

Star Wars Social Network Analysis

1st Matteo Gurrieri

*Tecniche Informatiche per la gestione dei dati
Libera Università Maria Santissima Assunta
Rome, Italy
matteogurrieri@gmail.com*

2nd Riccardo Bianchi

*Tecniche Informatiche per la gestione dei dati
Libera Università Maria Santissima Assunta
Rome, Italy
riccardobnc@live.it*

3rd Paolo Losacco

*Tecniche Informatiche per la gestione dei dati
Libera Università Maria Santissima Assunta
Rome, Italy
p.losacco@outlook.it*

Abstract—Questo studio presenta un'analisi dettagliata del social network basato sulla saga di Star Wars, utilizzando la Social Network Analysis (SNA). Attraverso l'applicazione di tecniche di analisi di rete e l'uso di librerie Python come NetworkX e Pyvis, vengono esplorate le intricate dinamiche e le complesse interazioni tra i personaggi.

Index Terms—Social Network Analysis (SNA), Star Wars Saga, Character Network, Network Visualization, Python Libraries (NetworkX, Pyvis)

I. INTRODUCTION

Nell'era digitale in cui le interazioni sociali avvengono sempre più frequentemente online, l'analisi dei social network ha assunto un ruolo di grande rilevanza. Attraverso questa disciplina, siamo in grado di sondare le articolate dinamiche e le complesse relazioni che si sviluppano all'interno delle comunità virtuali e non.

In questa relazione, presentiamo uno studio dettagliato sull'analisi del social network basato sulla celebre saga di Star Wars, avvalendoci delle tecniche di Social Network Analysis (SNA). La saga di Star Wars ha conquistato un vasto pubblico in tutto il mondo, e l'universo creato da George Lucas ha generato una ricca varietà di personaggi indimenticabili.

Lo scopo principale di questo studio è quello di condurre un'approfondita analisi del social network dei personaggi all'interno della saga, sfruttando un ricco set di dati proveniente dai sette film della serie (Episodi I-VII) che registra tutte le interazioni tra i personaggi.

Nel corso dello studio, applichiamo le procedure apprese nel corso di Social Network Analysis per analizzare il network e ottenere informazioni significative. A tal fine, faremo uso di potenti librerie Python, come NetworkX per la manipolazione dei dati di rete e Pyvis per la visualizzazione interattiva del social network.

II. DATASET DESCRIPTION AND CHOICE'S MOTIVATION

A. Dataset description

Il dataset è stato ottenuto da <https://icon.colorado.edu/#/networks> ed è fornito in formato JSON. Il file "starwars-

full-interactions-allCharacters.json" rappresenta l'intera rete sociale dei personaggi presenti in tutti i film, dall'Episodio I all'Episodio VII, dove i collegamenti tra i personaggi sono definiti dalle volte in cui i personaggi parlano nella stessa scena.

I nodi del social network sono caratterizzati dalle seguenti informazioni:

- Nome: Il nome del personaggio.
- Valore: Il numero di scene in cui il personaggio compare.
- Colore: Il colore associato nella visualizzazione del social network.

I collegamenti rappresentano le connessioni tra i personaggi e sono definiti dai seguenti attributi:

- Sorgente: Indice in base zero del personaggio che rappresenta un'estremità del collegamento, l'ordine dei nodi corrisponde all'ordine in cui sono elencati nell'elemento "nodi".
- Destinazione: Indice in base zero del personaggio che rappresenta l'altra estremità del collegamento.
- Valore: Il numero di scene in cui il "personaggio di origine" e il "personaggio di destinazione" del collegamento compaiono insieme.

È importante sottolineare che la rete è non diretta, il che significa che i collegamenti non hanno una direzione specifica. Questo dataset ci fornisce un quadro completo del social network di Star Wars, consentendoci di esplorare le relazioni e le dinamiche tra i vari personaggi.

B. Choice's motivation

Il nostro interesse personale per Star Wars come appassionati della saga ha giocato un ruolo fondamentale nella scelta di condurre questo studio. Siamo stati affascinati dalla vastità e dalla complessità del mondo di Star Wars, e desideriamo approfondire la comprensione delle relazioni tra i personaggi, l'importanza di alcuni personaggi nel network e il loro impatto sulla trama complessiva.

III. ANALYSIS DISCONNECTED GRAPH

Nella fase di analisi del nostro studio, abbiamo eseguito diverse operazioni per estrarre informazioni rilevanti dal grafo rappresentante la rete di personaggi. Inizialmente, abbiamo creato due rappresentazioni del grafo: una spring e una circolare.

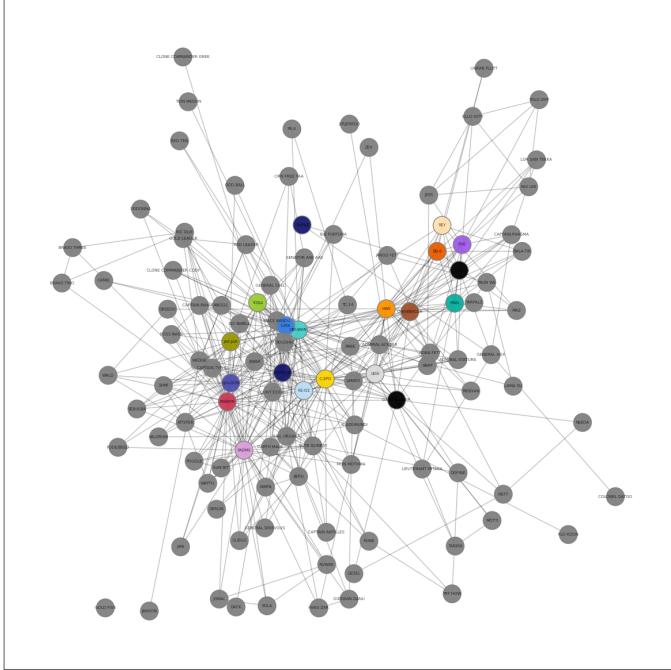


Fig. 1. Spring Layout Plot.

Successivamente, abbiamo calcolato diverse misure del grafo per ottenere una visione completa delle sue caratteristiche. Abbiamo determinato il numero di nodi presenti e il numero di collegamenti tra di essi.

TABLE I
STRUTTURA GRAFO

Nodi	Archi
112	450

Inoltre, abbiamo calcolato il grado medio del grafo e la densità, ovvero il rapporto tra i collegamenti effettivi e i collegamenti possibili. Questa misura ci ha aiutato a comprendere quanto il grafo fosse interconnesso.

Abbiamo anche analizzato la presenza di triangoli nel grafo, il che indica la tendenza dei personaggi a formare gruppi. Per valutare la struttura a grappolo della rete, abbiamo calcolato il coefficiente medio di clustering e la transitivity, che rappresenta il coefficiente globale di clustering. Questi valori ci hanno permesso di valutare quanto i personaggi fossero collegati tra di loro all'interno di gruppi.

Infine, abbiamo calcolato il coefficiente di assortatività, che descrive il grado di tendenza dei nodi a collegarsi con altri nodi dello stesso grado.

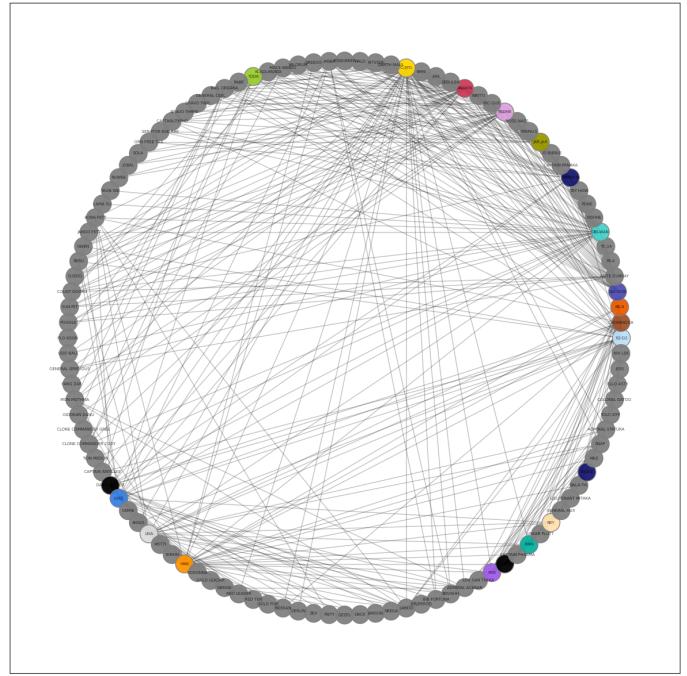


Fig. 2. Circular Layout Plot.

TABLE II
MISURE GRAFO

Misura	Valore
Densità	0.072
Degree Medio	8.036
Coefficiente medio clustering	0.687
Coefficiente globale di clustering	0.365
Coefficiente di assortatività	-0.186
Numero di triangoli	859

Densità: La densità del grafo è 0.072, che rappresenta la frazione di connessioni effettivamente presenti rispetto al numero massimo di connessioni possibili. In altre parole, indica quanto il grafo sia interconnesso. Una densità più alta indica una rete più densamente connessa. Nel caso di una densità del grafo pari a 0.072, significa che il grafo ha una bassa densità. Ciò indica che le connessioni tra i nodi nel grafo sono relativamente scarse o sparse.

Degree Medio: Il degree medio è 8.036, che rappresenta la media dei gradi di tutti i nodi nel grafo. Il grado di un nodo indica il numero di archi che sono collegati ad esso. Un degree medio più alto indica una rete con un numero medio elevato di connessioni per nodo.

Coefficiente medio di clustering: Il coefficiente medio di clustering è 0.687, che rappresenta la media dei coefficienti di clustering di tutti i nodi nel grafo. Il coefficiente di clustering di un nodo misura la tendenza dei suoi vicini ad essere anche collegati tra loro. Un coefficiente medio di clustering più alto indica una rete con una maggiore tendenza alla formazione di

gruppi o cluster di nodi interconnessi. Un coefficiente medio di clustering pari a 0.687 indica che, in media, i nodi nel grafo hanno una forte propensione a formare cluster locali. Questo valore suggerisce che vi è una tendenza significativa per i nodi ad essere collegati tra loro e a formare gruppi coesi all'interno del grafo.

Coefficiente globale di clustering: Il coefficiente globale di clustering è 0.365, che rappresenta la frazione di triangoli presenti nel grafo rispetto al numero massimo di triangoli possibili. Misura la presenza di clustering nella rete nel suo complesso. Un coefficiente di clustering globale più alto indica una rete con una maggiore tendenza alla formazione di cluster. Un coefficiente globale di clustering pari a 0.365 suggerisce che il grafo mostra una moderata tendenza alla formazione di cluster locali, ma potrebbe avere anche alcune regioni meno dense o meno connesse.

Coefficiente di assortatività: indica la tendenza dei nodi ad essere collegati a quelli con un grado simile. Un coefficiente di assortatività negativo indica una tendenza alla disassortatività, cioè i nodi con alto grado sono più propensi a essere collegati a quelli con basso grado e viceversa. Il coefficiente di assortatività di -0.186 indica una leggera tendenza verso la disassortatività nel grafo. Ciò significa che i nodi con attributi simili hanno meno probabilità di essere collegati tra loro rispetto ai nodi con attributi diversi.

Numero di triangoli: Il numero di triangoli nel grafo è 859. Questo rappresenta il numero totale di triple di nodi che formano un triangolo nella rete. I triangoli sono importanti per identificare le strutture di clustering e la coesione nella rete.

Per ulteriori analisi, abbiamo generato la matrice di adiacenza e la matrice di distanza del grafo. Queste matrici ci hanno fornito informazioni dettagliate sulla connettività tra i nodi e sulle distanze tra di essi.

A. Degree Centrality

La Degree Centrality (centralità di grado) è una misura di centralità dei nodi che indica il numero di collegamenti che un nodo ha nel grafo. In altre parole, misura il numero di archi di un nodo.

$$C_D(j) = \sum_{j=1}^N A_{ij} \quad (1)$$

I valori della Degree Centrality rappresentano la frazione del numero massimo possibile di collegamenti che un nodo ha rispetto a tutti gli altri nodi nel grafo. In altre parole, misurano la centralità del nodo basata sul numero di connessioni che ha con altri nodi.

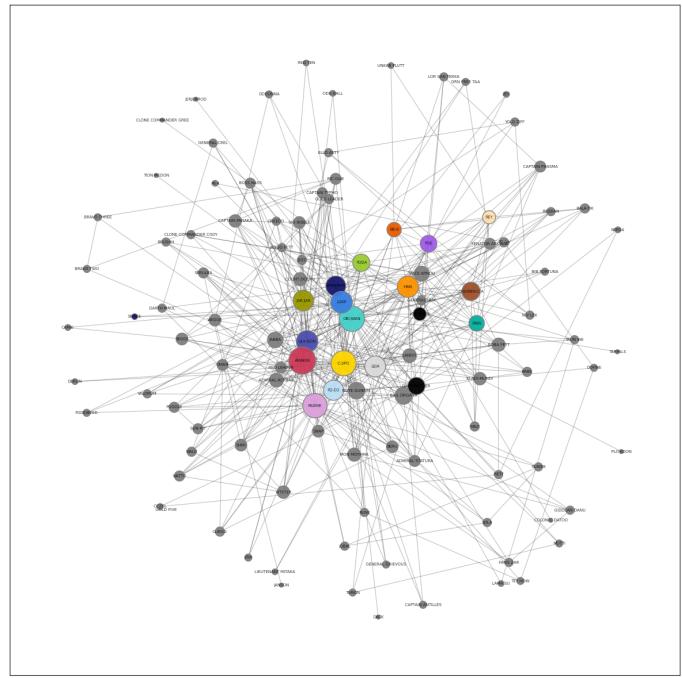


Fig. 3. Degree Centrality Plot.

TABLE III
TOP 8 DEGREE AND DEGREE CENTRALITY

Node	Degree	Degree Centrality
ANAKIN	42	0.378
OBI-WAN	37	0.333
C-3PO	36	0.324
PADME	34	0.306
QUI-GON	27	0.243
LUKE	27	0.243
HAN	26	0.234
LEIA	25	0.225

ANAKIN ha il grado più alto e la Degree Centrality più alta, indicando che è il nodo con il maggior numero di collegamenti nel grafo e quindi ha una maggiore centralità di grado rispetto agli altri nodi.

OBI-WAN, C-3PO e PADME seguono da vicino con gradi e Degree Centrality elevate.

QUI-GON e LUKE hanno lo stesso grado e Degree Centrality, indicando che hanno un numero uguale di collegamenti e quindi una centralità di grado identica.

HAN e LEIA hanno valori leggermente inferiori di grado e Degree Centrality rispetto agli altri nodi, ma comunque significativi, indicando che hanno un numero inferiore di collegamenti rispetto ai nodi precedenti.

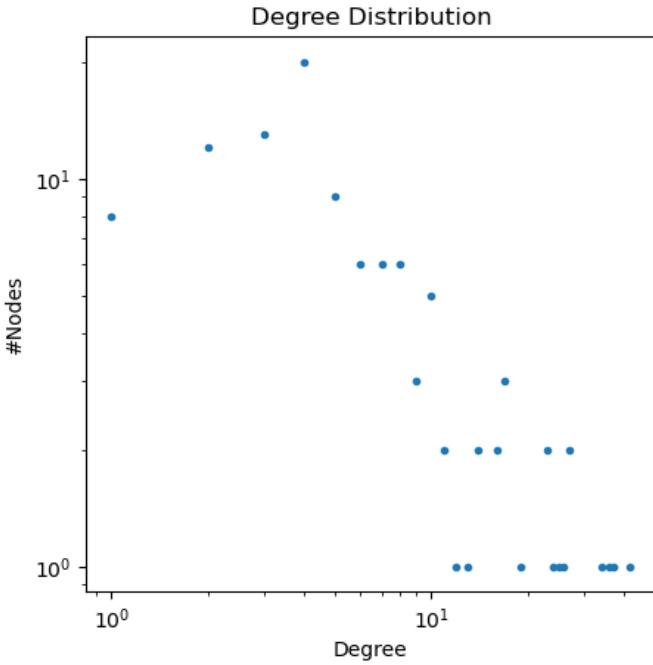


Fig. 4. Degree Distribution Scatterplot.

Abbiamo inoltre generato due rappresentazioni grafiche per illustrare la distribuzione dei gradi dei personaggi: uno scatterplot in scala logaritmica e un istogramma. Dall'analisi dei grafici emerge chiaramente che la maggior parte dei personaggi presenta un grado compreso tra 0 e 10, con particolare concentrazione sul grado 4.

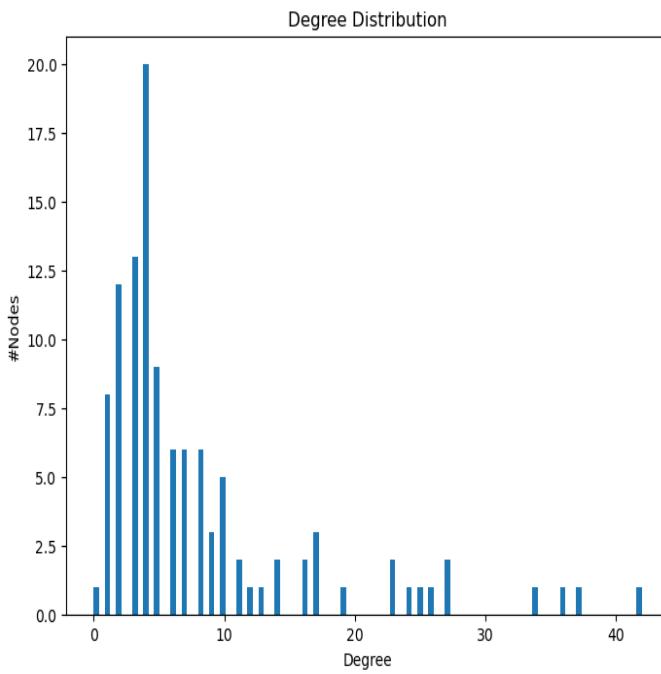


Fig. 5. Degree Distribution Histogram.

B. Betweenness Centrality

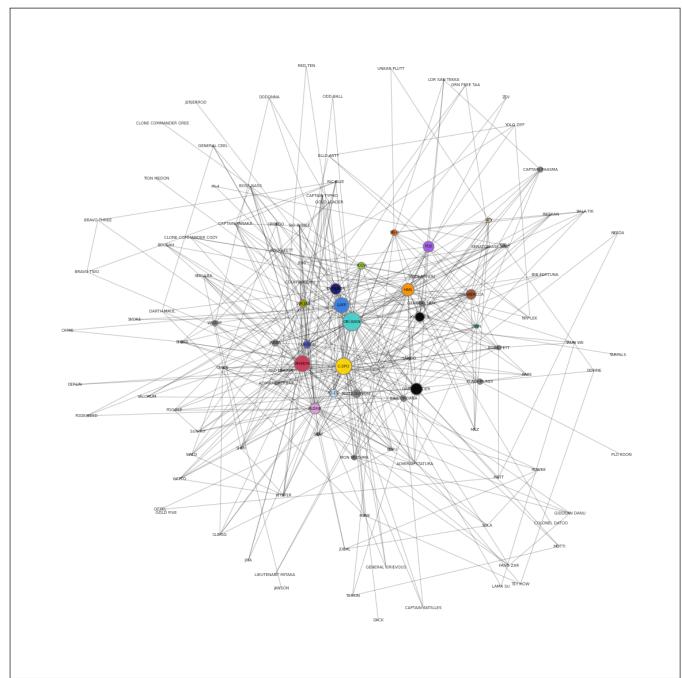


Fig. 6. Betweenness Centrality Plot.

La Betweenness Centrality (centralità di betweenness) è un'importante misura di centralità dei nodi in un grafo, che indica la quantità di influenza o controllo che un nodo ha nel trasferimento delle informazioni tra altri nodi.

$$C_i^{BET} = \sum_{i=1}^N \sum_{j=1}^N \frac{g_{ikj}}{g_{ij}} \quad (2)$$

I valori rappresentano la frazione dei cammini più brevi che passano attraverso ciascun nodo rispetto a tutti i cammini più brevi presenti nel grafo. Maggiore è il valore della Betweenness Centrality di un nodo, maggiore è la sua importanza nel controllo delle informazioni che fluiscono attraverso la rete.

TABLE IV
TOP 8 BETWEENNESS CENTRALITY

Node	Betweenness Centrality
OBI-WAN	0.205
C-3PO	0.168
ANAKIN	0.158
LUKE	0.142
HAN	0.096
DARTH VADER	0.085
EMPEROR	0.075
POE	0.073

OBI-WAN ha la più alta Betweenness Centrality, il che indica che ha un ruolo significativo nel facilitare la comunicazione tra altri nodi nel grafo.

Successivamente, emergono C-3PO e ANAKIN con valori di Betweenness Centrality elevati, suggerendo che anche loro

svolgono un ruolo importante nel trasferimento delle informazioni tra i nodi.

LUKE e HAN hanno valori leggermente più bassi, ma comunque significativi, nella Betweenness Centrality, indicando una certa influenza nel flusso delle informazioni nel grafo.

DARTH VADER, EMPEROR e POE hanno valori inferiori di Betweenness Centrality, indicando una minore influenza nel trasferimento delle informazioni rispetto agli altri nodi menzionati sopra.

C. Closeness Centrality

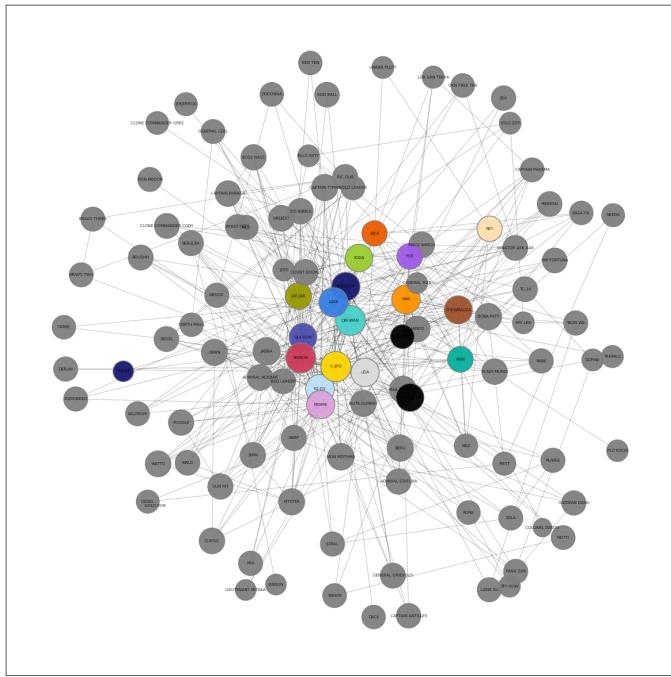


Fig. 7. Closeness Centrality Plot.

La Closeness Centrality (centralità di vicinanza) è una misura di centralità dei nodi che indica quanto un nodo è "vicino" a tutti gli altri nodi nel grafo in termini di distanza. Misura la reciproca della media delle distanze più brevi tra un nodo e tutti gli altri nodi nel grafo.

$$C_i^{CLOS} = \frac{N - 1}{\sum_{j=1}^N d_{ij}} \quad (3)$$

Più il valore della Closeness Centrality di un nodo è alto, più il nodo è "vicino" agli altri nodi in termini di distanza e quindi più rapido è il suo accesso alle informazioni che possono essere trasmesse attraverso il grafo.

TABLE V
TOP 8 CLOSENESS CENTRALITY

Node	Closeness Centrality
C-3PO	0.562
OBI-WAN	0.559
ANAKIN	0.551
LUKE	0.527
R2-D2	0.517
HAN	0.512
EMPEROR	0.502
LEIA	0.493

C-3PO ha la Closeness Centrality più alta, il che indica che è il nodo che è "più vicino" agli altri nodi in termini di distanza media. È in grado di raggiungere più rapidamente gli altri nodi nel grafo rispetto agli altri.

A seguire troviamo OBI-WAN e ANAKIN con valori di Closeness Centrality elevati, indicando che anche loro sono "vicini" agli altri nodi e hanno un buon accesso alle informazioni che passano attraverso il grafo.

LUKE e R2-D2 presentano valori leggermente inferiori di Closeness Centrality, tuttavia rimangono comunque significativi, indicando che hanno un accesso relativamente rapido alle informazioni e sono strettamente collegati agli altri nodi nel grafo.

HAN, EMPEROR e LEIA hanno valori inferiori di Closeness Centrality, indicando che potrebbero richiedere un po' più di tempo per accedere alle informazioni dagli altri nodi rispetto a quelli con Closeness Centrality più alta.

In generale, osservando attentamente il grafico, possiamo notare come tutti i nodi presentino valori simili, i quali diminuiscono man mano che ci spostiamo verso la periferia del grafo. È interessante notare che un nodo in particolare, "GOLD FIVE", risulta praticamente invisibile nel grafo in quanto è un nodo isolato e presenta un valore di centralità pari a 0.

D. Page Rank

Sebbene l'algoritmo di PageRank sia comunemente utilizzato per i grafi direzionali, è possibile applicarlo anche ai grafi non direzionali. Abbiamo deciso di utilizzarlo al fine di determinare quale personaggio fosse considerato il più "importante" nel contesto del nostro grafo.

La PageRank Centrality (centralità di PageRank) è una misura di centralità dei nodi che indica l'importanza di un nodo all'interno di una rete. È basata sull'algoritmo di PageRank, sviluppato da Google per valutare l'importanza delle pagine web nel Web.

$$PR[A] = \frac{1-d}{N} + d \left(\sum_{i=1}^N \frac{PR[P_k]}{C[P_k]} \right) \quad (4)$$

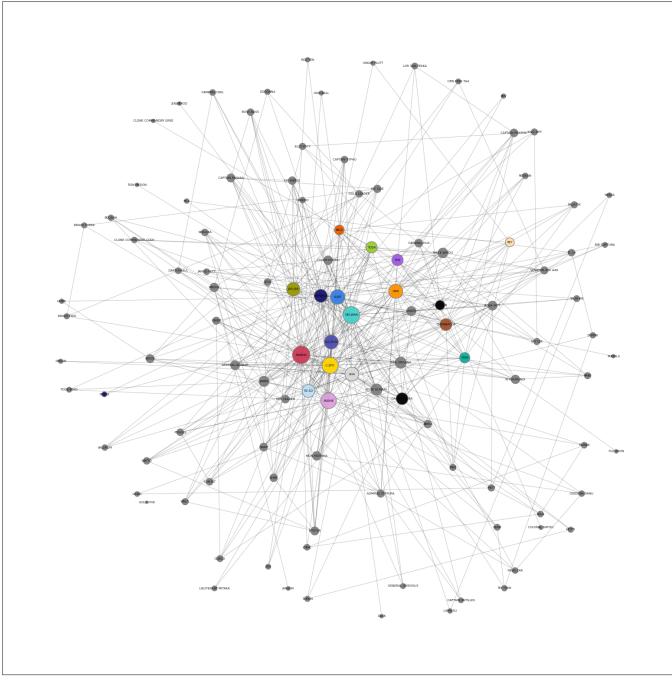


Fig. 8. PageRank Centrality Plot.

I valori della PageRank Centrality generalmente rappresentano l'importanza di un nodo all'interno del grafo, considerando sia il numero di link in ingresso che la qualità di tali link (ovvero l'importanza dei nodi che li generano). Più alto è il valore della PageRank Centrality di un nodo, maggiore è la sua importanza nella rete.

TABLE VI
TOP 8 PAGE RANK CENTRALITY

Node	Page Rank Centrality
ANAKIN	0.042
OBI-WAN	0.039
C-3PO	0.036
PADME	0.034
LUKE	0.029
QUI-GON	0.027
HAN	0.027
LEIA	0.026

ANAKIN ha il valore più alto di PageRank Centrality, indicando che è il nodo più importante all'interno del grafo secondo l'algoritmo di PageRank.

OBI-WAN, C-3PO e PADME hanno valori simili di PageRank Centrality, indicando che hanno un'importanza comparabile nel grafo.

LUKE ha un valore leggermente inferiore di PageRank Centrality, ma comunque significativo, indicando che ha un'importanza riconosciuta nel grafo.

QUI-GON, LUKE e LEIA hanno valori inferiori di PageRank Centrality, indicando una minore importanza rispetto agli altri nodi menzionati sopra.

Come abbiamo già osservato, la Degree Centrality è un'altra misura utilizzata per identificare i nodi più "importanti". La

differenza tra queste due misure potrebbe essere dovuta al fatto che il PageRank tiene conto anche dell'importanza delle "pagine" che collegano i nodi, mentre la Degree Centrality considera solo il numero di connessioni. Tuttavia, per entrambe le metriche, Anakin si è classificato al primo posto in modo costante.

E. Eigenvector Centrality

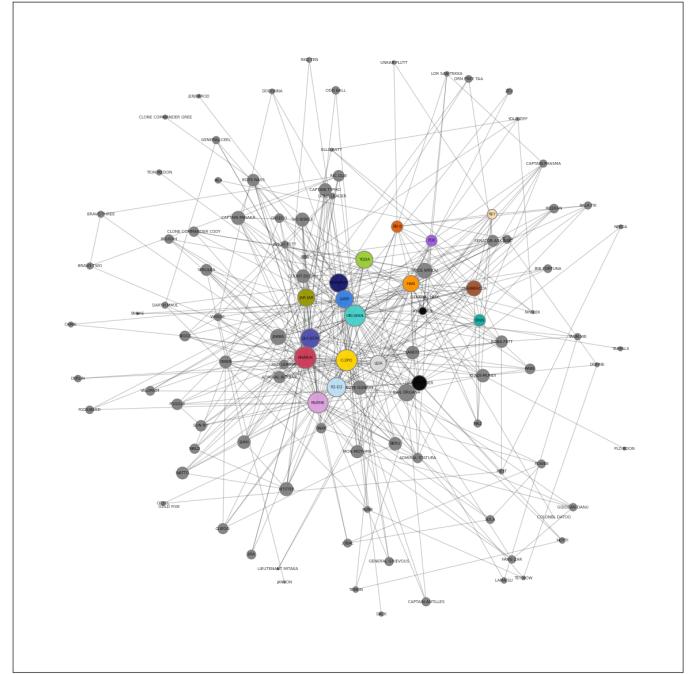


Fig. 9. Eigenvector Centrality Plot.

La Eigenvector Centrality (centralità dell'autovettore) è una misura che valuta l'importanza di un nodo in un grafo basandosi sulla sua connessione con altri nodi importanti. La centralità dell'autovettore di un nodo dipende dalla centralità dei nodi ad esso collegati.

$$x_i = \frac{1}{\lambda} \sum_k a_{k,i} x_k \quad (5)$$

TABLE VII
TOP 8 EIGENVECTOR CENTRALITY

Node	Eigenvector Centrality
ANAKIN	0.297
OBI-WAN	0.281
C-3PO	0.275
PADME	0.258
QUI-GON	0.229
R2-D2	0.228
EMPEROR	0.211
LUKE	0.191

ANAKIN ha una Eigenvector Centrality più alta, il che indica che è collegato ad altri nodi di alta importanza nel grafo.

OBI-WAN e C-3PO si posizionano in prossimità l'uno dell'altro con valori elevati di Eigenvector Centrality, evidenziando la loro forte connessione con altri nodi di rilievo nel grafo.

PADME, QUI-GON e R2-D2 hanno valori leggermente inferiori di Eigenvector Centrality, ma comunque rilevanti.

EMPEROR e LUKE hanno valori inferiori di Eigenvector Centrality, sebbene in misura leggermente inferiore rispetto agli altri nodi evidenziati.

F. Communities

Abbiamo condotto un'analisi delle comunità all'interno del grafo utilizzando la funzione "best partition" che è basata sull'algoritmo di Louvain. Quest'ultimo calcola la partizione dei nodi del grafo che massimizza la modularità (o prova a farlo) utilizzando l'euristiche di Louvain. I risultati hanno mostrato l'esistenza di sei comunità distintive.

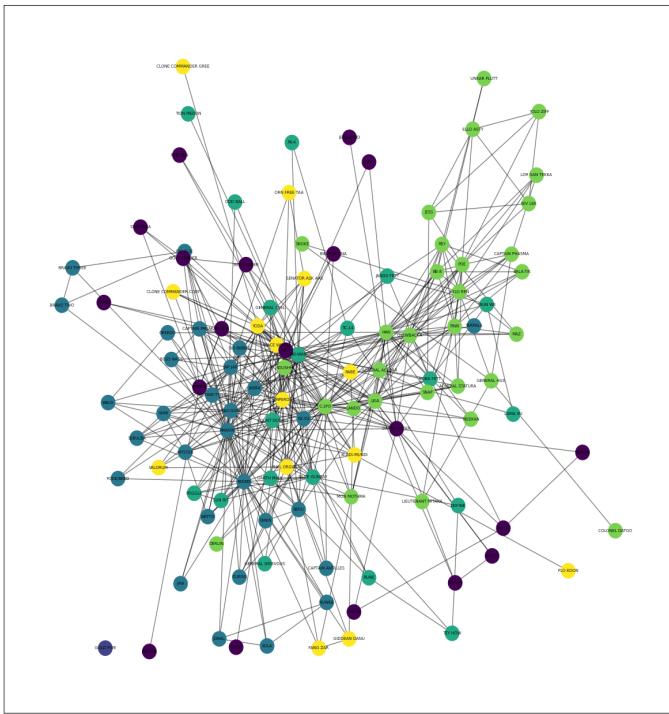


Fig. 10. Best Partition Plot.

TABLE VIII
BEST PARTITION COMMUNITIES

Communities		
Community 0	Community 1	Community 2
DARTH VADER LUKE CAMIE BIGGS MOTTI TARKIN DODONNA GOLD LEADER WEDGE RED LEADER RED TEN ZEV PIETT OZZEL DACK JANSON NEEDA JERJERROD BIB FORTUNA	GOLD FIVE	R2-D2 QUI-GON CAPTAIN PANAKA SIO BIBBLE JAR JAR TARPALS BOSS NASS PADME RIC OLINE WATTO ANAKIN SEBULBA JIRA SHMI KITISTER WALD FODE/BEED JABBA GREEDO BRAVO TWO BRAVO THREE CAPTAIN TYPHO SOLA JOBAL RUWEE OWEN BERU CLIEGG CAPTAIN ANTILLES
NUTE GUNRAY PK-4 TC-14 OBI-WAN DOFINE RUNE TEY HOW DARTH MAUL GENERAL CEEL TAUN WE LAMA SU BOBA FETT JANGO FETT COUNT DOOKU SUN RIT POGGLE ODD BALL GENERAL GRIEVOUS TION MEDON	CHEWBACCA BB-8 C-3PO MON MOTHMA LEIA HAN RIEKAN DERLIN LANDO BOUSHH ADMIRAL ACKBAR LOR TEKKA POE KYLO REN CAPTAIN PHASMA FINN UNKAR PLUTT REY GENERAL HUX LIEUTENANT MITAKA BALA-TIK SNOKE MAZ SNAP ADMIRAL STATURA YOLO ZIFF COLONEL DATOO ELLO ASTY JESS NIV LEK	EMPEROR VALORUM MACE WINDU KI-ADI-MUNDI YODA RABE BAIL ORGANA SENATOR ASK AAK ORN FREE TAA PLO Koon FANG ZAR GIDDEAN DANU CLONE COMMANDER GREE CLONE COMMANDER CODY

La prima comunità è caratterizzata dalla presenza di personaggi appartenenti al quarto film di Star Wars, "A New Hope". Al centro di questa comunità si trovano personaggi come DARTH VADER, e nella lista compaiono anche piloti, compagni di LUKE e altri personaggi che hanno lavorato sulla Morte Nera insieme a DARTH VADER. Sorprendentemente, BIB FORTUNA si trova in questa comunità nonostante non sia apparso nel quarto film della saga. Tuttavia, probabilmente è stato inserito in questa comunità poiché è un personaggio legato a LUKE, dato il loro incontro nel sesto film della saga.

La seconda comunità è composta da un solo personaggio, GOLD FIVE. Essendo un pilota, avremmo presumibilmente incluso questo personaggio nella prima comunità. Tuttavia, è stato collocato in una comunità separata poiché, analizzando il grafo, si è constatato che non ha alcuna connessione con gli altri personaggi.

La terza comunità è caratterizzata dai personaggi che compaiono nel primo film della saga, "Phantom Menace", come QUI-GON e JAR JAR. Questi personaggi sono stati

raggruppati insieme perché sono associati a quel particolare episodio della storia.

La quarta comunità rappresenta i personaggi legati ai prequel, in particolare a OBI-WAN. Molti dei personaggi in questa comunità sono i suoi antagonisti o personaggi con cui ha avuto interazioni significative durante quei film.

La quinta comunità è costituita dai personaggi legati al settimo film, "The Force Awakens". Questi personaggi sono stati raggruppati insieme perché sono associati a quella fase della saga.

Infine, la sesta e ultima comunità include personaggi di grande importanza legati a ruoli di leadership e autorità politica, tra cui i Jedi. Questa comunità rappresenta personaggi che ricoprono posizioni di potere e influenzano gli eventi chiave all'interno dell'universo di Star Wars.

Abbiamo inoltre applicato l'algoritmo di Girvan-Newman fornito dal pacchetto "community". Questo algoritmo rileva le comunità rimuovendo gradualmente gli archi del grafo originale. Ad ogni passo, viene rimosso l'arco considerato "più significativo", generalmente l'arco con la maggiore Betweenness Centrality. Utilizzando questo algoritmo, abbiamo identificato con successo tre comunità distinte all'interno del grafo.

La prima comunità comprende i personaggi che compaiono dal quarto al settimo film della saga di Star Wars.

La seconda comunità è composta dai personaggi dei prequel, ovvero i film dal primo al terzo.

Infine, l'ultima comunità è caratterizzata dalla presenza di GOLD FIVE come unico personaggio. Questo personaggio risulta sempre disconnesso dall'intero grafo, il che significa che non ha collegamenti con nessun altro personaggio.

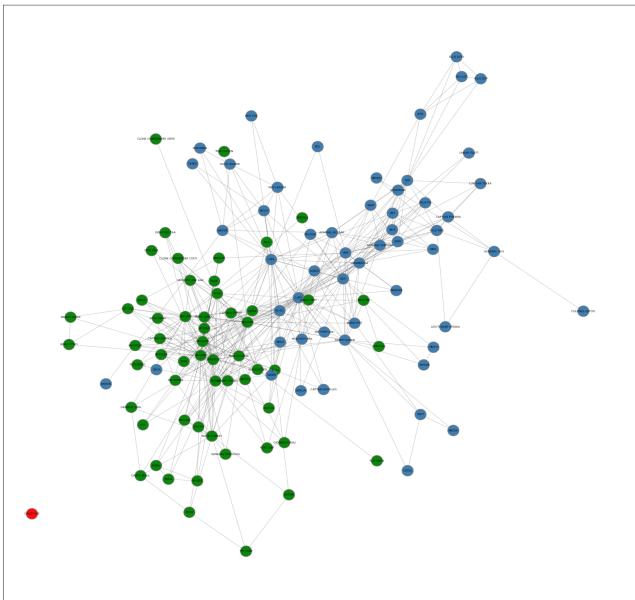


Fig. 11. Girvan Newman Plot.

TABLE IX
GIRVAN NEWMAN COMMUNITIES

Communities		
Community 0	Community 1	Community 2
LIEUTENANT MITAKA	TC-14	GOLD FIVE
BERU	RUWEE	
YOLO ZIFF	BRAVO TWO	
NIV LEK	EMPEROR	
NEEDA	GREEDO	
RED LEADER	JABBA	
CAPTAIN PHASMA	TEY HOW	
MOTTI	PK-4	
SNOKE	DARTH MAUL	
R2-D2	FODE/BEED	
BOBA FETT	PADME	
JERJERROD	JANGO FETT	
DARTH VADER	GENERAL GRIEVOUS	
POE	SOLA	
RIEKAN	TION MEDON	
DACK	KI-ADI-MUNDI	
DODONNA	KITSTER	
WEDGE	CLIEGG	
CHEWBACCA	CAPTAIN TYPHO	
BALA-TIK	SENATOR ASK AAK	
UNKAR PLUTT	WALD	
BIGGS	JOBAL	
DERLIN	TARPALS	
CAPTAIN ANTILLES	LAMA SU	
REY	BOSS NASS	
JESS	SHMI	
HAN	BRAVO THREE	
BIB FORTUNA	WATTO	
ADMIRAL STATURA	VALORUM	
ELLO ASTY	SEBULBA	
OZZEL	DOFINE	
FINN	COUNT DOOKU	
JANSON	BAIL ORGANA	
BOUSHH	YODA	
KYLO REN	FANG ZAR	
COLONEL DATOO	ANAKIN	
LUKE	PLO KOON	
BB-8	CLONE COMMANDER CODY	
GOLD LEADER	ODD BALL	
ZEV	RABE	
C-3PO	QUI-GON	
LANDO	NUTE GUNRAY	
MAZ	GIDDEAN DANU	
ADMIRAL ACKBAR	CAPTAIN PANAKA	
LOR SAN TEKKA	GENERAL CEEL	
GENERAL HUX	POGGLE	
LEIA	JAR JAR	
SNAP	CLONE COMMANDER GREE	
MON MOTHMA	RUNE	
PIETT	OBI-WAN	
TARKIN	JIRA	
RED TEN	ORN FREE TAA	
CAMIE	SIO BIBBLE	
OWEN	RIC OLIE	

G. Bridges

Abbiamo analizzato il grafo per identificare le coppie di nodi connesse da un bridge, distinguendo tra bridge globale (coppie con bridge) e bridge locali (coppie con local bridge). Abbiamo evidenziato i bridge con il colore rosso e i local bridge con il colore verde.

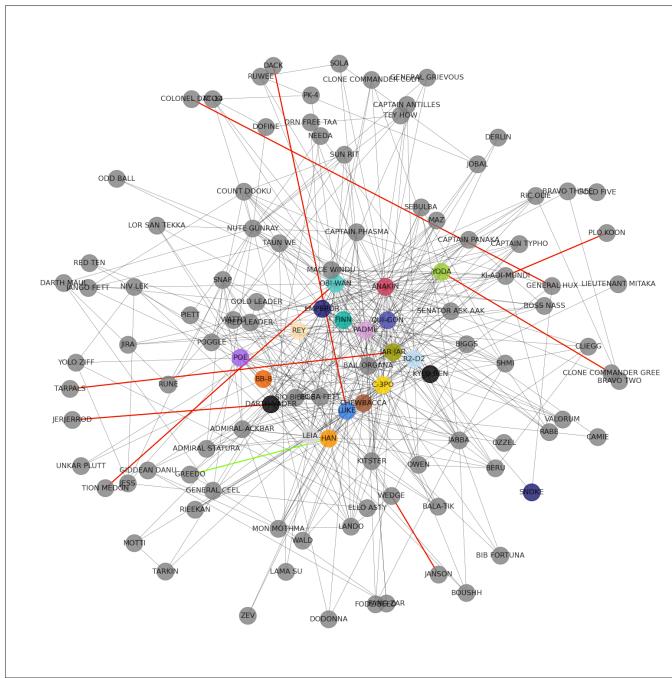


Fig. 12. Global and Local Bridges Plot.

È importante notare che tutti i bridge sono anche local bridge, ma non tutti i local bridge sono bridge. La coppia tra GREEDO e HAN, è un local bridge e non è considerato un bridge. Questo indica che la connessione tra GREEDO e HAN ha un'importanza rilevante solo all'interno di una porzione specifica del grafo e potrebbe non influenzare le altre parti in modo significativo.

H. Ego Network

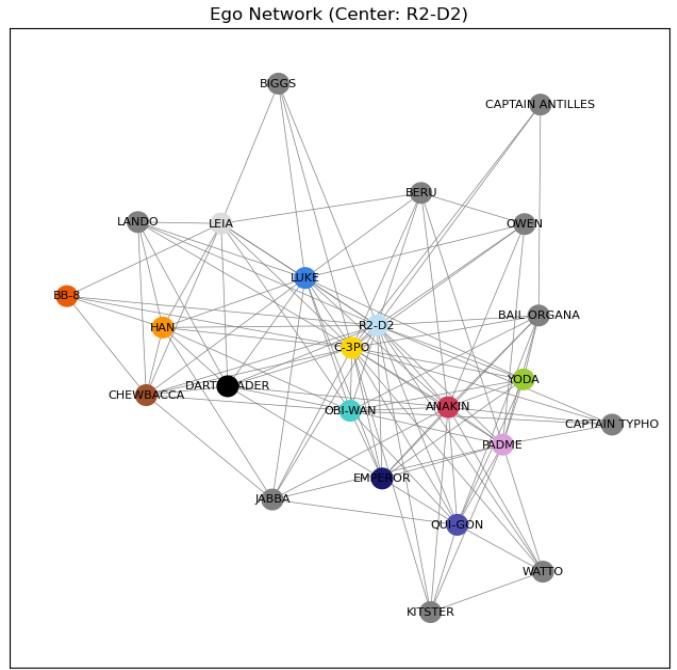


Fig. 13. Ego Network Plot.

TABLE X
BRIDGES

Bridges	
	Local Bridges
OBI-WAN — TION MEDON	OBI-WAN — TION MEDON
JAR JAR — TARPALS	JAR JAR — TARPALS
KI-ADI-MUNDI — PLO KOON	KI-ADI-MUNDI — PLO KOON
YODA — CLONE COMMANDER GREE	YODA — CLONE COMMANDER GREE
DARTH VADER — JERJERROD	DARTH VADER — JERJERROD
LUKE — DACK	LUKE — DACK
WEDGE — JANSON	WEDGE — JANSON
GENERAL HUX — COLONEL DATOO	GENERAL HUX — COLONEL DATOO

Le coppie con bridge rappresentano connessioni tra nodi che svolgono il ruolo di ponti nel grafo. Questi collegamenti sono rilevanti perché permettono di unire parti del grafo che altrimenti sarebbero separate.

Le coppie con local bridge, invece, indicano connessioni tra nodi che fungono da ponti locali all'interno di una specifica porzione del grafo. Queste connessioni sono significative all'interno di una porzione specifica del grafo, ma potrebbero non avere un impatto globale sulle altre parti.

Abbiamo poi sviluppato un codice Python per generare e visualizzare un ego network basato sull'input fornito, ossia il personaggio specificato. Un ego network rappresenta il sottoinsieme di nodi e archi che sono direttamente connessi al nodo di interesse, chiamato "ego".

E' possibile inserire il nome di un personaggio come input e ottenere un grafo che mostra le connessioni dirette di quel personaggio con altri personaggi nel contesto dell'universo di Star Wars.

Questo ci consente di studiare l'importanza del personaggio all'interno della rete sociale di Star Wars e comprendere meglio le sue interazioni con altri personaggi chiave.

I. Shortest Path Length tra 2 Personaggi

Sempre in Python, abbiamo creato un codice per calcolare la shortest path length, ovvero la lunghezza del percorso più breve, tra due personaggi specificati come input.

Questa misura quantifica il numero minimo di archi che devono essere attraversati per raggiungere un personaggio dall'altro.

La shortest path length ci fornisce una metrica utile per comprendere la vicinanza e la connettività tra due personaggi nell'universo di Star Wars.

J. Grafo Interattivo

Abbiamo poi deciso di utilizzare la libreria Pyvis per costruire un grafo interattivo, rendendo così la visualizzazione dei dati più dinamica e intuitiva.

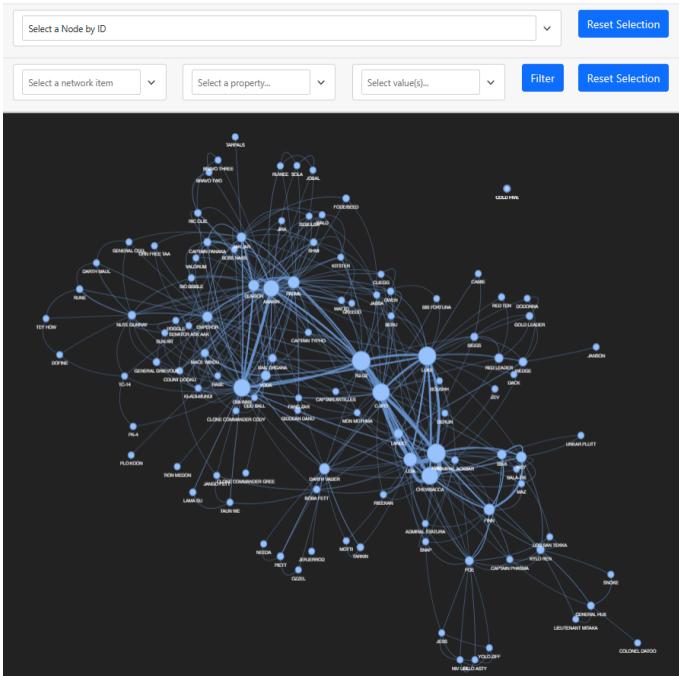


Fig. 14. Interactive Plot.

IV. ANALYSIS CONNECTED GRAPH

Durante l'analisi, abbiamo individuato un nodo disconnesso chiamato "GOLD FIVE". Dopo aver rimosso questo nodo, siamo stati in grado di calcolare il diametro del grafo, che rappresenta la massima distanza tra due nodi, e la lunghezza media del percorso più breve.

TABLE XI
STRUTTURA GRAFO

Grafo	Nodi	Archi
Disconnesso	112	450
Connesso	111	450

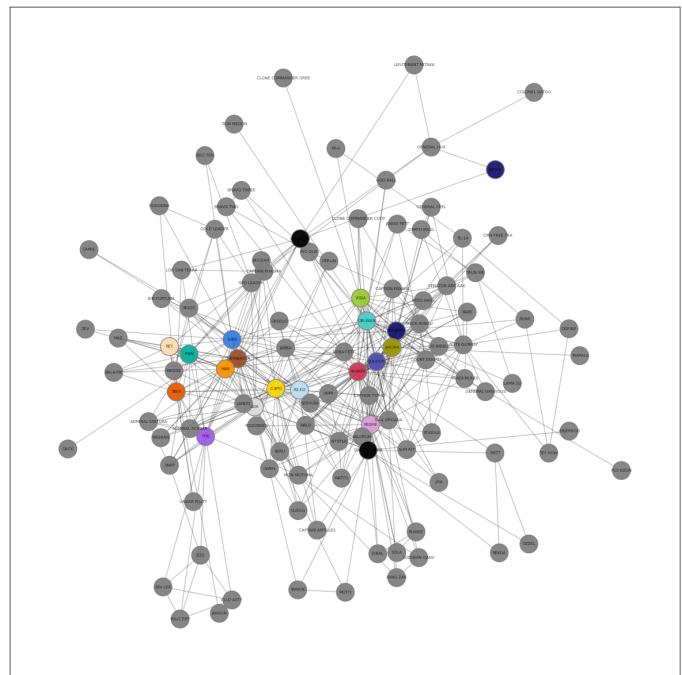


Fig. 15. Spring Layout Plot.

TABLE XII
MISURE GRAFO

Misura	Disconnesso	Connesso
Densità	0.072	0.074
Degree Medio	8.036	8.108
Coefficiente medio clustering	0.687	0.693
Coefficiente globale di clustering	0.365	0.365
Coefficiente di assortatività	-0.186	-0.186
Numero di triangoli	859	859
Diametro	NA	6
Lunghezza media del percorso più breve	NA	2.664

Abbiamo esaminato le metriche di degree, betweenness e Closeness Centrality sul grafo connesso, al fine di confrontare i valori ottenuti con quelli del grafo non connesso. Dall'analisi, abbiamo notato che i cambiamenti nelle metriche sono minimi e generalmente indicano variazioni leggere.

Questo risultato può essere spiegato considerando che abbiamo rimosso un singolo nodo che non aveva connessioni con gli altri nodi. La sua rimozione ha avuto un impatto limitato sulle misure di centralità. In altre parole, l'assenza di connessioni di quel nodo ha avuto un effetto trascurabile sulle metriche globali.

Ciò significa che i nodi rimanenti mantengono le loro posizioni relative all'interno delle misure di centralità. I nodi che avevano un alto grado di centralità prima della rimozione del nodo isolato rimangono generalmente alti dopo la rimozione, mentre i nodi con una centralità più bassa continuano a mantenere una posizione inferiore.

A. Degree Centrality

TABLE XIII
TOP 8 DEGREE CENTRALITY

Node	Connesso	Disconnesso
ANAKIN	0.382	0.378
OBI-WAN	0.336	0.333
C-3PO	0.327	0.324
PADME	0.309	0.306
QUI-GON	0.245	0.243
LUKE	0.245	0.243
HAN	0.236	0.234
LEIA	0.227	0.225

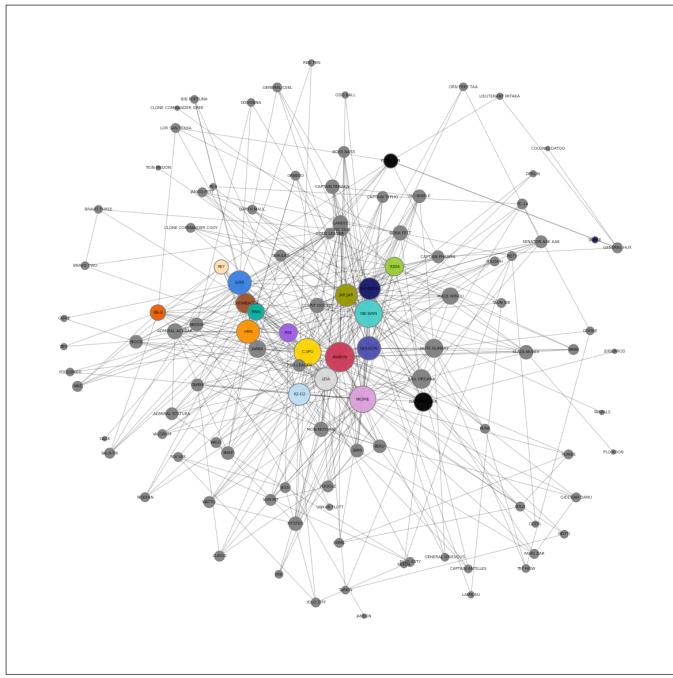


Fig. 16. Degree Centrality Plot.

B. Betweenness Centrality

TABLE XIV
TOP 8 BETWEENNESS CENTRALITY

Node	Connesso	Disconnesso
OBI-WAN	0.209	0.205
C-3PO	0.171	0.168
ANAKIN	0.161	0.158
LUKE	0.144	0.142
HAN	0.098	0.096
DARTH VADER	0.087	0.085
EMPEROR	0.076	0.075
POE	0.074	0.073

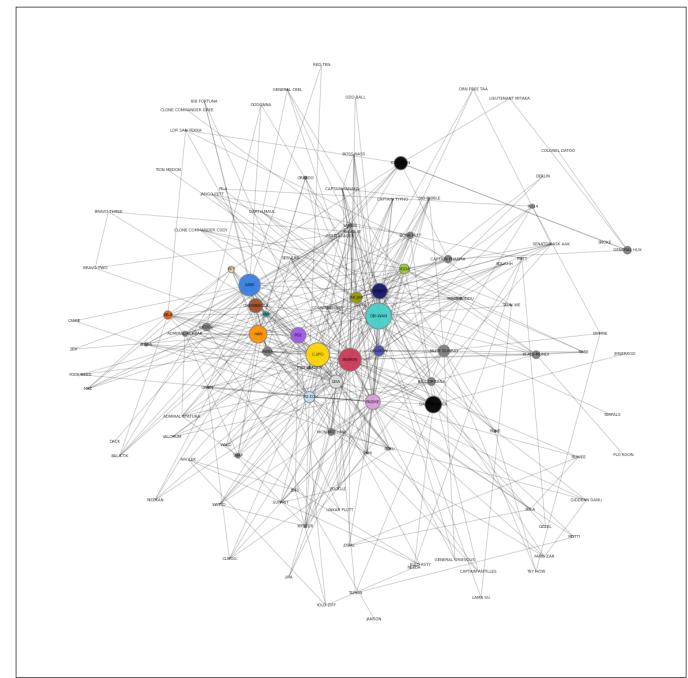


Fig. 17. Betweenness Centrality Plot.

C. Closeness Centrality

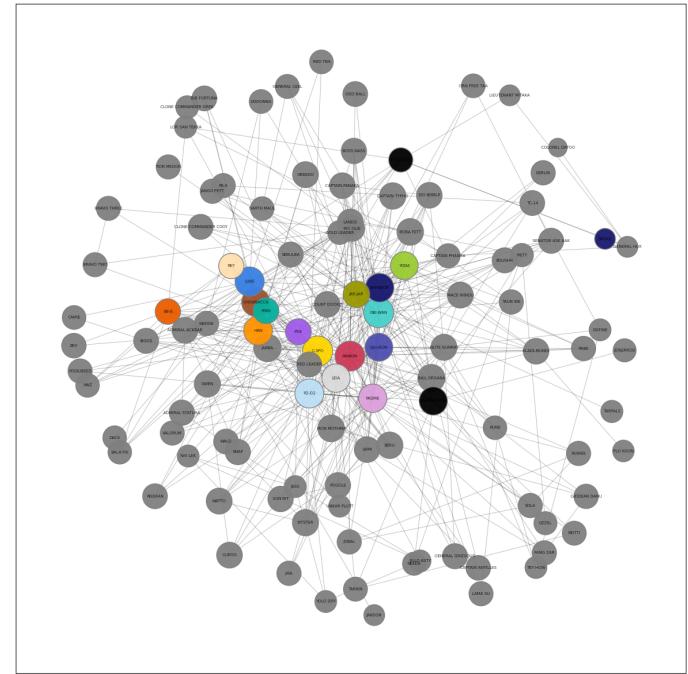


Fig. 18. Closeness Centrality Plot.

TABLE XV
TOP 8 CLOSENESS CENTRALITY

Node	Connesso	Disconnesso
C-3PO	0.567	0.562
OBI-WAN	0.564	0.559
ANAKIN	0.556	0.551
LUKE	0.531	0.527
R2-D2	0.521	0.517
HAN	0.516	0.512
EMPEROR	0.507	0.502
LEIA	0.498	0.493

In conclusione, i cambiamenti minimi nelle misure di centralità dopo la rimozione di un singolo nodo non connesso indicano che il ruolo di quel nodo era marginale nell'influenzare le metriche di centralità complessive del grafo connesso.

D. Communities

Abbiamo eseguito un'ulteriore analisi delle comunità all'interno del grafo utilizzando sempre la funzione "best partition" basata sull'algoritmo di Louvain. I risultati hanno evidenziato la presenza di sette comunità distintive.

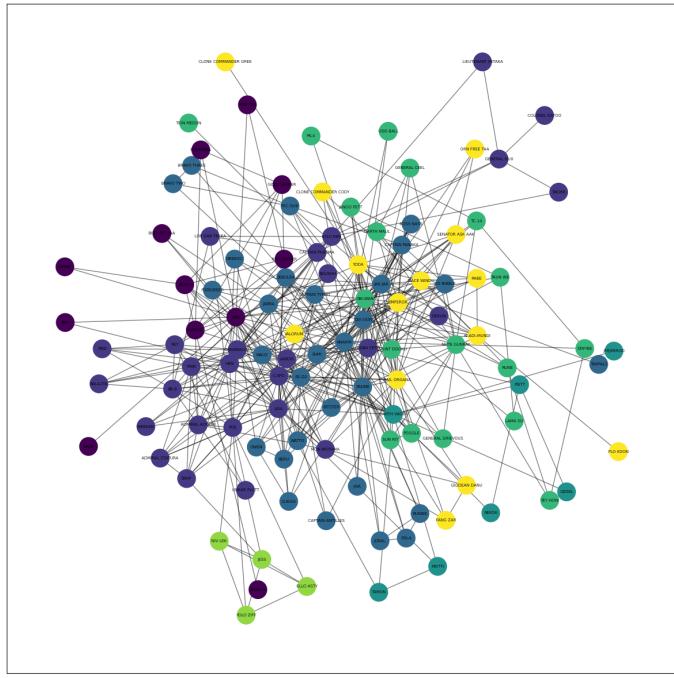


Fig. 19. Best Partition Plot.

TABLE XVI
BEST PARTITION COMMUNITIES

Communities		
Community 0	Community 1	Community 2
LUKE CAMEL BIGGS DODONNA GOLD LEADER WEDGE RED LEADER RED TEN ZEV DACK JANSON BIB FORTUNA	CHEWBACCA BB-8 C-3PO BOBA FETT MON MOTHMA LEIA HAN RIEKEKAN DERLIN LANDO BOUSHH ADMIRAL ACKBAR LOR SAN TEKKA POE KYLO REN CAPTAIN PHASMA FINN UNKAR PLUTT REY GENERAL HUX LIEUTENANT MITAKA BALA-TIK SNOKE MAZ SNAP ADMIRAL STURIA COLONEL DATOO	R2-D2 QUI-GON CAPTAIN PANAKA SIO BIBBLE JAR JAR TARPALS BOSS NASS PADME RIC OLINE WATTO ANAKIN SEBULBA JIRA SHMI KITSTER WALD FODE/BEED JABBA GREEDO BRAVO TWO BRAVO THREE CAPTAIN TYPHO SOLA JOBAL RUWEE OWEN BERU CLIEGG CAPTAIN ANTILLES
Community 3	Community 4	Community 5
DARTH VADER MOTTI TARKIN PIETT OZZEL NEEDA JERJERROD	NUTE GUNRAY PK-4 TC-14 OBI-WAN DOFINE RUNE TEY HOW DARTH MAUL GENERAL CEL TAUN WE LAMA SU JANGO FETT COUNT DOOKU SUN RIT POGGLE ODD BALL GENERAL GRIEVOUS TION MEDON	YOLO ZIFF ELLO ASTY JESS NIV LEK
Community 6		
EMPEROR VALORUM MACE WINDU KI-ADI-MUNDI YODA RABE BAIL ORGANA SENATOR ASK AAK ORN FREE TAA PLO KOON FANG ZAR GIDDEAN DANU CLONE COMMANDER GREE CLONE COMMANDER CODY		

La prima comunità rappresenta i personaggi legati a LUKE, inclusi i piloti ribelli e BIB FORTUNA, un personaggio correlato nel sesto film.

La seconda comunità è costituita dai personaggi del settimo film, "The Force Awakens", equivalente alla quinta community precedente.

La terza comunità comprende i personaggi del primo film, "Phantom Menace", come nel grafo non connesso.

La quarta comunità rappresenta i personaggi del primo film, in particolare quelli legati a DARTH VADER e che collaborano con lui sulla Morte Nera.

La quinta comunità coincide con la quarta nel grafo non connesso e comprende i personaggi dei prequel, tra cui OBI-WAN e quelli ad esso correlati.

La sesta comunità è formata da quattro piloti che fanno parte della ribellione.

Infine, la settima e ultima comunità coincide con la sesta nel grafo non connesso e include personaggi importanti con ruoli di leadership e autorità politica, tra cui i Jedi.

Le communities ottenute con l'algoritmo Girvan Newman rimangono identiche a quelle del grafo non connesso, ad eccezione della community che includeva il singolo nodo GOLD FIVE, che non è presente.

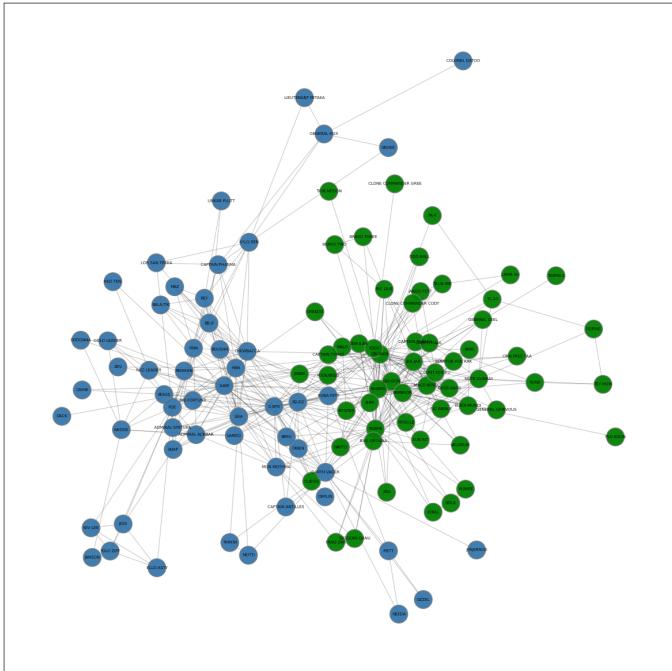


Fig. 20. Girvan Newman Plot.

TABLE XVII
GIRVAN NEWMAN COMMUNITIES

Communities	
Community 0	Community 1
LOR SAN TEKKA	BRAVO TWO
REY	TION MEDON
LANDO	DOFINE
BERU	ANAKIN
LEIA	JANGO FETT
BOUSHH	GIDDEAN DANU
NIV LEK	ODD BALL
ELLO ASTY	WALD
BALA-TIK	TARPALS
RIEEKAN	BRAVO THREE
DARTH VADER	TAUN WE
CAPTAIN ANTILLES	CAPTAIN PANAKA
COLONEL DATOO	PADME
RED TEN	OBI-WAN
YOLO ZIFF	JABBA
R2-D2	TEY HOW
ADMIRAL STATURA	KITSTER
ADMIRAL ACKBAR	SOLA
LUKE	BOSS NASS
MON MOTHMA	PK-4
PIETT	SENATOR ASK AAK
WEDGE	NUTE GUNRAY
DODONNA	TC-14
HAN	POGGLE
BIGGS	CLOSE COMMANDER CODY
NEEDA	SHMI
CHEWBACCA	RUNE
RED LEADER	JOBAL
OZZEL	JIRA
JESS	GENERAL GRIEVOUS
GOLD LEADER	FANG ZAR
SNOKE	MACE WINDU
LIEUTENANT MITAKA	RUWEE
MAZ	SUN RIT
BB-8	QUI-GON
MOTTI	YODA
JANSON	CAPTAIN TYPHO
BOBA FETT	BAIL ORGANA
DERLIN	RIC OLIE
CAMIE	CLIEGG
JERJERROD	DARTH MAUL
TARKIN	ORN FREE TAA
OWEN	SIO BIBBLE
UNKAR PLUTT	SEBULBA
BIB FORTUNA	VALORUM
POE	LAMA SU
CAPTAIN PHASMA	KI-ADI-MUNDI
KYLO REN	COUNT DOOKU
GENERAL HUX	RABE
FINN	CLOSE COMMANDER GREE
ZEV	GREEDO
C-3PO	WATTO
DACK	FODE/BEED
SNAP	EMPEROR

E. Grafo Interattivo

Abbiamo poi di nuovo optato per l'impiego della libreria Pyvis per creare un grafico interattivo, questa volta però dopo aver eliminato dal nostro grafo il nodo isolato.

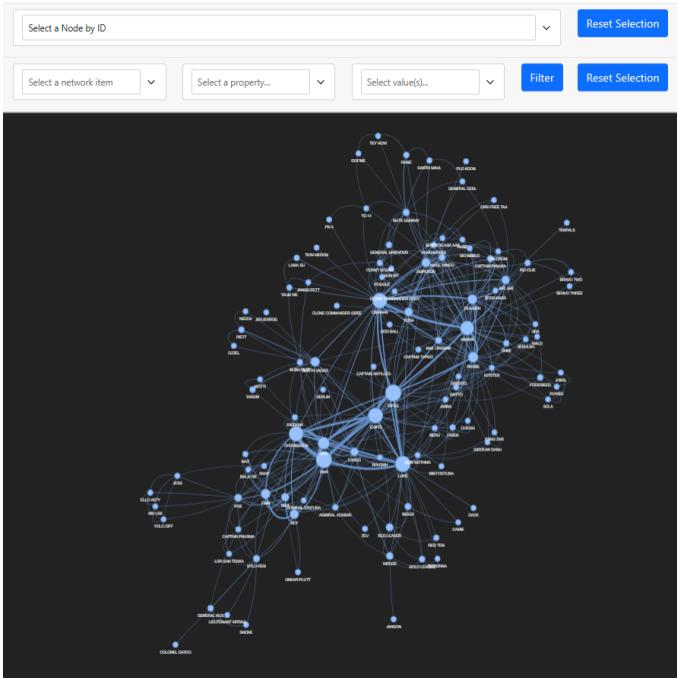


Fig. 21. Interactive Plot.

V. STREAMLIT

Utilizzando Streamlit, abbiamo progettato una dashboard interattiva per lo studio dei dataset sulle reti sociali di Star Wars. La dashboard offre strumenti avanzati e visualizzazioni interattive per esplorare e analizzare le interazioni tra i personaggi.

Oltre alle visualizzazioni, mettiamo a disposizione metriche chiave come la Degree Centrality, la Closeness Centrality e la Betweenness Centrality per valutare l'importanza e l'influenza dei personaggi all'interno della rete.

Inoltre, forniamo informazioni generali sui grafi, inclusi il numero totale di nodi e archi, la densità del grafo e la distribuzione dei gradi dei personaggi. Questi dati consentono agli utenti di ottenere una comprensione approfondita delle reti sociali di Star Wars e di esplorare le dinamiche dei personaggi in modo interattivo.

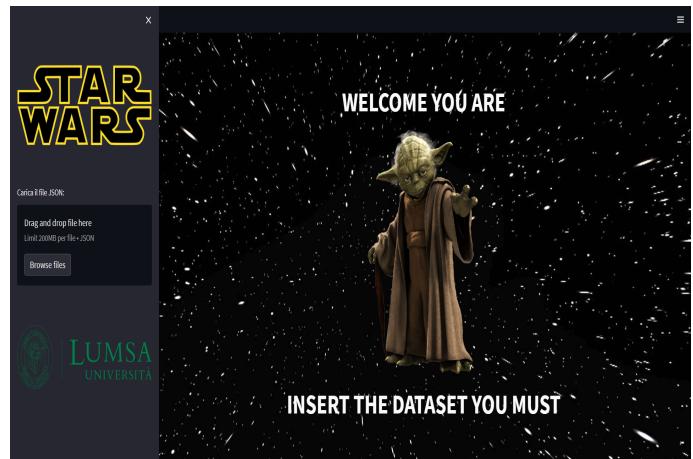


Fig. 22. Streamlit App Intro.

Per importare i dataset, è necessario utilizzare il box appositamente designato all'interno della barra laterale. È possibile importare il file "starwars-episode-N-interactions.json", dove "N" indica il numero del film per cui si desidera analizzare la rete sociale.

Il nostro studio è stato condotto sul grafo che include tutti i personaggi, ma è possibile visualizzare le stesse metriche e grafi interattivi selezionando il film di interesse.

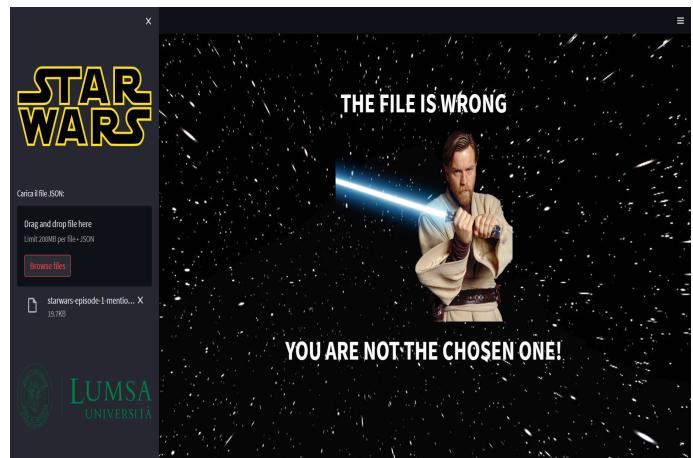


Fig. 23. Streamlit App Error.

Se il file JSON selezionato non è tra quelli supportati per l'analisi, verrà visualizzata una schermata di errore.

VI. CONCLUSIONS

In conclusione, l'analisi condotta all'interno di questa relazione ha fornito una visione approfondita della rete dei personaggi nell'universo di Star Wars.

Nella prima parte dell'analisi, abbiamo esaminato il grafo disconnesso, che ci ha permesso di identificare i personaggi principali e le loro interazioni all'interno della rete. Successivamente, nella seconda parte dell'analisi, abbiamo rimosso il nodo isolato, che ha reso il grafo connesso.

Nonostante questa differenza nella struttura del grafo, i risultati delle metriche di centralità hanno mostrato solo lievi variazioni.

Gli ordini di importanza dei personaggi sono rimasti costanti tra il grafo disconnesso e quello connesso. I risultati ottenuti dalle misure di centralità hanno contribuito a identificare i personaggi più importanti nella rete.

TABLE XVIII
BEST CHARACTERS

Metriche	Nodo
Degree Centrality	ANAKIN
Betweenness Centrality	OBI-WAN
Closeness Centrality	C-3PO
PageRank Centrality	ANAKIN
Eigenvector Centrality	ANAKIN

ANAKIN SKYWALKER è emerso come il personaggio più significativo in base alla Degree Centrality, alla PageRank Centrality e all'Eigenvector Centrality, indicando un elevato numero di connessioni e un ruolo cruciale nel collegare gli altri personaggi.

OBI-WAN KENOBI ha ottenuto una rilevante Betweenness Centrality, suggerendo che agisce come principale mediatore tra i personaggi.

Infine, C-3PO si è distinto per la Closeness Centrality, implicando una forte accessibilità ad altri personaggi.

Questi tre personaggi si distinguono costantemente e compaiono sempre tra i primi tre posti in tutte le metriche analizzate.

Tuttavia, è importante notare che le community all'interno della rete possono variare leggermente a seconda dell'algoritmo utilizzato per la loro identificazione.

In particolare, utilizzando Best Partition, le comunità potrebbero essere leggermente diverse tra il grafo connesso e quello disconnesso. Ciò può dipendere dalle specificità dell'algoritmo e dalla sua interpretazione delle connessioni e delle interazioni tra i personaggi.

Nonostante queste differenze, entrambe le parti dell'analisi hanno contribuito a una migliore comprensione della struttura e delle dinamiche della rete dei personaggi di Star Wars.