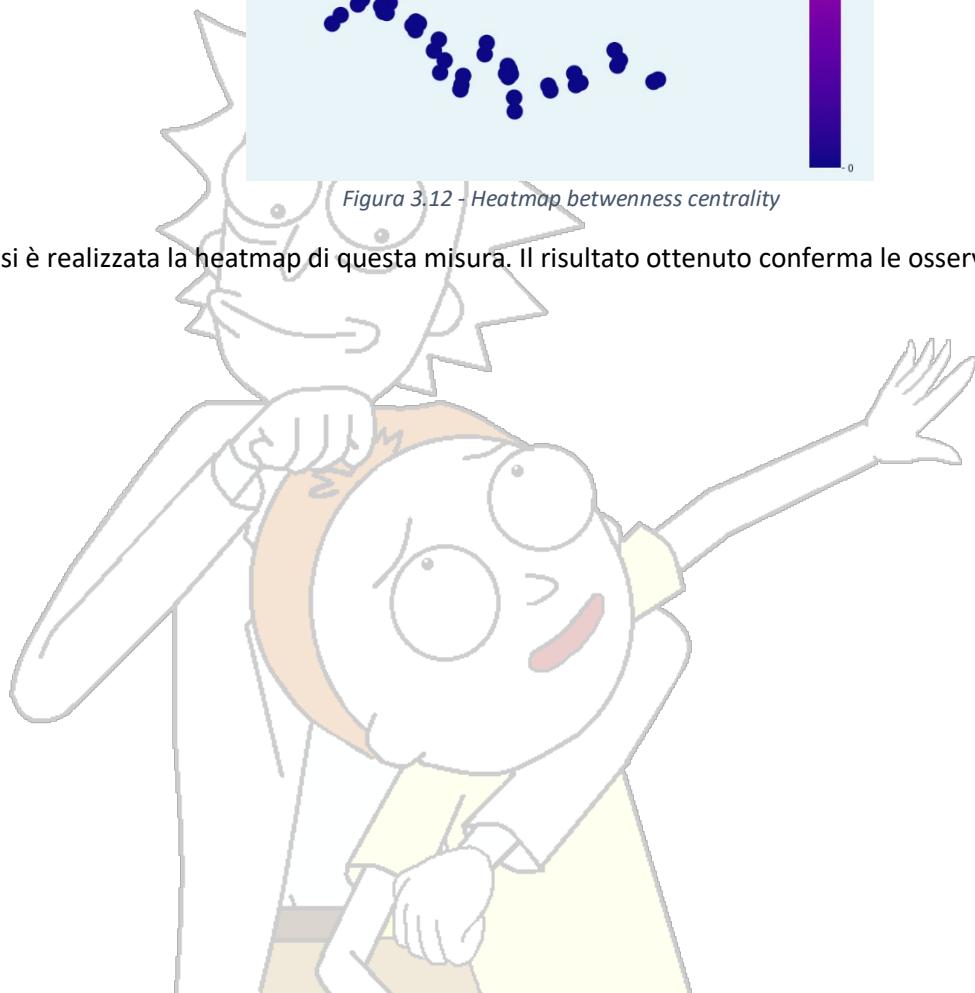


Figura 3.12 - Heatmap betweenness centrality

Infine, si è realizzata la heatmap di questa misura. Il risultato ottenuto conferma le osservazioni precedenti.





4 Analisi delle strutture

In questa parte del progetto ci siamo concentrati sui sottografi presenti nella rete che potrebbero dare luogo a strutture quali le clique e le triadi, che sono note per la loro coesione e robustezza.

Abbiamo poi deciso di applicare il concetto di Ego Network ai due personaggi protagonisti della serie (come si può intuire dal nome della stessa, abbiamo scelto Rick e Morty) per farci un'idea del "vicinato" di tali personaggi e del loro campo di azione.

4.1 Triadi

Un insieme di tre nodi è quello che viene definita una triade e se tutti i nodi, di tale insieme, sono tra loro connessi, la triade è definita chiusa.

Si presta particolare attenzione alle triadi perché esse sono un sinonimo di stabilità per quel che riguarda la diffusione delle informazioni, se infatti si dovesse eliminare un arco da una triade chiusa tutti i suoi nodi sarebbero ancora in grado di comunicare, e quindi l'informazione riuscirebbe comunque a raggiungere tutti i nodi.

Le triadi chiuse trovate sono state 801.

Abbiamo successivamente filtrato le triadi per prendere in considerazione solamente quelle che contenevano sia Rick che Morty e abbiamo così ridotto il numero a 32.

Le triadi chiuse contenenti sia Rick che Morty sono: 32

e sono:

```
('Rick Sanchez', 'Diane Sanchez', 'Morty Smith')
('Rick Sanchez', 'Beth Smith (Cronenberg dimension)', 'Morty Smith')
('Rick Sanchez', 'Space Beth', 'Morty Smith')
('Rick Sanchez', 'Jerry Smith (Cronenberg dimension)', 'Morty Smith')
('Rick Sanchez', 'Summer Smith (Cronenberg dimension)', 'Morty Smith')
('Rick Sanchez', 'Jerry Smith (Replacement Dimension)', 'Morty Smith')
('Rick Sanchez', 'Beth Smith (Replacement Dimension)', 'Morty Smith')
('Rick Sanchez', 'Summer Smith (Replacement Dimension)', 'Morty Smith')
('Rick Sanchez', 'Morty Jr.', 'Morty Smith')
...
('Rick Sanchez', 'Morty Smith', 'Beth Smith (C-132)')
('Rick Sanchez', 'Morty Smith', 'Sticky')
```

Figura 4.1 – Triadi chiuse contenenti sia Rick che Morty

4.2 Clique

Se si estende il concetto di triade chiusa a più di 3 nodi si parla allora di clique, essa è dunque un generico insieme di nodi totalmente connesso, e si può quindi assimilare ad un insieme di triadi tra loro sovrapposte. Risultano di notevole importanza in quanto rappresentano la massima espressione di coesione tra i nodi di una rete.

Il numero di clique trovate è 6159 di cui 252 sono massimali.

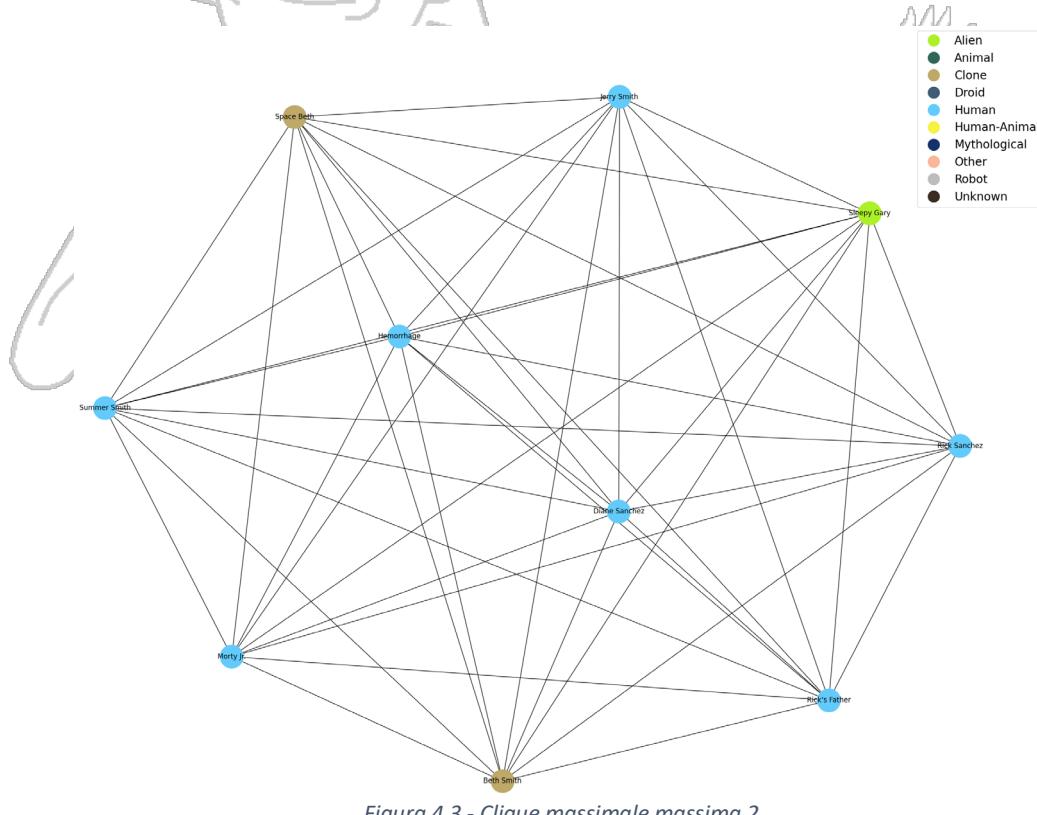
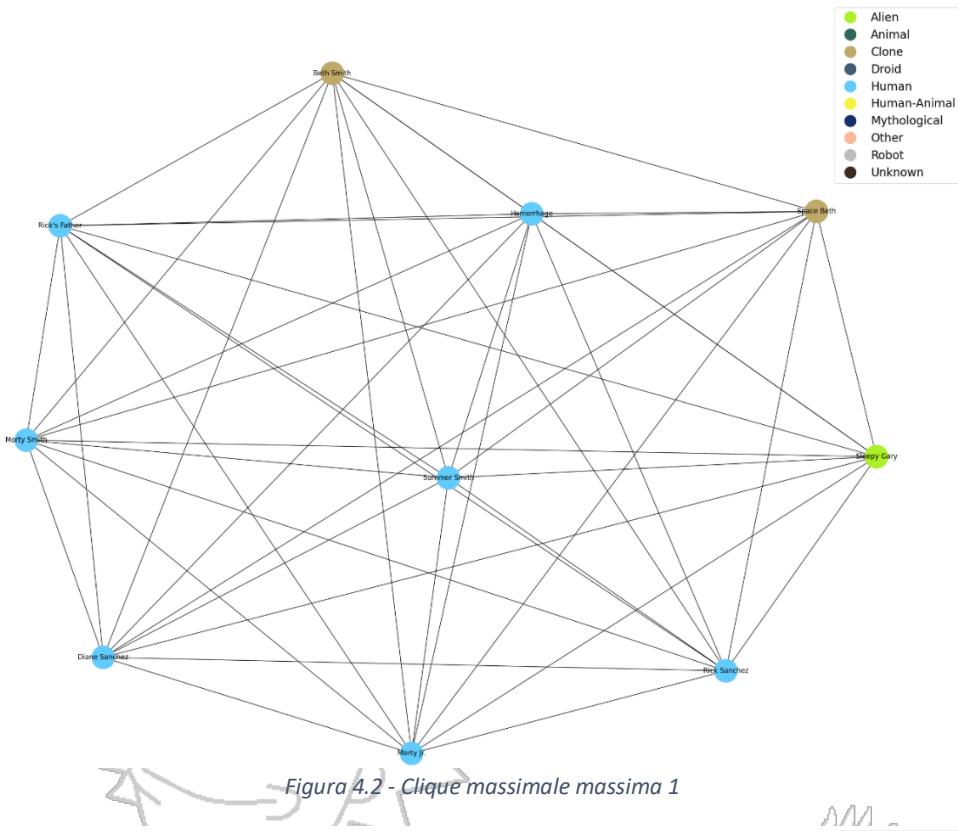
Le clique massimali e massime (ovvero con il maggior numero di elementi) in questa rete sono composte da 10 elementi e sono 2.

Clique 1. ['Rick Sanchez', 'Morty Smith', 'Summer Smith', 'Beth Smith', 'Morty Jr.', 'Rick's Father', 'Space Beth', 'Hemorrhage', 'Diane Sanchez', 'Sleepy Gary'];

Clique 2. ['Rick Sanchez', 'Jerry Smith', 'Summer Smith', 'Beth Smith', 'Morty Jr.', 'Space Beth', 'Rick's Father', 'Hemorrhage', 'Diane Sanchez', 'Sleepy Gary']



SOCIAL NETWORK ANALYSIS RICK AND MORTY



È possibile notare che le due cliques in questione è costituita essenzialmente dai membri della famiglia di Rick Sanchez. Saltano all'occhio due personaggi che non sono della famiglia in senso stretto e sono *Hemorrhage* e *Sleepy Gary*, dei quali però il primo instaura legami di parentela con tutta la famiglia a seguito del matrimonio con *Summer Smith* nel secondo episodio della terza stagione e il secondo invece instaura legami con tutta la famiglia in quanto alieno parassita che vuole convincere tutti di essere parte delle loro vite.



SOCIAL NETWORK ANALYSIS RICK AND MORTY

4.3 K-core

Rilassando il concetto di clique, consentendo ad un nodo di unirsi al gruppo se è connesso a k nodi, indipendentemente da quanti sono i nodi a cui esso non è connesso, si arriva alla definizione di k-core.

Al variare della dimensione di k, ovviamente anche le strutture che risultano sono differenti.

Sostanzialmente un k-core è un gruppo massimale di nodi ciascuno dei quali connessi ad almeno k altri nodi del gruppo.

Il sottografo in Figura 4.4 è stato calcolato per mezzo del metodo `k_core()`, al quale non è stato passato il parametro k al fine di ottenere il “main-core” del grafo di partenza.

Il numero dei nodi appartenenti a questo k-core è 17 e il valore di k è pari a 11. Se si presta attenzione ai nodi di tale grafo si può notare che tali nodi sono quelli delle clique massime con l’aggiunta di altri 6 nodi.

I nodi in più rispetto alle clique sono: [*'Jerry Smith (C-131)', 'Unnamed Uncle', 'Leonard Smith', 'Joyce Smith', 'Naruto Smith', 'Gwendolyn'*].

Si ha un numero maggiore di nodi in quanto nei k-core, non c’è più il vincolo di connessione diretta a ciascun altro nodo, dunque, il percorso tra due nodi qualsiasi può comprendere più di un arco.

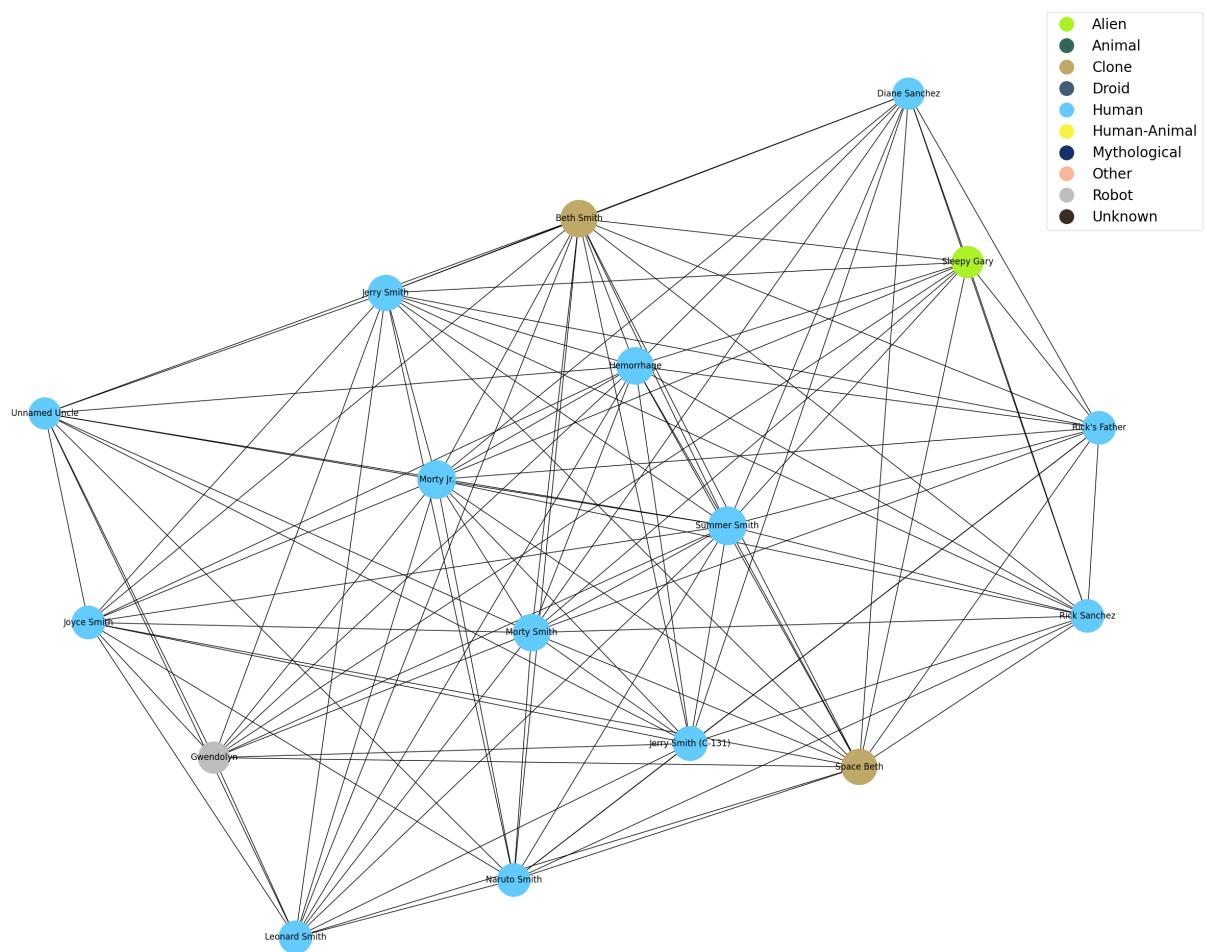


Figura 4.4 - K-core

I nodi in più rispetto alle clique, rappresentano dei personaggi che possono comunque essere considerati della famiglia di Rick Sanchez ma in un senso più esteso e soprattutto sono personaggi che nella serie compaiono poche volte, non rendendoli fondamentali e legati con tutti gli altri delle clique.



SOCIAL NETWORK ANALYSIS RICK AND MORTY

4.4 Ego Network

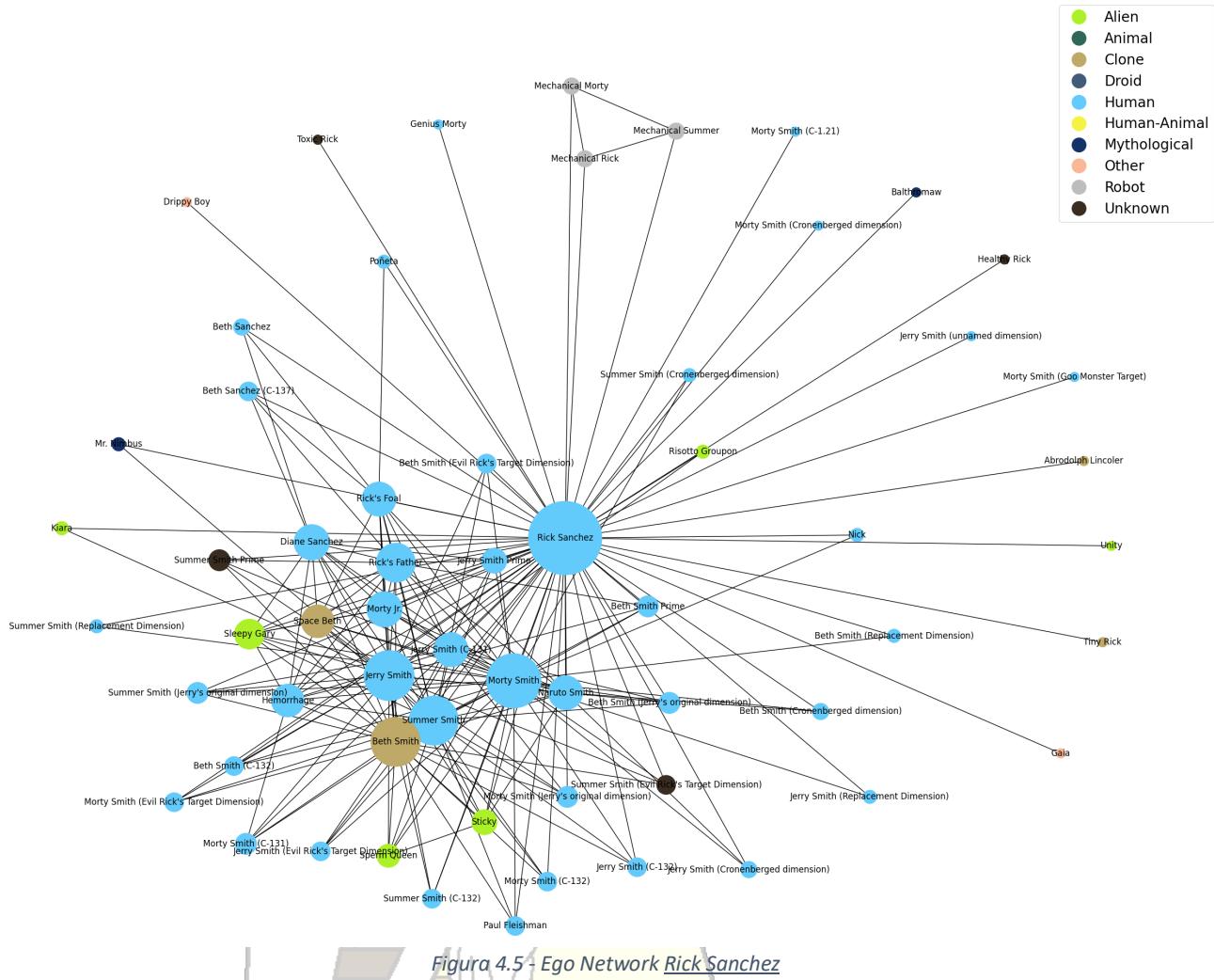
Una Ego Network altro non è che una sottorete, ma centrata su uno specifico nodo, ed è di notevole importanza per avere maggiori informazioni sul nodo centrale, in quanto vedendo i vicini di tale nodo si possono ricavare informazioni che altrimenti non si avrebbero.

Formalmente la struttura di una Ego Network si basa sul concetto di omofilia, ovvero la tendenza delle persone di circondarsi di quelle a loro più simili.

Per fare ciò ci siamo serviti del metodo `ego_graph()`, messo a disposizione da NetworkX.

4.4.1 Rick

Si è proceduto costruendo l'ego network relativa al personaggio di Rick, e successivamente si è analizzata la rete risultante.



Osservando l'ego-network riportata nella Figura 4.5, emerge un aspetto particolare, ovvero la presenza di due regioni con caratteristiche assai differenti.

Nell'area in basso, si può vedere che i vicini del nodo "ego" presentano molte connessioni reciproche, formando una struttura alquanto fitta nella quale sicuramente le informazioni transitano in maniera più efficiente. Questa regione è caratterizzata dai membri della famiglia, intesa in senso ampio.

Invece nella parte in alto, è visibile come i vicini di Rick Sanchez (che non fanno parte della famiglia), non siano propensi ad essere connessi tra loro, probabilmente perché provenienti da dimensioni differenti, domina dunque l'ego e le sembianze sono quelle di una rete a stella, e Rick funge da "intermediatore" in quanto può viaggiare attraverso le dimensioni tramite la sua spara-porte.

4.4.2 Morty

Si è proceduto costruendo l'ego network relativa al personaggio di Rick, e successivamente si è analizzata la rete risultante.

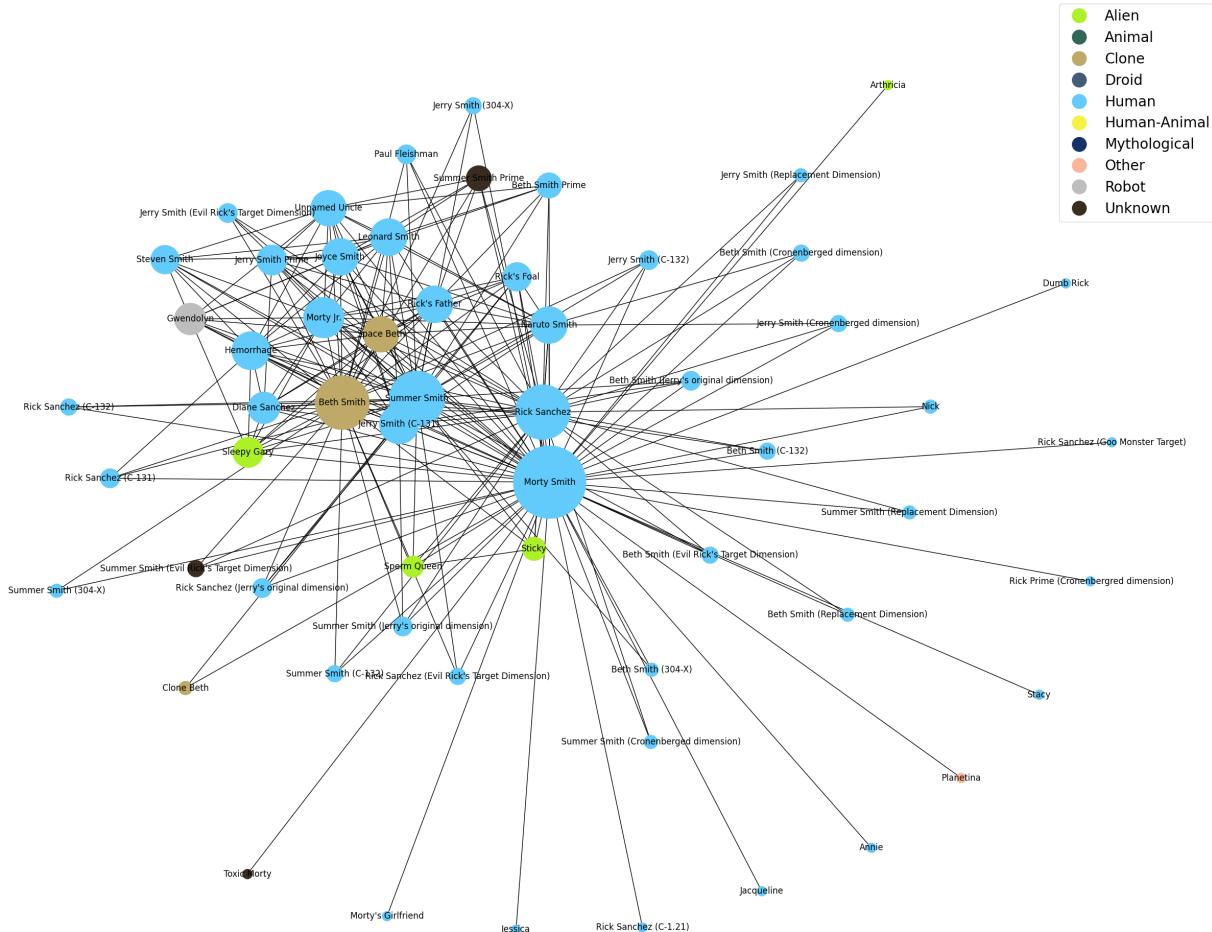


Figura 4.6 - Ego Network Morty Smith

Osservando l'ego-network riportata nella Figura 4.6, emerge un aspetto particolare, ovvero la presenza di due regioni con caratteristiche assai differenti.

Nell'area in basso, si può vedere che i vicini del nodo “ego” presentano molte connessioni reciproche, formando una struttura alquanto fitta nella quale sicuramente le informazioni transitano in maniera più efficiente. Questa regione è caratterizzata dai membri della famiglia, intesa in senso ampio.

Si può notare come la parte densa sia alquanto simile a quella della ego network di Rick Sanchez.

Invece nella parte in alto, è visibile come i vicini di Morty Smith (che non fanno parte della famiglia), non siano propensi ad essere connessi tra loro. Questa parte è costituita prettamente dalle ex-fidanzate o interessi amorosi del nodo centrale.



5 Group Centrality

L'ultimo tipo di analisi che si è deciso di realizzare è quella relativa al calcolo dei vari tipi di centralità per varie famiglie che compaiono all'interno della serie, per capire quali fossero i gruppi più significativi. Le metriche di centralità di gruppo assumono lo stesso significato di quelle già viste nelle sezioni precedenti, ma gli algoritmi utilizzati non lavorano con i singoli nodi, ma con insiemi di nodi. Queste funzioni appartengono alla libreria networkX, e sono:

- **Group_degree_centrality**: viene calcolata come il rapporto tra la somma del numero dei nodi non appartenenti al gruppo, ma connessi ad uno dei membri, e il numero totale di nodi che non appartengono al gruppo.
- **Group_closeness_centrality**: viene calcolata come il rapporto fra il numero di nodi non appartenenti al gruppo e la somma delle distanze tra il gruppo e ogni nodo non appartenente al gruppo.
- **Group_betweenness_centrality**: viene calcolata come il rapporto tra il numero di shortest path tra le coppie di nodi del grafo che passano per almeno uno dei nodi del gruppo e il totale di shortest path di tutte le possibili coppie del grafo.

Per comprendere meglio i risultati ottenuti, mostriamo di seguito l'albero genealogico che definisce le varie relazioni tra i personaggi. Per semplificare la lettura non vengono riportati tutti i personaggi, ma solo i più importanti.

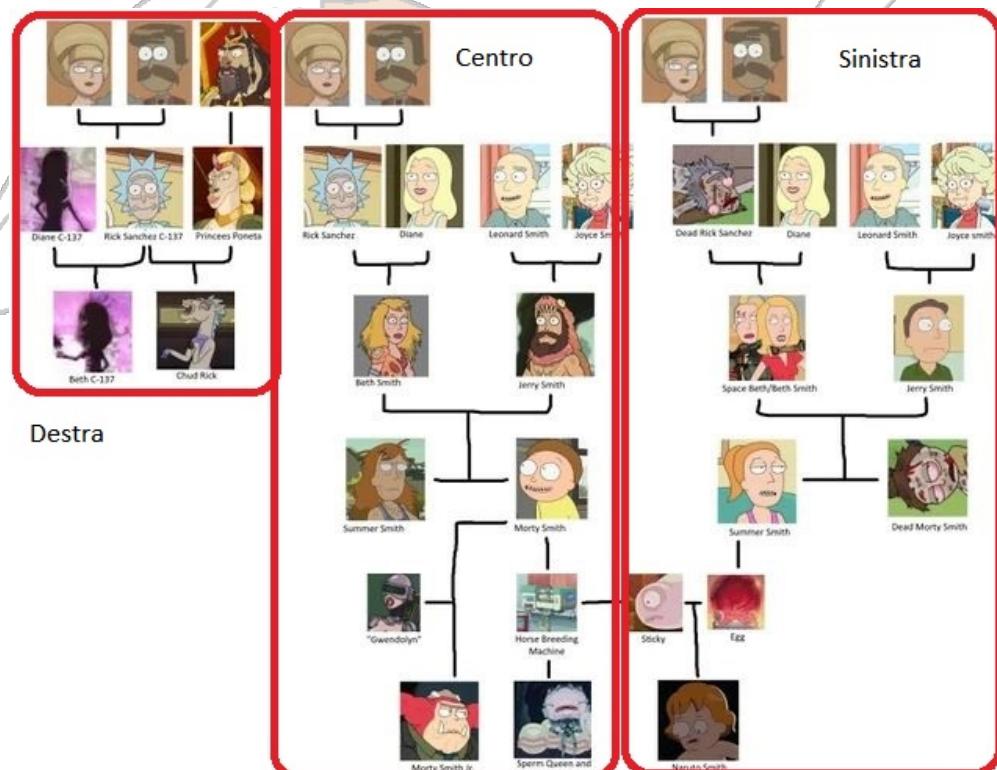


Figura 5.1 - Albero genealogico dei personaggi

Come si può osservare dalla figura sono stati considerati tre gruppi. Le tre partizioni sono state chiamate per semplicità, partizione destra, sinistra e partizione di centro. Sulla base di questi gruppi sono state calcolate



.social Network analysis RICK AND MORTY

le metriche di centralità, le quali poi sono state riportate in tre diversi diagrammi a barre in modo da poter visualizzare meglio i risultati ottenuti. Di seguito riportiamo i risultati ottenuti.

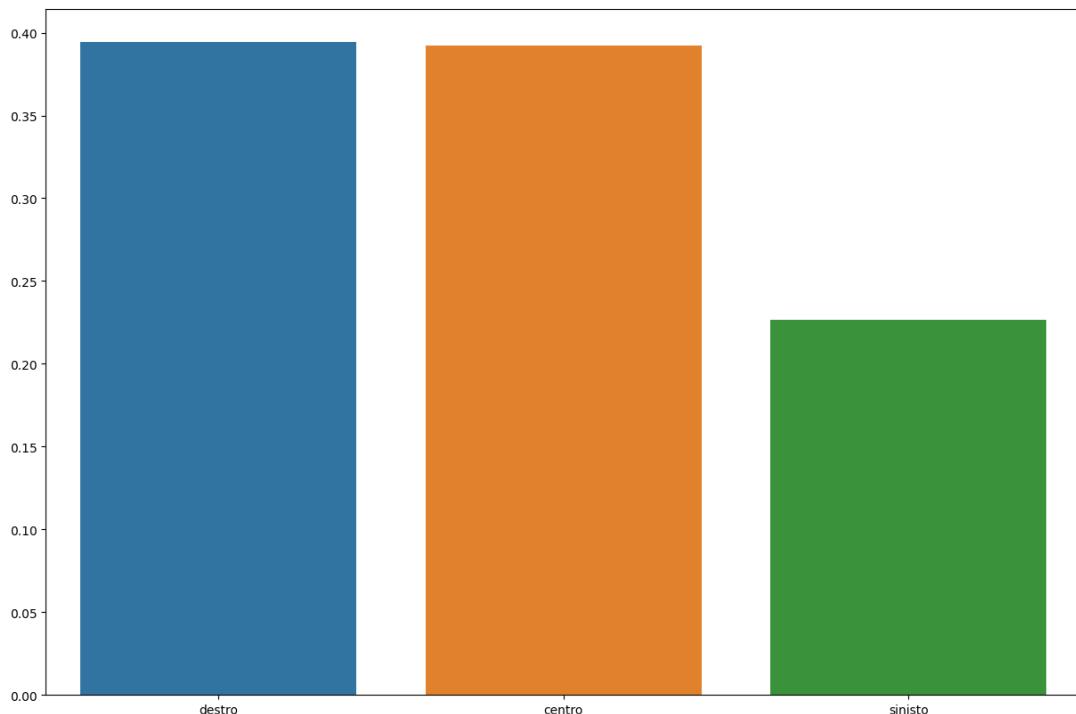


Figura 5.2 - Risultati group degree centrality

Questi sono i risultati ottenuti dall'analisi della group degree centrality. Osserviamo come i valori maggiori siano quelli relativi ai gruppi di nodi che contengono i protagonisti. Infatti, nel centro sono presenti Rick e Morty, i due protagonisti della serie animata, che compaiono in tutti gli episodi, e in tutte le stagioni. Sono anche i gruppi con il maggior numero di personaggi importanti.

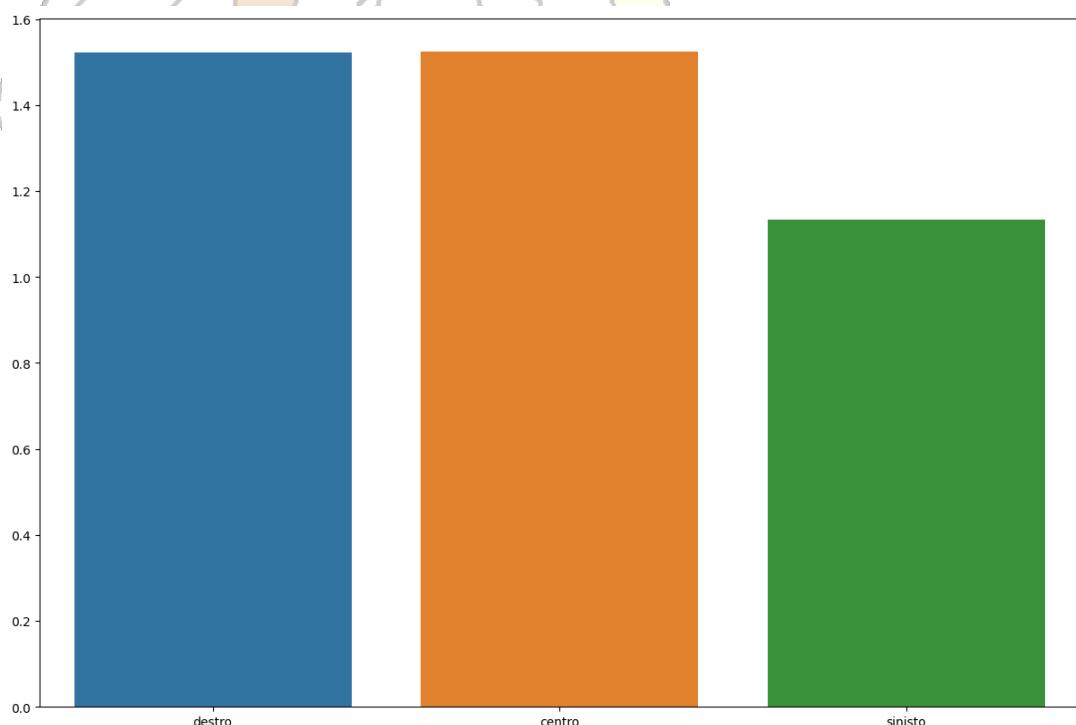


Figura 5.3 - Risultati group closeness centrality



.social Network analysis RICK AND MORTY

Questi sono i risultati ottenuti dall'analisi della group closeness centrality. I risultati ottenuti sono molto simili a quelli precedenti, infatti notiamo come i gruppi che contengono i protagonisti sono anche quelli con i valori maggiori.

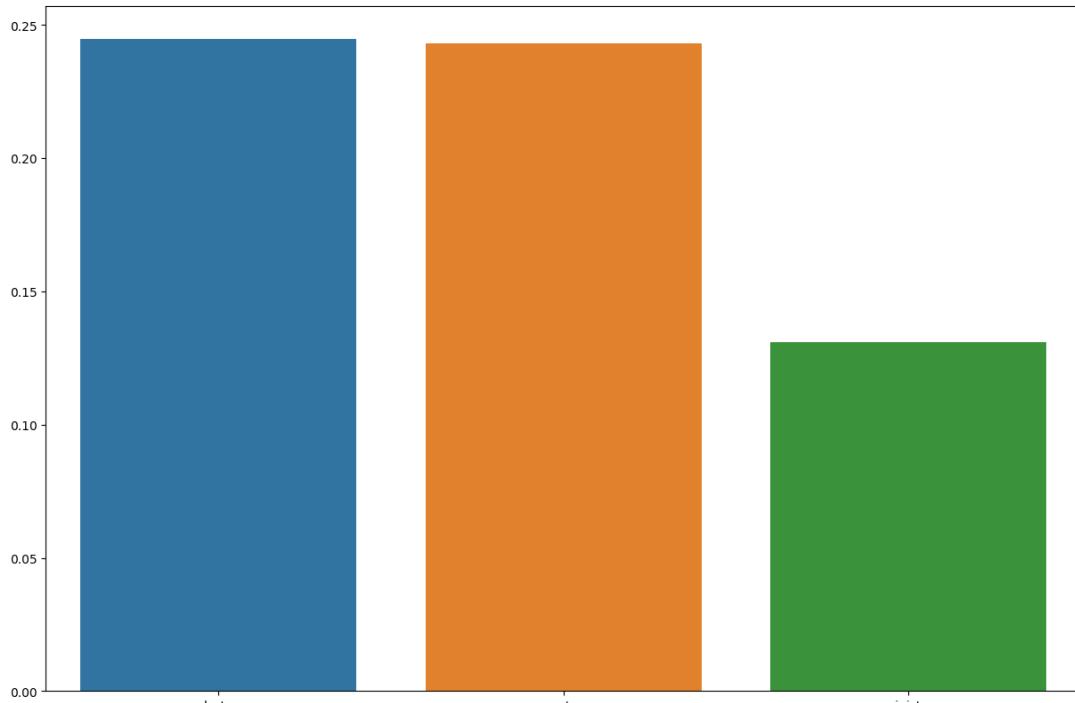


Figura 5.4 – Risultati group betweenness centrality

Infine, questi sono i risultati ottenuti dall'analisi della group betweenness centrality. Osserviamo come, anche questi risultati, sono molto simili ai precedenti, per gli stessi motivi descritti precedentemente. È possibile notare un valore abbastanza basso del gruppo sinistro, questo significa che al suo interno la maggior parte dei nodi non collegano parti abbastanza lontane della rete.



Link al repository GitHub:
<https://github.com/Simone-Scalella/networkxProject>
<https://simone-scalella.github.io/networkxProject/>

**grazie per
l'attenzione**



Zhang
Yihang

Simone
Scalella

Chiara Amalia
Caporusso

Margherita
Galeazzi