

ALMA MATER STUDIORUM · UNIVERSITÀ DI
BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea Magistrale in Matematica

Otello fra amore, morte e virtù:
un'analisi dei campi semantici
dell'opera tramite la teoria dei grafi

Tesi di Laurea in teoria dei grafi

Relatore:
Chiar.mo Prof.
Marco Lenci

Correlatore:
Chiar.mo Prof.
Rocco Coronato

Presentata da:
Andrea Magnavacchi

II Sessione
Anno Accademico 2017/2018

Introduzione

Negli ultimi anni, l'ampia disponibilità di opere letterarie digitalizzate ha dato origine a un approccio computazionale per l'analisi di questi testi. Questo approccio è stato usato, a volte, in combinazione con tecniche di analisi letteraria più tradizionali, per cogliere meglio le complessità di diverse opere letterarie. Con il maturare del campo, nuovi approcci e idee hanno dato origine all'uso di tecniche, come le reti sociali, solitamente riservate ai campi quantitativi per acquisire nuove conoscenze sulle opere. Vari esempi sono citati nel capitolo 4. In tutti questi esempi, le tecniche estratte dalla teoria dei grafi e applicate alle opere letterarie vengono utilizzate per trovare risultati quantitativi sui personaggi dell'opera, cioè chi è più importante, quali sono i gruppi o le comunità nel testo, ma nessuno di questi affronta il tema degli argomenti trattati. Questo è ciò che contraddistingue il nostro lavoro dai precedenti. Le tecniche che useremo sono già state usate, ma è innovativo lo spazio di ricerca. Creare delle reti di argomenti non è naturale come creare reti di personaggi, reti sociali, ma i risultati raggiunti sono incoraggianti, confermando quanto siano duttili gli strumenti che derivano dalla teoria dei grafi e dalla matematica in generale.

Questa tesi è suddivisa in due parti principali. La prima parte comprende i primi tre capitoli ed è strettamente matematica, mentre la seconda verte sui risultati ottenuti applicando le tecniche, in particolare le misure di centralità, ai grafi che abbiamo estrapolato dal testo *The Tragedy of Othello, Moor of Venice* [22]. In particolare, nella parte teorica abbiamo dato alcune nozioni fondamentali riguardanti la teoria dei grafi, per poi concentrarci sull'aspetto

spettrale di questa teoria, analizzando soprattutto lo spettro del Laplaciano, per semplicità, relativo a grafi non orientati e non pesati. Per affrontare in maniera completa questa recente teoria si possono consultare [11] o [9]. Dopodichè ci siamo concentrati sulle misure di centralità, grazie alle quali si riescono ad estrapolare dati quantitativi dai grafi.

Nella seconda parte abbiamo effettuato un breve excursus sui lavori precedenti, evidenziandone le differenze, per poi concentrarci su ciò che abbiamo fatto noi. Grazie al professore Rocco Coronato dell'Università degli Studi di Padova, che ci ha fornito una lista di *topics*, o campi semantici, presenti nel testo di Othello, ognuno con una lista di parole di riferimento, siamo riusciti a costruire dei grafi, prima indiretti e poi diretti, sui quali abbiamo applicato le varie misure di centralità. In particolare per ogni campo semantico è stato generato un grafo indiretto, con i personaggi rappresentati dai nodi collegati se nelle loro interazioni parlavano del *topic* in questione. Successivamente abbiamo analizzato il testo anche temporalmente, effettuando operazioni del tutto simili alla precedente. In tutto ciò i grafi sono sempre stati non orientati, mentre per l'ultima parte di lavoro abbiamo generato dei grafi diretti, dove i nodi rappresentavano i *topics*, e venivano collegati fra loro con uno spigolo orientato dal *topic* i verso il *topic* j se nel testo la menzione di j seguiva quella di i , all'interno di un limite di caratteri. Tutto questo lo abbiamo fatto temporalmente e, calcolando le solite misure di centralità, abbiamo classificato i vari campi semantici, cercando di capire quali fossero quelli predominanti nel testo. Per finire abbiamo cambiato alcuni parametri, come la quantità di parole chiave legati ad ogni *topic*, o il limite caratteriale in cui considerare legati i *topics* che si susseguono nel testo, per verificare quanto fossero stabili ed indipendenti dai parametri scelti i risultati ottenuti. Tutti i dettagli di questo lavoro si trovano nel capitolo 5. Non abbiamo inserito tutti i grafi, tabelle e grafici ottenuti all'interno del capitolo, ma solo i più significati a titolo d'esempio. I restanti li abbiamo inseriti nell'appendice.

La ricerca testuale delle parole per la costruzione dei grafi è stata fatta tramite Python, mentre per l'analisi quantitativa, cioè calcolo delle misure

di centralità, classifiche e grafici dei risultati, è stato utilizzato Matlab.

Indice

Introduzione	vi
1 Cenni di teoria dei grafi	1
1.1 Introduzione alla teoria dei grafi	1
1.2 Teorema di Perron-Forbenius	3
2 Teoria spettrale dei grafi	9
2.1 Matrici associate a un grafo	9
2.2 Teoria spettrale del Laplaciano	10
2.3 Disuguaglianza di Cheeger	15
3 Misure di centralità	23
3.1 Closeness Centrality	23
3.2 Betweenness Centrality	24
3.3 Eingenvektor Centrality	25
3.4 PageRank Centrality	26
3.5 Hubs e Authority Centrality	28
4 Misure di centralità applicate alla narrativa	31
5 L’Otello di Shakespeare	37
5.1 Personaggi importanti per campi semantici	38
5.2 Centralità dei <i>topics</i>	47
Conclusioni	67

A	71
A.1 Grafi personaggi per campo semantico	71
A.2 Classifiche personaggi per campo semantico	85
A.3 Grafici personaggi per campo semantico	99
A.4 Tabelle campi semantici	109
A.5 Grafici campi semantici	111
Bibliografia	113

Capitolo 1

Cenni di teoria dei grafi

1.1 Introduzione alla teoria dei grafi

Definizione 1.1. Un Grafo G è una coppia ordinata $G = (V, E)$, dove V è un insieme di $|V| = n$ elementi chiamati *nodi*, mentre $E = \{(i, j) \in V \times V : i \text{ è connesso a } j\}$ è l'insieme degli *spigoli* o *archi*.

Definizione 1.2. Se gli elementi di E sono coppie ordinate, il grafo si dice *diretto* o orientato, altrimenti si dice *indiretto* o non orientato. Quindi un grafo indiretto non distingue (i, j) da (j, i) , ma essi vengono identificati.

Il grafo può essere rappresentato da una matrice $n \times n$ $A = (a_{ij})$ chiamata *matrice di adiacenza* e definita come segue: $a_{ij} = \begin{cases} 1 & \text{se } (i, j) \in E \\ 0 & \text{altrimenti} \end{cases}$

Osservazione 1. G è un grafo indiretto $\iff A$ simmetrica

Definizione 1.3. Una tupla $(u_1, \dots, u_k) \in V^k$ t.c. $u_l \neq u_h \quad \forall l, h \in \{1, \dots, k\}$ e t.c. si ha uno spigolo fra u_l e $u_{l+1} \quad \forall l = 1, \dots, k-1$ si chiama *cammino* (se il grafo fosse diretto, lo spigolo deve essere diretto). k è la lunghezza del cammino. Denotiamo con P_k l'insieme dei cammini lunghi k nel caso di grafo indiretto. Nel caso diretto usiamo P_k^+ .

Diciamo che un grafo G indiretto è *connesso* se esiste un cammino che attraversa ogni nodo.

Un grafo G diretto è *fortemente connesso* se esiste un cammino diretto che attraversa ogni nodo.

Un grafo si dice *semplice* se $(i, i) \notin E, \quad \forall i \in V$

Definizione 1.4. Un grafo *pesato* è $G = (V, E, \omega_G)$ con

$$\omega_g : E \rightarrow \mathbb{R}$$

$$(i, j) \rightarrow \omega_{ij}$$

$\omega_G(E)$ può essere visto come l'insieme dei pesi associati ad ogni spigolo.

La matrice di adiacenza in questo caso può essere definita come $W = (w_{ij})$.

Osservazione 2. Un grafo non pesato è un grafo pesato con pesi unitari.

Definiamo ora una grandezza fondamentale nella teoria dei grafi:

Definizione 1.5. Sia G un grafo indiretto con n nodi. Preso un nodo $i \in G$ definiamo il *grado* di i come

$$d_i = \sum_{j=1}^n a_{ij}. \quad (1.1)$$

Osservazione 3. Sia $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^n$. Possiamo definire il vettore dei gradi dei nodi come $d = A\mathbf{1}$.

Questo modo di definire il grado è utile in quanto necessita unicamente della matrice di adiacenza.

Se il grafo è diretto non possiamo considerare solo il grado precedente, ma dobbiamo tenere conto che gli spigoli hanno un verso. Definiamo quindi l'out-degree e l'in-degree:

- $d_i^{out} = [A\mathbf{1}]_i$;
- $d_i^{in} = [A^T\mathbf{1}]_i$.

Osservazione 4. Se il grafo fosse indiretto, avremmo $d_i^{out} = d_i^{in} = d_i$.

Notiamo che queste definizioni funzionano sia nel caso di grafi pesati che non. Nel caso di grafi non pesati, il grado ci dice quanti 'vicini' entrano in un nodo o 'escono' dal nodo, oppure, nel caso indiretto, quanti sono i 'vicini' del nodo i .

1.2 Teorema di Perron-Forbenius

Riportiamo un importante risultato per matrici positive, poi esteso per matrici non negative irriducibili, che verrà utilizzato per le misure di centralità. Scriveremo $A \geq 0$ se $a_{ij} \geq 0 \quad \forall i, j = 1, \dots, n$ e $A > 0$ se la disuguaglianza è stretta.

Teorema 1.2.1. *Sia A una matrice $n \times n$. Se $A > 0$ con $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ l'insieme degli autovalori di A tali che $\rho(A) = |\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$, allora le seguenti affermazioni sono vere:*

1. $\lambda_1 > 0$ (radice di Perron) e $\lambda_1 > |\lambda_i|$ per $i = 2, \dots, n$;
2. $\exists v > 0$ e $w > 0$ vettori tali che $Av = \lambda_1 v$ e $w^T A = \lambda_1 w^T$ (vettori di Perron)
3. λ_1 ha molteplicità algebrica 1
4. Sia v_i (w_i) un autovettore destro (sinistro) associato a λ_i . Se $v_i > 0$ ($w_i > 0$) allora $i = 1$.
5. $\lim_{k \rightarrow \infty} \frac{A^k}{\lambda_1^k} = v w^T =: \mu$ (proiezione di Perron) con v e w normalizzati in modo che $w^T v = 1$. La velocità di convergenza è dell'ordine di $(\frac{\lambda_2}{\lambda_1})^k$;
6. $\lambda_1 \leq \max_{i=1, \dots, n} \sum_j a_{ij} =: \|A\|_\infty$

Dimostrazione. Notiamo innanzitutto che se $\rho(A) = 0$, allora A è nilpotente per qualche $n \in \mathbb{N}$, visto che $\|A^n\|_\infty = 0$ e ciò contraddice la positività di A , quindi $\rho(A) > 0$. Sia $PJP^{-1} = A$ la decomposizione di Jordan di A . Si mostra facilmente che $PJ^kP^{-1} = A^k$. Denotiamo con J_{λ_i} il blocco di Jordan $m_i \times m_i$ associato all'autovalore λ_i e $N_i = (n_{js})$ la matrice $m_i \times m_i$ nilpotente tale che:

$$n_{js} = \begin{cases} 1 & \text{se } s = j + 1 \\ 0 & \text{altrimenti} \end{cases}$$

Allora $N^{m_i} = 0$ e $N, \lambda_i I$ commutano; ora applichiamo il teorema binomiale:

$$J_{\lambda_i}^k = (\lambda_i I + N)^k = \sum_{l=0}^k \binom{k}{l} \lambda_i^{k-l} N^l = \sum_{l=1}^{\min(k, m_i-1)} \binom{k}{l} \lambda_i^{k-l} N^l$$

Se $\rho(A) < 1$ allora

$$\lim_{k \rightarrow \infty} \sum_{l=1}^{\min(k, m_i-1)} \binom{k}{l} \frac{\lambda_i^{k-l}}{\lambda_1^k} = 0 \quad \forall i = 1, \dots, n.$$

Quindi $\exists \lim_{k \rightarrow \infty} J^k = 0 \Rightarrow \exists \lim_{k \rightarrow \infty} P J^k P^{-1} = \lim_{k \rightarrow \infty} A^k = 0$. Questo lo useremo dopo, mentre adesso assumiamo $\rho(A) = 1$. Se $Av = \lambda_1 v$ e con la notazione $|v| = (|v_i|)_{i=1, \dots, n}$ abbiamo:

$$\|v\| = |\lambda_1| \|v\| = \|\lambda_1 v\| = \|Av\| \leq \|A\|_{\infty} \|v\| \Rightarrow |\lambda_1| \leq \|A\|_{\infty}$$

e

$$|v| = |\lambda_1| |v| = |\lambda_1 v| = |Av| \leq |A| |v| \Rightarrow y := |A| |v| - |v| \geq 0.$$

Supponiamo per assurdo che

$$y > 0 \Rightarrow Ay > 0 \Rightarrow z := A|v| > 0 \Rightarrow \exists \epsilon > 0 :$$

$$Ay = A(|A||v| - |v|) = |A|z - z > \epsilon z \Rightarrow \frac{A}{1+\epsilon} z > z$$

Sia $B = \frac{A}{1+\epsilon}$ e moltiplicando per B la disuguaglianza otteniamo la catena di disuguaglianze

$$B^k z > B^{k-1} z > \dots > Bz > z \quad \forall k \in \mathbb{N}.$$

Abbiamo che $\rho(B) = \rho\left(\frac{A}{1+\epsilon}\right) = \frac{1}{1+\epsilon} \rho(A) = \frac{1}{1+\epsilon} < 1$, quindi $\lim_{k \rightarrow \infty} B^k = 0$ da cui $z < \lim_{k \rightarrow \infty} B^k z = 0$, ma questo è assurdo $\Rightarrow y = 0$, così $|v|$ è un autovettore di A relativo all'autovalore $1 = \rho(A) \Rightarrow 1 = \lambda_1 > 0$.

Se $\rho(A) \neq 1$ allora consideriamo la matrice $\frac{1}{|\lambda_1|} A$. Abbiamo:

$$\rho\left(\frac{1}{|\lambda_1|} A\right) = 1 \Rightarrow \exists v > 0 : \frac{A}{|\lambda_1|} v = v \Rightarrow Av = |\lambda_1| v$$

quindi v è un autovettore di A relativo a $|\lambda_1|$ e come prima $\lambda_1 > 0$.

Ora mostriamo che $\forall i = 2, \dots, n$ tale che $\lambda_1 \neq |\lambda_i|$ si ha $\lambda_1 > |\lambda_i|$ e che la molteplicità algebrica di λ_1 è 1. Come prima possiamo supporre che $\lambda_1 = 1$. In ordine dimostreremo che:

1. 1 è l'unico autovalore con valore assoluto uguale a 1;
2. Il suo indice (cioè la dimensione del suo blocco di Jordan nella decomposizione di Jordan di A) è 1, facendo vedere così che la sua molteplicità algebrica è 1;
3. La sua molteplicità geometrica è 1.

Supponiamo che $\exists i$ per cui $|\lambda_i| = 1$. Come prima, se $Ax = \lambda_i x$ allora $|A||x| = |x| > 0$. In particolare:

$$\sum_{j=1, \dots, n} a_{kj} |x_j| = \sum_{j=1, \dots, n} |a_{kj} x_j| = |x_k| = |\lambda_i| |x_k| = |\lambda_i x_k| = |(Ax)_k| = \left| \sum_{j=1, \dots, n} a_{kj} x_j \right|.$$

Quindi la disuguaglianza triangolare in questo caso è un'uguaglianza ed inoltre vale che preso z vettore non nullo n -dim $\left\| \sum_j z_j \right\|_2 = \sum_j \|z_j\|_2$ se e solo se $\forall j = 1, \dots, n$ $z_j = \alpha_j z_1$ per qualche $\alpha_j > 0$. Quindi $\exists \alpha_j > 0 : a_{kj} x_j = \alpha_j (a_{k1} x_1) \quad \forall j = 2, \dots, n$

$$\Rightarrow x = x_1 \left(1, \frac{\alpha_2 a_{k1}}{a_{k2}}, \dots, \frac{\alpha_n a_{k1}}{a_{kn}} \right)^T =: x_1 p > 0$$

Allora $A(x_1 p) = \lambda_i x_1 p \Rightarrow \lambda_i p = Ap = |Ap| = |\lambda_i p| = p \Rightarrow \lambda_i = 1$.

Per assurdo assumiamo che l'indice di 1 sia $m > 1$. Per la decomposizione di Jordan $P^{-1}AP = J$ e per ipotesi esiste un blocco di Jordan J_1 $m \times m$ associato all'autovalore 1. Allora $\lim_{k \rightarrow \infty} \|J_1^k\|_\infty = \infty \Rightarrow \lim_{k \rightarrow \infty} \|J^k\|_\infty = \infty$.

ora

$$\begin{aligned} \|J^k\|_\infty &= \|(P^{-1}Ap)^k\|_\infty \leq \|P^{-1}\|_\infty \|A^k\|_\infty \|P\|_\infty \\ \Rightarrow \|A^k\|_\infty &\geq \frac{\|J^k\|_\infty}{\|P^{-1}\|_\infty \|P\|_\infty} \Rightarrow \|A^k\|_\infty \rightarrow \infty \end{aligned}$$

Sia i_k l'indice che denota la colonna per cui $\|A^k\|_\infty = \sum_j a_{i_k j}^{(k)}$. Sappiamo che $\exists p = (p_i) > 0 : Ap = p$, quindi:

$$\|p\|_\infty \geq p_{i_k} = \sum_{j=1, \dots, n} a_{i_k j}^{(k)} p_j \geq \sum_{j=1, \dots, n} a_{i_k j}^{(k)} \min_{i=1, \dots, n} p_i = \|A^k\|_\infty \min_{i=1, \dots, n} p_i \rightarrow \infty$$

con $k \rightarrow \infty \Rightarrow \lim_{k \rightarrow \infty} \|p\|_\infty = \infty$, ma questo è assurdo poichè p non dipende da $k \Rightarrow m = 1$.

Visto che $m_{geo}(1)$ e $m_{alg}(1)$ sono rispettivamente il numero di blocchi di Jordan associati a 1 e la somma delle loro dimensioni, abbiamo provato che sono uguali.

Se $m := m_{geo}(1) = \dim Ker(A - \lambda_1 I) = m_{alg}(1) > 1 \Rightarrow \exists v_1, \dots, v_m$ vettori positivi e linearmente indipendenti : $Av_i = v_i \quad \forall i = 1, \dots, m$. In particolare, $v_i \neq \alpha v_j \Rightarrow \exists l \in \{ \dots, n \} : v_{il} \neq v_{jl} \Rightarrow z := v_i - \frac{v_{il}}{v_{jl}} v_j$ soddisfa $Az = z \Rightarrow A|z| = |z| > 0$ ma questo è impossibile poichè $z_l = 0 \Rightarrow m = 1 = \dim Ker(A - I) \Rightarrow \exists v > 0$ autovettore di 1 che è proprio il vettore di Perron.

Tutto la dimostrazione può essere ripetuta per trovare il vettore sinistro di Perron $w > 0$ sostituendo ad A con A^T .

Se l'autovettore v_i associato a λ_i è positivo e w è il vettore sinistro di Perron, allora $w^T v_i > 0$ e visto che $w^T A = w^T$, si ha $w^T v_i = w^T A v_i = w^T \lambda_i v_i \Rightarrow \lambda_i = 1$.

Ora studiamo il limite di $\frac{1}{\lambda_1^k} A^k$. Come prima possiamo decomporre A e studiare i blocchi di Jordan $(\frac{1}{\lambda_1} J_{\lambda_i})^k \quad \forall i = 1, \dots, n$.

Se $i \neq 1$ abbiamo, esattamente come nella prima parte della dimostrazione,

$$\left(\frac{J_{\lambda_i}}{\lambda_1} \right)^k = \sum_{l=1}^{\min(k, m_i - 1)} \binom{k}{l} \frac{\lambda_i^{k-l}}{\lambda_1^k} N^l \quad (1.2)$$

Dato che $\lambda_1 > \lambda_i$ per $i \neq 1$, si ha $\lim_{k \rightarrow \infty} \binom{k}{l} \frac{\lambda_i^{k-l}}{\lambda_1^k} = 0 \Rightarrow \lim_{k \rightarrow \infty} \left(\frac{J_{\lambda_i}}{\lambda_1} \right)^k = 0$.

Se $i = 1$ allora $J_{\lambda_1}^k = \lambda_1^k \Rightarrow \left(\frac{J_{\lambda_1}}{\lambda_1} \right)^k = 1$. Quindi abbiamo che:

$$\exists \lim_{k \rightarrow \infty} \left(\frac{J_{\lambda_1}}{\lambda_1} \right)^k = (1, 0, \dots, 0)^T (1, 0, \dots, 0) \Rightarrow \lim_{k \rightarrow \infty} P \left(\frac{J_{\lambda_1}}{\lambda_1} \right)^k P^{-1} = \lim_{k \rightarrow \infty} \frac{A^k}{\lambda_1^k},$$

con velocità di convergenza dell'ordine di $(\frac{\lambda_2}{\lambda_1})^k$ come si osserva dalla 1.2.

Ora normalizzando i vettori di Perron v e w^T in modo tale che $w^T v = 1$. Con abuso di notazioni richiamiamo i vettori normalizzati v e w^T . Possiamo scrivere la decomposizione di Jordan di A come $PAP^{-1} = J$, dove la prima colonna di P è v e la prima riga di P^{-1} è w^T . Quindi

$$P^{-1} \frac{J^k}{\lambda_1^k} P = \frac{A^k}{\lambda_1^k} \Rightarrow \lim_{k \rightarrow \infty} \frac{A^k}{\lambda_1^k} = \lim_{k \rightarrow \infty} P^{-1} \frac{J^k}{\lambda_1^k} P.$$

Osserviamo che facendo il limite in $\frac{1}{\lambda_1^k} J^k$ sopravvive solo $J_{11}^{(k)} = 1$ mentre il resto tende a 0, quindi con un calcolo diretto otteniamo

$$\lim_{k \rightarrow \infty} P^{-1} \frac{J^k}{\lambda_1^k} P = v w^T.$$

□

Definizione 1.6. La matrice $n \times n$ A si dice *irriducibile* se e soltanto se sono verificate le seguenti affermazioni:

1. A non è simile a una matrice triangolare superiore, cioè non esiste una matrice di permutazione P tale che

$$P^{-1} A P = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix}$$

con X e Z matrici quadrate;

2. $(I + A)^{n-1} > 0$;
3. $\forall i, j = 1, \dots, n \quad \exists m \in \mathbb{N}$ tale che $(A^m)_{ij} \neq 0$

Definizione 1.7. Sia A una matrice $n \times n$, $A \geq 0$. Definiamo il *periodo* dello stato i come $MCD\{m \in \mathbb{N} : (A^m)_{ii} > 0\}$. Se A è irriducibile, allora tutti gli stati hanno lo stesso periodo. Se il periodo è 1 allora si dice che A è *aperiodica*, altrimenti *periodica*.

Osservazione 5. Se A è la matrice di adiacenza di un grafo G , allora $A \geq 0$ e $(A^k)_{is} = \#\{w \in P_k : w_1 = i, w_k = s\}$. In altre parole la k -esima potenza della matrice di adiacenza ci restituisce il numero di cammini di lunghezza k tra qualsiasi coppia di nodi. Si può mostrare che un grafo (diretto) è (fortemente) connesso se e solo se la sua matrice di adiacenza è irriducibile.

Teorema 1.2.2. *Sia A una matrice $n \times n$, $A \geq 0$, irriducibile con periodo h . Consideriamo l'insieme degli autovalori di A come nel teorema 1.2.1. Allora le seguenti affermazioni sono vere:*

1. $\lambda_1 \in \mathbb{R}^+$ (radice di Perron);
2. $\lambda_1, \lambda_2 = \lambda_1 e^{2\pi i \frac{1}{h}}, \dots, \lambda_h = \lambda_1 e^{2\pi i \frac{h-1}{h}} \in \Lambda$; hanno tutti molteplicità algebrica 1 e $\lambda_i = |\lambda_i| \quad \forall i = 1, \dots, h$.
3. $\exists v > 0$ e $w > 0$ vettori tali che $Av = \lambda_1 v$ e $w^T A = \lambda_1 w^T$ (vettori di Perron)
4. Sia v_i (w_i) un autovettore destro (sinistro) associato a λ_i . Se $v_i > 0$ ($w_i > 0$) allora $i \in \{1, \dots, h\}$.
5. Se $h = 1$ allora $\lim_{k \rightarrow \infty} \frac{A^k}{\lambda_1^k} = vw^T =: \mu$ (proiezione di Perron) con v e w normalizzati in modo che $w^T v = 1$. La velocità di convergenza è dell'ordine di $(\frac{\lambda_2}{\lambda_1})^k$;

Per altri risultati e dimostrazioni si può fare riferimento al capitolo 8 di [23].

Capitolo 2

Teoria spettrale dei grafi

2.1 Matrici associate a un grafo

Consideriamo inizialmente $G = (V, E)$ un grafo indiretto non pesato. Per lavorare con la teoria spettrale dei grafi, vediamo i vettori $x \in \mathbb{R}^n$ come funzioni dai vertici del grafo ai numeri reali. Applicando la matrice di adiacenza A_G ad una funzione x otteniamo $(A_G x)(i) = \sum_{j:(i,j) \in E} x_j$.

Abbiamo visto come la matrice di adiacenza rappresenti il grafo, ma questa non è la sola matrice a lui legata. La prima matrice che introduciamo è la *matrice di diffusione* di G o *walk matrix* di G ; per definirla consideriamo D la matrice diagonale che ha come elementi della diagonale i gradi dei nodi, cioè

$$D_{ij} = \begin{cases} d_i & \text{se } i = j \\ 0 & \text{altrimenti} \end{cases}$$

Definizione 2.1. La *matrice di diffusione* di G è definita come $W_G = A_G D_G^{-1}$ e applicandola ad un vettore x otteniamo $(W_G x)(i) = \sum_{j:(i,j) \in E} \frac{x_j}{d_j}$

Un altro tipo di rappresentazione matriciale molto ricorrente è la matrice Laplaciana;

La *matrice Laplaciana* del grafo G è definita come $L = D - A$. Il suo nome non è casuale, infatti la matrice Laplaciana è un analogo discreto dell'operatore laplaciano continuo. L'idea è discretizzare l'operatore laplaciano,

rappresentando lo spazio come una griglia di nodi. Per i dettagli si veda la prima sezione dell'articolo [8].

Ad ogni matrice A si può associare una forma quadratica $x^T Ax$. Nel caso del Laplaciano

$$x^T Lx = \sum_{(i,j) \in E} (x_i - x_j)^2. \quad (2.1)$$

Molte volte può essere utile normalizzare il Laplaciano tramite i gradi. Il *Laplaciano normalizzato* di G è definito da

$$L_N = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

2.2 Teoria spettrale del Laplaciano

Quello che faremo in questa sezione è studiare lo spettro della matrice Laplaciana, cercando di ottenere delle limitazioni per suoi autovalori e studiandone la rappresentazione. Per fare questo diamo alcune definizioni preliminari:

Definizione 2.2. Una matrice simmetrica A si dice *semidefinita positiva* se $\forall x \in \mathbb{R}^n$ si ha $x^T Ax \geq 0$, cioè se la sua forma quadratica associata è non negativa.

Osservazione 6. Il Laplaciano è una matrice semidefinita positiva, segue banalmente dalla sua forma quadratica.

Una matrice semidefinita positiva ha tutti gli autovalori non negativi, infatti basta considerare la matrice diagonalizzata $A = Q^T \Lambda Q$ con Λ la matrice diagonale con gli autovalori di A come elementi, $x^T \Lambda x = \sum_i \lambda_i x_i^2 \geq 0 \quad \forall y \in \mathbb{R}^n$ se e solo se $\lambda_i \geq 0 \quad \forall i$.

Quindi possiamo ordinare gli autovalori del Laplaciano nel modo seguente: $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

Un modo alternativo per mostrare che il Laplaciano è semidefinito positivo utilizza il seguente lemma:

Lemma 2.2.1. *A è semidefinita positiva se e solo se esiste una matrice B tale che $A = B^T B$. B può non essere quadrata e soprattutto non è unica.*

L si può fattorizzare nel modo precedente usando la *matrice di incidenza* ∇ $m \times n$ con m il numero di spigoli e n il numero di vertici, definita come segue:

$$\nabla_{v,e} = \begin{cases} 1, & \text{se } e = (v, w) \text{ per qualche } w > v \\ -1, & \text{se } e = (v, w) \text{ per qualche } w < v \\ 0, & \text{altrimenti} \end{cases}$$

Teorema 2.2.2. $L = \nabla^T \nabla$

Dimostrazione. $(\nabla^T \nabla)_{ij} = (i\text{-esima colonna di } \nabla)(j\text{-esima colonna di } \nabla) = \sum_e \nabla_{e,v_i} \nabla_{e,v_j}$. Abbiamo tre casi:

1. se $i = j$, allora

$$(\nabla^T \nabla)_{ij} = \sum_e (\nabla_{e,v_i})^2 = \sum_{e \text{ incidenti a } v_i} 1 = \text{deg}(i)$$

2. se $i \neq j$ e non esiste uno spigolo che collega v_i e v_j , allora

$$(\nabla^T \nabla)_{ij} = \sum_e (\nabla_{e,v_i})(\nabla_{e,v_j}) = 0$$

3. se $i \neq j$ ed esiste uno spigolo e_1 fra v_i e v_j , allora

$$(\nabla^T \nabla)_{ij} = \sum_e (\nabla_{e,v_i})(\nabla_{e,v_j}) = \sum_e (\nabla_{e_1,v_i})(\nabla_{e_1,v_j}) = -1$$

□

Osservazione 7. $\mathbf{1} = (1, \dots, 1)$ è un autovettore di L relativo all'autovalore 0, infatti $(Lx)_i = d_i(x_i - \frac{\sum_{j:(i,j) \in E} x_j}{d_i})$ e con $x = (1, \dots, 1)$ il termine nella parentesi è 0.

Riepilogando abbiamo che gli autovalori di L possono essere scritti come $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ e l'autovettore relativo a 0 è $\mathbf{1}$.

Per questo motivo l'analisi degli autovalori si concentra su λ_2 . Per trovare qualche caratteristica di λ_2 studiamo la dimensione dello spazio N dato dai vettori x tali che $Lx = 0$.

Proposizione 2.2.3. *Se G è connesso, allora lo spazio N ha dimensione 1 ed è generato dal vettore $\mathbf{1}$.*

Dimostrazione. Sia $x \in N$, allora

$$x^T Lx = \sum_{(i,j) \in E} (x_i - x_j)^2 = 0. \quad (2.2)$$

La 2.2 implica che $x_i = x_j \quad \forall (i, j) \in E$. Visto che G è connesso, abbiamo che tutti gli x_i sono uguali, quindi ogni vettore dello spazio N è multiplo del vettore $\mathbf{1}$. \square

Corollario 2.2.4. *Se G è connesso, $\lambda_2 > 0$*

Inoltre si dimostra facilmente che la dimensione dello spazio N altro non è che il numero di componenti connesse di G .

Quello che vogliamo fare adesso è cercare delle limitazioni per gli autovalori di L .

Una prima limitazione banale può essere la seguente:

Lemma 2.2.5. *Sia G un grafo con n vertici ognuno con grado d_i . Sia $d_{max} = \max_i d_i$, L laplaciano con autovalori λ_i , allora*

$$\sum_i \lambda_i = \sum_i d_i \leq d_{max} n$$

Dimostrazione. La disuguaglianza è banale, mentre l'uguaglianza è data dal fatto che sia $\sum_i d_i$, sia $\sum_i \lambda_i$ sono espressioni della traccia. \square

Lemma 2.2.6. *Nella notazione del lemma 2.2.5, si ha*

$$\lambda_2 \leq \frac{\sum_i d_i}{n-1} \quad (2.3)$$

$$\lambda_n \geq \frac{\sum_i d_i}{n-1} \quad (2.4)$$

Dimostrazione. Per il lemma 2.2.5 e visto che $\lambda_1 = 0$ si ha $\sum_{i=2}^n \lambda_i = \sum_i d_i$. visto che $\lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_n$ si ha $(n-1)\lambda_2 \leq \sum_i d_i$. Nello stesso modo si dimostra la 2.4. \square

Definizione 2.3. Definiamo il *Quoziente di Rayleigh* di un vettore non nullo x rispetto a una matrice simmetrica A come

$$Q_R := \frac{x^T Ax}{x^T x} \quad (2.5)$$

Courant-Fischer Sono riusciti a caratterizzare tutti gli autovalori di una matrice simmetrica con il seguente teorema:

Teorema 2.2.7 (Courant-Fischer). *Sia A una matrice simmetrica $n \times n$ con autovalori $\lambda_1 \leq \dots \leq \lambda_n$ e con v_1, \dots, v_n i corrispondenti autovettori. Allora per $1 \leq k \leq n$, preso $S_k = \text{span}\{v_1, \dots, v_n\}$ (con $S_0 = 0$) e S_k^\perp il suo ortogonale,*

$$\lambda_k = \min_{\substack{\|x\|=1 \\ x \in S_{k-1}^\perp}} x^T Ax = \min_{\substack{x \neq 0 \\ x \in S_{k-1}^\perp}} \frac{x^T Ax}{x^T x}.$$

Inoltre

$$\lambda_{max} = \max_{\|x\|=1} x^T Ax = \max_{x \neq 0} \frac{x^T Ax}{x^T x}$$

Dimostrazione. Sia $A = Q^T \Lambda Q$ la diagonalizzazione di A . Come prima ci basta considerare $A = \Lambda$ e $x^T Ax = \sum_{i=1}^n \lambda_i x_i^2$.

Notiamo che quando A è diagonale, gli autovettori di A sono $v_k = e_k$, la base canonica di \mathbb{R}^n . La condizione $x \in S_{k-1}^\perp$ implica che $x \perp e_i$ per $i = 1, \dots, k-1$, quindi $x_i = \langle x, e_i \rangle = 0$. Perciò, per $x \in S_{k-1}^\perp$ con $\|x\| = 1$, abbiamo:

$$x^T Ax = \sum_{i=k}^n \lambda_i x_i^2 \geq \lambda_k \sum_{i=k}^n x_i^2 = \lambda_k \|x\|^2 = \lambda_k$$

D'altra parte, imponendo $x = e_k \in S_{k-1}^\perp$ otteniamo $x^T Ax = (e_k)^T A e_k = \lambda_k$. Grazie a questo abbiamo che

$$\lambda_k = \min_{\substack{x \in S_{k-1}^\perp \\ \|x\|=1}} x^T Ax$$

Allo stesso modo, per $\|x\| = 1$, si ha:

$$x^T Ax = \sum_{i=1}^n \lambda_i x_i^2 \leq \lambda_{max} \sum_{i=1}^n x_i^2 = \lambda_{max} \|x\|^2 = \lambda_{max}.$$

Ora, prendendo $x = e_n$ otteniamo $x^T A x = (e_n)^T A e_n = \lambda_{max}$. Quindi possiamo concludere che

$$\lambda_{max} = \max_{\|x\|=1} x^T A x.$$

□

Corollario 2.2.8. *Sia G un grafo indiretto, non pesato con laplaciano L , allora:*

$$\lambda_1 = 0 \quad v_1 = 1 \tag{2.6}$$

$$\lambda_2 = \min_{\substack{x \perp v_1 \\ x \neq 0}} \frac{x^T L x}{x^T x} = \min_{\substack{x \perp 1 \\ x \neq 0}} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i \in V} x_i^2} \tag{2.7}$$

$$\lambda_{max} = \max_{x \neq 0} \frac{x^T L x}{x^T x} = \max_{x \neq 0} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i \in V} x_i^2} \tag{2.8}$$

Possiamo interpretare la formula per λ_2 come mettendo delle molle su ogni spigolo (con delle condizioni al limite che corrispondono alla normalizzazione) e minimizzando l'energia potenziale della configurazione.

Inoltre si può ottenere una approssimazione di λ_2 e anche di λ_{max} costruendo rispettivamente un vettore x che renda piccolo il quoziente di Rayleigh e uno che lo renda grande.

Un altro teorema caratterizza gli autovalori di una matrice simmetrica in termini di massimizzazione e minimizzazione del quoziente di Rayleigh. Questo risultato gioca una parte fondamentale nella teoria spettrale dei grafi.

Teorema 2.2.9. *Sia A una matrice simmetrica con autovalori ordinati come nei precedenti teoremi. Allora*

$$\lambda_k = \max_{\substack{S \subset (R)^n \\ \dim(S)=k}} \min_{\substack{x \in S \\ x \neq 0}} Q_R = \min_{\substack{T \subset (R)^n \\ \dim(T)=n-k+1}} \max_{\substack{x \in T \\ x \neq 0}} Q_R.$$

Il massimo nella prima espressione è fatto su tutti i sottospazi di dimensione k , mentre il minimo nella seconda su tutti i sottospazi di dimensione $n - k + 1$.

La dimostrazione di questo teorema la si trova in [25]

2.3 Disuguaglianza di Cheeger

Un altro risultato importante che riguarda il secondo autovalore di L si tratta della disuguaglianza di Cheeger. Abbiamo già scritto che il laplaciano di un grafo è anche interpretabile come la discretizzazione dell'operatore laplaciano su varietà differenziabili. I risultati in teoria dei grafi e in geometria spettrale Riemanniana sono dunque collegati, e permettono di ottenere nuovi risultati in entrambi i campi. Il concetto di costante isoperimetrica, che defineremo in questa sezione, per esempio, è presente anche in geometria riemanniana, ed è legato al più piccolo autovalore non nullo dell'operatore laplaciano tramite la Disuguaglianza di Cheeger. Questa disuguaglianza è stata trattata ampiamente nella letteratura, si veda ad esempio [10] nel caso di varietà differenziabili, mentre nell'ambito della teoria dei grafi, ad esempio, da Chung in [7], che utilizza quattro dimostrazioni diverse per ottenere la disuguaglianza usando però il Laplaciano normalizzato. Noi invece la dimostreremo per il Laplaciano non normalizzato, seguendo l'approccio in [6].

Prendiamo $G = (V, E)$ un grafo e consideriamo $S \subset V$ un insieme di vertici di G e il suo complementare $S^c = V \setminus S$. Sia $e(S, S^c)$ il numero di spigoli fra S e S^c . Una partizione in due insiemi disgiunti come la precedente è chiamata *taglio* $S - S^c$

Definizione 2.4. Definiamo ϕ di un taglio $S - S^c$ come

$$\phi(S) = \frac{e(S, S^c)}{\min(|S|, |S^c|)} \quad (2.9)$$

Il numero isoperimetrico del grafo G si definisce come

$$\phi(G) = \min_{S \subset V} \phi(S).$$

Osservazione 8. Si vede facilmente che $\min(|S|, |S^c|) \leq \frac{2}{n}|S||S^c|$, poichè $|S| + |S^c| = n$.

Come detto all'inizio del capitolo, ogni vettore x deve essere considerato una mappa dall'insieme dei vertici V ad \mathbb{R} . In questo modo, il nostro grafo può essere immaginato sulla retta dei reali e ad un taglio nel grafo si ha un corrispettivo taglio in \mathbb{R} .

Associamo ora ad ogni taglio $S - S^c$ un vettore x tale che:

$$x_i = \begin{cases} 1, & \text{se } i \in S \\ -1, & \text{se } i \in S^c \end{cases}.$$

Allora abbiamo che

$$e(S, S^c) = \frac{1}{4} \sum_{(i,j) \in E} (x_i - x_j)^2.$$

Inoltre

$$|S||S^c| = \left(\sum_{i \in V} \mathbf{1}_{i \in S} \right) \left(\sum_{j \in V} \mathbf{1}_{j \in S^c} \right) = \sum_{i,j \in V} \mathbf{1}_{i \in S, j \in S^c} = \frac{1}{2} \sum_{i,j \in V} \mathbf{1}_{x_i \neq x_j} = \frac{1}{4} \sum_{i < j} (x_i - x_j)^2.$$

Quindi

$$\min_{x \in \{-1,1\}^n} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i < j} (x_i - x_j)^2} = \min_{S \subset V} \frac{e(S, S^c)}{|S||S^c|}. \quad (2.10)$$

Consideriamo ora

$$\min_{x \in \mathbb{R}^n} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i < j} (x_i - x_j)^2}.$$

Possiamo traslare $x \in \mathbb{R}^n$ in modo tale che $x \perp \mathbf{1}$, cioè $\sum_{i=1}^n x_i = 0$. Allora si ha che $\sum_{i < j} (x_i - x_j)^2 = n \sum_{i=1}^n x_i^2$, ed utilizzando il corollario 2.2.8

otteniamo

$$\min_{x \in \mathbb{R}^n} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i < j} (x_i - x_j)^2} = \min_{\substack{x \in \mathbb{R}^n \\ x \perp \mathbf{1}}} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i=1}^n x_i^2} = \frac{\lambda_2}{n}. \quad (2.11)$$

Mettendo insieme l'osservazione 8, la 2.10 e la 2.11 otteniamo

$$\begin{aligned} \phi(G) &= \min_{S \subset V} \frac{e(S, S^c)}{\min(|S|, |S^c|)} \geq \frac{n}{2} \min_{S \subset V} \frac{e(S, S^c)}{|S||S^c|} = \frac{n}{2} \min_{x \in \{-1,1\}^n} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i < j} (x_i - x_j)^2} \\ &\geq \frac{n}{2} \min_{x \in \mathbb{R}^n} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{\sum_{i < j} (x_i - x_j)^2} = \frac{n}{2} \min_{\substack{x \in \{-1,1\}^n \\ x \perp \mathbf{1}}} \frac{\sum_{(i,j) \in E} (x_i - x_j)^2}{n \sum_{i=1}^n x_i^2} = \frac{\lambda_2}{2} \end{aligned}$$

Abbiamo ottenuto la parte destra della disuguaglianza di Cheeger, che è la seguente:

Teorema 2.3.1. *Preso un grafo G ,*

$$\frac{\phi(G)^2}{2d_{max}} \leq \lambda_2 \leq 2\phi(G)$$

dove d_{max} è il massimo grado di G .

Per mostrare la disequazione di sinistra, dimostreremo un teorema più forte:

Teorema 2.3.2. *Per ogni $x \perp \mathbf{1}$, con la convenzione $x_1 \leq x_2 \leq \dots \leq x_n$, esiste i per cui*

$$\frac{x^T L x}{x^t x} \geq \frac{\phi(G)^2}{2d_{max}}.$$

Si vede facilmente che con $x = v_2$ è proprio la disuguaglianza cercata.

Prima di iniziare la dimostrazione facciamo qualche osservazione. Supponiamo per semplicità che n sia dispari e definiamo $m = \frac{n+1}{2}$ e il vettore y come $y_i = x_i - x_m$. Per costruzione $y_m = 0$.

Osservazione 9.

$$\frac{x^T Lx}{x^T x} \geq \frac{y^T Ly}{y^T y}$$

Dimostrazione.

$$x^T Lx = \sum_{(i,j) \in E} (x_i - x_j)^2 = \sum_{(i,j) \in E} ((y_i + x_m) - (y_j + x_m))^2 = \sum_{(i,j) \in E} (y_i - y_j)^2 = y^T Ly.$$

Quindi il numerato è lo stesso. Visto che $x \perp \mathbf{1}$ si ha per il denominatore

$$y^T y = (x + x_m \mathbf{1})^T (x + x_m \mathbf{1}) = x^T x + 2x_m (x^T \mathbf{1}) + x_m^2 (\mathbf{1}^T \mathbf{1}) = x^T x + nx_m^2 \geq x^T x$$

□

Ora ogni spigolo (i, j) con $i \leq m \leq j$ lo sostituiamo con altri due spigoli (i, m) e (m, j) . Chiamiamo quindi il nuovo insieme di spigoli E' . Questa operazione ci servirà per separare vertici negativi e positivi, ma prima facciamo un'altra osservazione:

Osservazione 10.

$$\frac{\sum_{(i,j) \in E} (y_i - y_j)^2}{\sum_{i \in V} y_i^2} \geq \frac{\sum_{(i,j) \in E'} (y_i - y_j)^2}{\sum_{i \in V} y_i^2}.$$

Dimostrazione. L'unica differenza è che nel numeratore di sinistra la somma è sugli spigoli (i, j) , che sono divisi in (i, m) e (m, j) in quello di destra.

$$\begin{aligned} (y_j - y_i)^2 &= (y_j - y_m + y_m - y_i)^2 = (y_j - y_m)^2 + (y_m - y_i)^2 + 2(y_j - y_m)(y_m - y_i) \\ &\geq (y_j - y_m)^2 + (y_m - y_i)^2 \end{aligned}$$

infatti $(y_j - y_m)(y_m - y_i) \geq 0$

□

Adesso vogliamo dividere la somma in due parti, una con i numeri positivi, l'altra con quelli negativi. Chiamiamo quindi E'_- l'insieme degli spigoli (i, j) con $i, j \leq m$ e E'_+ l'insieme degli spigoli (i, j) con $i, j \geq m$. Abbiamo:

$$\frac{\sum_{(i,j) \in E'} (y_i - y_j)^2}{\sum_i y_i^2} = \frac{\sum_{(i,j) \in E'_-} (y_i - y_j)^2 + \sum_{(i,j) \in E'_+} (y_i - y_j)^2}{\sum_{i=1}^m y_i^2 + \sum_{i=m}^n y_i^2}. \quad (2.12)$$

Lemma 2.3.3. Per ogni $a, b, c, d > 0$ vale

$$\frac{a+b}{c+d} \geq \min\left(\frac{a}{c}, \frac{b}{d}\right)$$

Dimostrazione. Supponiamo che $\frac{b}{d} \geq \frac{a}{c}$. Allora moltiplicando per c e per d e sommando ac ad entrambi i termini otteniamo $ca + cb \geq ac + ad$, da cui $\frac{a+b}{c+d} \geq \frac{a}{c} = \min(\frac{a}{c}, \frac{b}{d})$. Analogo è il caso in cui il minimo sia $\frac{b}{d}$. \square

Il lemma 2.3.3 precedente può essere applicato alla frazione 2.12 trovando che è limitata o da $\frac{\sum_{(i,j) \in E'_-} (y_i - y_j)^2}{\sum_{i=1}^m y_i^2}$ oppure da $\frac{\sum_{(i,j) \in E'_+} (y_i - y_j)^2}{\sum_{i=m}^n y_i^2}$. Per semplicità supponiamo che sia limitata dal primo.

Dopo queste osservazioni, siamo pronti per dimostrare un lemma fondamentale che ci servirà per la dimostrazione principale: Sia C_i il numero di spigoli che attraversano x_i , cioè il numero di spigoli nel taglio se dovessimo prendere $S = \{1, \dots, i\}$. Osserviamo che prendendo appunto $S = \{1, \dots, i\}$, abbiamo $C_i \geq \phi i$ per $i \leq \frac{n}{2}$ e $C_i \geq \phi(n - i)$ per $i \geq \frac{n}{2}$ con $\phi = \phi(G)$.

Lemma 2.3.4. Per ogni $z_1 \leq \dots \leq z_m = 0$ si ha

$$\sum_{(i,j) \in E'_-} |z_i - z_j| \geq \phi \sum_{i=1}^{j-1} |z_i|.$$

Dimostrazione. Per ogni $(i, j) \in E'_-$ con $i < j$, scriviamo

$$|z_i - z_j| = z_j - z_i = \sum_{k=i}^{j-1} (z_{k+1} - z_k).$$

Ogni termine $z_{k+1} - z_k$ appare esattamente C_k volte, allora:

$$\sum_{(i,j) \in E'_-} |z_i - z_j| = \sum_{k=1}^{m-1} C_k (z_{k+1} - z_k) \geq \phi \sum_{k=1}^{m-1} k (z_{k+1} - z_k).$$

Visto che $z_i \leq z_m = 0$, si ha $|z_i| = -z_i$ per $1 \leq i \leq m$. Allora

$$\begin{aligned} \sum_{k=1}^{m-1} k(z_{k+1} - z_k) &= (z_2 - z_1) + 2(z_3 - z_2) + \dots + (m-1)(z_m - z_{m-1}) = \\ &= -z_1 - z_2 - \dots - z_{m-1} + (m-1)z_m = \sum_{i=1}^m |z_i|, \end{aligned}$$

quindi

$$\sum_{(i,j) \in E'_-} |z_i - z_j| \geq \phi \sum_{i=1}^m |z_i|.$$

□

Ora abbiamo tutti gli strumenti per dimostrare la disuguaglianza di Cheeger

Dimostrazione. Per prima cosa consideriamo y normalizzato in modo tale che $\sum_{i=1}^m y_i^2 = 1$.

Ora, vogliamo ottenere dei quadrati nella nostra espressione ed è per questo che applichiamo il lemma 2.3.4 al nuovo vettore z definito come $z_i = -y_i^2$. In questo modo otteniamo:

$$\sum_{(i,j) \in E'_-} |y_i^2 - y_j^2| \geq \phi \sum_{i=1}^m |y_i^2| = \phi.$$

Ora, vogliamo qualcosa simile a $(y_i - y_j)^2$ invece che $y_i^2 - y_j^2$, quindi usiamo la disuguaglianza di Caucht-Schwarz:

$$\begin{aligned} \sum_{(i,j) \in E'_-} |y_i^2 - y_j^2| &= \sum_{(i,j) \in E'_-} |y_i - y_j| |y_i + y_j| \\ &\leq \left(\sum_{(i,j) \in E'_-} (y_i - y_j)^2 \right)^{\frac{1}{2}} \left(\sum_{(i,j) \in E'_-} (y_i + y_j)^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Adesso vogliamo liberarci della parte con $(y_i + y_j)^2$, quindi cercheremo di trovare una sua maggiorazione. Possiamo notare che il massimo numero

di volte che ogni y_i^2 compare nella sommatoria è il grado massimo di ogni vertice:

$$\sum_{(i,j) \in E'_-} (y_i + y_j)^2 \leq 2 \sum_{(i,j) \in E'_-} (y_i^2 + y_j^2) \leq 2 \sum_{i=1}^m d_{max} y_i^2 \leq 2d_{max}.$$

Quindi otteniamo che

$$\frac{\sum_{(i,j) \in E'_-} (y_i - y_j)^2}{\sum_{i=1}^m y_i^2} \geq \frac{\left(\sum_{(i,j) \in E'_-} (y_i^2 - y_j^2) \right)^2}{\sum_{(i,j) \in E'_-} (y_i + y_j)^2} \geq \frac{\phi^2}{2d_{max}}.$$

In modo analogo possiamo mostrare che anche

$$\frac{\sum_{(i,j) \in E'_+} (y_i - y_j)^2}{\sum_{i=1}^m y_i^2} \geq \frac{\phi^2}{2d_{max}}.$$

Allora si ha

$$\frac{x^T Lx}{x^T x} \geq \frac{y^T Ly}{y^T y} \geq \min \left(\frac{\sum_{(i,j) \in E'_-} (y_i - y_j)^2}{\sum_{i=1}^m y_i^2}, \frac{\sum_{(i,j) \in E'_+} (y_i - y_j)^2}{\sum_{i=1}^m y_i^2} \right) \geq \frac{\phi^2}{2d_{max}}.$$

□

Capitolo 3

Misure di centralità

In questa sezione di occupiamo della questione di determinare l'importanza di un nodo in un grafo. Così posta, ovviamente, la questione è molto vaga, ma vedremo che, in funzione di ciò che si vuole studiare, esistono delle quantità, dette misure di centralità, che in maniera più o meno soddisfacente, danno risposte quantitative per questo tipo di analisi.

In realtà abbiamo già incontrato una di queste misure nel capitolo 1. Il grado di un nodo, definito nella 1.5, è proprio una misura di centralità, che conta, nel caso di un grafo non pesato, quanti sono i vicini di un determinato nodo. Nelle prossime sezioni tratteremo solo alcune misure di centralità, in particolare quelle che abbiamo utilizzato nel nostro lavoro. Altre misure sono trattate in [1], [15], [14], [12] e [13].

3.1 Closeness Centrality

Definizione 3.1. Sia G grafo connesso indiretto. La *distanza* fra due nodi i e j è $d(i, j) = \min\{k \in \mathbb{N} \text{ t.c. } \omega \in P_k, \omega_1 = i, \omega_k = j\}$, cioè la lunghezza del percorso più breve che connette i due nodi. Definiamo ora la *Closeness Centrality* di i come

$$C_i = \frac{1}{\sum_{j \in V} d(i, j)} \quad (3.1)$$

Più un nodo è vicino agli altri (distanza piccola), più la sua Closeness Centrality sarà grande.

Se G non è connesso si definisce la Centralità armonica di i come

$$H_i = \sum_{j \in V} \frac{1}{d(i, j)}$$

con $d(i, j) = \infty$ se i non è connesso con j e con la convenzione che $\frac{1}{\infty} = 0$.

Si può anche modificare leggermente la definizione considerando la dimensione n della rete e il numero n_i di nodi raggiungibili da i in questo modo:

$$c_i = \left(\frac{n_i}{n-1}\right)^2 C_i$$

Nel caso di grafi diretti fortemente connessi definiamo

$$c_i^{out} = \left(\frac{n_i}{n-1}\right)^2 \frac{1}{\sum_{j \in V} d^+(i, j)}$$

$$c_i^{in} = \left(\frac{m_i}{n-1}\right)^2 \frac{1}{\sum_{j \in V} d^+(i, j)}$$

con n_i e m_i rispettivamente il numero di nodi che raggiungono i e che sono raggiunti da i e con $d^+(i, j)$ la distanza fra i nodi i e j considerando i cammini $\omega \in P_k^+$.

3.2 Betweenness Centrality

In molte reti l'importanza è vista come la capacità di un nodo di trovarsi in mezzo a molti percorsi brevi che collegano gli altri nodi. L'importanza ci dice se un nodo è un buon mediatore o intermediario nella rete.

Definizione 3.2. Sia G un grafo diretto. Definiamo la *Betweenness Centrality* di i come

$$B_i = \sum_{l, s \in V, l \neq s} \frac{g(l, i, s)}{g(l, s)} \quad (3.2)$$

dove $g(l, i, s)$ è il numero dei percorsi più brevi tali che $w_1 = l, w_k = s$ ed esiste un $j \in \{2, \dots, k-1\}$ per cui $w_j = i$, mentre $g(l, s) = \sum_{j \in V} g(l, j, s)$.

Possiamo vedere i termini della somma B_i come la probabilità di passare da i attraverso il percorso più breve che collega l ed s . B_i quantifica cioè quanto sia presente i nei percorsi più brevi fra ogni coppia di nodi.

Nel caso di grafo indiretto abbiamo $g(l, s) = g(s, l)$ e $g(l, i, s) = g(s, i, l)$ $\forall l, s \in V$, con $l \neq s$. Si definisce allora $B_i^u = \frac{1}{2}B_i$.

3.3 Eigenvector Centrality

Sia G un grafo indiretto; consideriamo i, j due nodi con lo stesso grado, cioè $d_i = d_j$; Per la degree centrality questi due nodi hanno esattamente la stessa importanza, ma spesso non è così. Uno dei due nodi potrebbe essere legato a nodi più importanti rispetto all'altro. Con l'eigenvector centrality vogliamo fare in modo che il grado di i cresca se quest'ultimo fosse connesso con nodi 'importanti'.

Definizione 3.3. Sia $G = (V, E)$ (in)diretto fortemente connesso con matrice di adiacenza A . Definiamo l'*Eigenvector Centrality* del nodo i come $E_i = p_i$ dove p è l'autovettore relativo all'autovalore più grande di A , cioè p soddisfa $A\lambda_1 = \lambda_1 p$.

In [2] la definizione è leggermente differente: si definisce l'eigenvector centrality x_i di un nodo i come

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n A_{ij} x_j,$$

con λ costante. Preso x il vettore delle centralità, possiamo riscrivere l'equazione in forma matriciale come $\lambda x = Ax$ e quindi notare che x è un autovettore relativo all'autovalore λ . Volendo la misura di centralità non negativa, allora, usando il teorema 1.2.1, si mostra che λ deve essere il più grande autovalore di A , ottenendo quindi la definizione 3.3.

Questa misura di centralità risolve il problema citato ad inizio paragrafo. Questo lo si capisce bene costruendo il vettore di centralità x seguendo l'approccio in [1], definendo una successione di vettori (x_k) , con x_{k+1} una

versione migliorata di x_k nel senso che aumentiamo la centralità $(x_k)_i$ del nodo i se questo è connesso a nodi importanti.

Poichè vale il teorema 1.2.1, abbiamo che $\lambda_1 \geq |\lambda_i| \quad \forall i \neq 1$; definiamo allora $x_0 = \frac{1}{\lambda_1} \mathbf{1}$ con $\mathbf{1} = (1, \dots, 1)$. Inizialmente ogni nodo ha la stessa importanza. Definiamo poi $x_1 = \frac{1}{\lambda_1} Ax_0 = \frac{1}{\lambda_1} d$.

In generale

$$(x_{k+1})_i = \frac{1}{\lambda_1} \sum_{j/(i,j) \in E} (x_k)_j = \frac{1}{\lambda_1} \sum_j a_{ij} (x_k)_j$$

quindi

$$x_{k+1} = Ax_k = \frac{1}{\lambda_1^k} A^k x_0.$$

Il limite di questa successione sarà il vettore x cercato. Sappiamo che $x_0 = \sum_i c_i v_i$ con v_i autovettore di A . Allora $x_k = A^k x_0 = A^k \sum_i c_i v_i = \sum_i c_i A^k v_i = \sum_i c_i \lambda_i^k v_i = \lambda_1^k \sum_i c_i \frac{\lambda_i^k}{\lambda_1^k} v_i$

allora

$$\lim_{k \rightarrow \infty} \frac{1}{\lambda_1^k} x_k = c_1 v_1.$$

Il vettore trovato è proporzionale a v_1 , questo significa che x soddisfa $Ax = \lambda_1 x$, cioè x è autovettore relativo all'autovalore λ_1 . Per il teorema 1.2.1 di Perron-Forbenius sappiamo, ovviamente quando le ipotesi sono verificate, che esiste un autovettore $p > 0$ relativo a λ_1 . Definendo allora l'engivector centrality di i come $E_i = p_i$, otteniamo proprio la nostra definizione.

3.4 PageRank Centrality

PageRank è l'algoritmo utilizzato da Google per ordinare l'elenco dei risultati dei motori di ricerca. Possiamo vedere che al World Wide Web si associa un grafo in cui i nodi sono i siti Web e gli spigoli sono i collegamenti da un sito a un altro. Un'osservazione ovvia ma importante da fare è che se il sito web ha un link al sito j non è garantito che j abbia un link a i . Quindi il grafo può essere diretto e questo è uno dei punti di forza del metodo. Anche se è nato per lo scopo del sito web, PageRank costituisce una misura

di centralità che può essere utilizzata in ogni rete e può essere vista come una misura migliore rispetto a quella dell'autovalore. È chiaro che questo metodo fornisce risultati sensati per i grafi non orientati. Per definire questa misura abbiamo seguito l'approccio suggerito in [3] e spiegato in [4].

Sia $G = (V, E)$ un grafo orientato e A la sua matrice di adiacenza. Il nostro obiettivo finale è costruire una matrice stocastica irriducibile in modo che possiamo usare il teorema di Perron-Frobenius per trovare una misura di centralità positiva per i nodi. Definiamo innanzitutto una matrice $H = (h_{ij})$ che rappresenti la probabilità per un nodo i di essere raggiunto (in un passo) dal nodo j ; se non c'è alcun collegamento da j a i la possibilità è 0, altrimenti è $\frac{1}{d_j^{out}}$. Sia $D = (d_{ij})$ la matrice diagonale tale che

$$d_{ij} = \begin{cases} d_j^{out} & \text{se } i = j \text{ e } d_j^{out} > 0 \\ 1 & \text{se } i = j \text{ e } d_j^{out} = 0. \\ 0 & \text{altrimenti} \end{cases}$$

Notiamo che la condizione apparentemente non necessaria su j tale che $d_j^{out} = 0$ assicura che D sia invertibile. Sia $H = A^T D^{-1} = (h_{ij}) = (a_{ji} d_{jj}^{-1})$. Osserviamo che se esiste j tale che $d_j^{out} = 0$, allora $a_{ji} = 0 \forall i \in \{1, \dots, n\}$ e questo implica che $h_{ij} = 0 \forall i \in \{1, \dots, n\}$.

Questi nodi sono chiamati *dangling nodes* e corrispondono alle colonne di H fatte di tutti gli zeri. L'obiettivo è rendere possibile il passaggio dai *dangling nodes* a qualche altro nodo. Per fare ciò introduciamo una distribuzione discreta uniforme per i *dangling nodes* come segue: denotiamo con I il vettore degli indici dei *dangling nodes* e definiamo a nel modo seguente:

$$a_i = \begin{cases} 1 & \text{se } i \notin I \\ 0 & \text{altrimenti} \end{cases}.$$

Definiamo $S = H + \frac{1}{n} \mathbf{1} a^T$.

In altre parole, stiamo ipoteticamente aggiungendo $(j, i) \in E \forall i \in \{1, \dots, n\}, \forall j \in I$. Questa può ancora essere una matrice riducibile, quindi

introduciamo una costante $\alpha \in (0, 1)$ (fattore di attenuazione di Google) e definiamo la matrice di Google:

$$G_\alpha = \alpha S + (1 - \alpha) \frac{1}{n} \mathbf{1}\mathbf{1}^T. \quad (3.3)$$

Ora abbiamo tutti gli elementi per definire la PageRank centrality, grazie anche al teorema 1.2.1 di Perron-Forbenius applicabile alla matrice G_α .

Definizione 3.4. Sia G_α come in 3.3. Definiamo la misura *PageRank* del nodo i come $p(\alpha)_i$ con $p(\alpha)$ il vettore di Perron associato all'autovalore 1.

Osserviamo che S rappresenta un cammino casuale su G , o equivalentemente un grafo diretto pesato in cui ogni spigolo (i, j) in G ha come peso la probabilità di raggiungere j da i , e dove abbiamo collegato tutti i nodi a tutti i dangling nodes j con peso $s_{ij} = \frac{1}{n}$. Ma avremmo potuto ancora avere nodi non raggiungibili da altri. Quindi abbiamo introdotto un nodo immaginario che con probabilità $1 - \alpha$ è raggiunto da qualsiasi altro nodo con probabilità uniforme $\frac{1}{n}$ quindi il grafo G diventa fortemente connesso e questo ci porta all'irriducibilità di G_α , grazie alla quale abbiamo potuto applicare il teorema 1.2.1 trovando il vettore di Perron $p(\alpha)$ associato all'autovalore 1, che ci ha permesso di definire la misura PageRank. Avremmo potuto scegliere qualsiasi altro vettore di probabilità invece di quello uniforme, ma questa scelta nel nostro caso è la più ragionevole. Questa è una procedura che mostra come "modifichiamo" le proprietà del grafo considerando G_α anziché G , per avere un ambiente matematico migliore.

La scelta del valore della costante α è fatta in modo prevalentemente empirico. Brin e Page suggerirono il valore di 0.85, valore ancora maggiormente usato e che utilizzeremo anche noi.

3.5 Hubs e Authority Centrality

In alcune reti può essere utile introdurre altre due misure di centralità.

Ad esempio, in una rete di articoli scientifici, uno di essi può citare altri articoli che sono fonti autorevoli per informazioni su un particolare argomen-

to. Una recensione può contenere poche informazioni sull'argomento, ma ci dice dove trovarle e questo da solo la rende utile. Allo stesso modo, ci sono molti esempi di pagine web che consistono principalmente in collegamenti ad altre pagine su un determinato argomento e tale pagina di collegamenti potrebbe essere molto utile anche se non contiene informazioni esplicite sull'argomento in questione.

Allora ci sono due tipi di nodi importanti in queste reti: le autorità sono nodi che contengono informazioni utili su un argomento di interesse; gli hubs sono nodi che ci dicono dove si trovano le migliori autorità. Un'autorità può anche essere un hub, e viceversa: gli articoli di revisione spesso contengono utili discussioni sulla visualizzazione di alcune eccezioni ad altre discussioni. Chiaramente hubs e autorità esistono solo nelle reti dirette, poiché nel caso non orientato non c'è distinzione tra il puntamento a un vertice e l'essere puntati.

Definiamo ora l'authority e la hubs centrality seguendo l'approccio in [1].

L'*authority* di un vertice i è definita per essere proporzionale alla somma delle hubs centrality dei vertici che puntano ad esso:

$$x_i = \alpha \sum_{j \in V} A_{ij} y_j$$

con α costante.

In modo analogo definiamo l'*hub* di un vertice i

$$y_i = \beta \sum_{j \in V} A_{ji} x_j$$

con β costante.

In forma matriciale abbiamo quindi $x = \alpha A y$ e $y = \beta A^T x$. Combinandole e con $\lambda = (\alpha\beta)^{-1}$ otteniamo $AA^T x = \lambda x$ e $A^T A y = \lambda y$.

Ciò significa che l'authority e l'hub centrality corrispondono rispettivamente ad un autovettore di AA^T e $A^T A$ relativi allo stesso autovalore. Anche in questo caso, si può dimostrare che si deve prendere l'autovalore massimo.

Capitolo 4

Misure di centralità applicate alla narrativa

Prima di mostrare ciò che abbiamo fatto e i risultati ottenuti, facciamo un breve excursus lungo i lavori principali dove, con tecniche legate alla teoria dei grafi, si è cercato di estrapolare informazioni da opere letterali.

In [17] Beveridge e Shan hanno applicato gli algoritmi di rete sui social network da loro generati da *A Storm of Swords*, il terzo romanzo di George R.R. Martin, serie *A Song of Ice and Fire* (che è l'origine letteraria della serie tv HBO *Game of Thrones*). Misure di centralità come PageRank, Betweenness, Closeness e modularità hanno fornito un approccio empirico per determinare comunità e personaggi chiave all'interno della rete. In questa tesi non è stata usata la modularità, quindi rinviamo a [14] per approfondimenti, mentre qui ne diamo una breve trattazione. Denotato con $\omega_{ij} \geq 0$ il peso dello spigolo che collega i nodi i e j consideriamo il grado pesato d_i del nodo i come definito nella 1.5. Definiamo la modularità Q di una partizione dei nodi del grafo C_1, \dots, C_l come

$$Q = \frac{1}{2n} \sum_{k=1}^l \sum_{i,j \in C_k} \left(\omega_{ij} - \frac{d_i d_j}{2n} \right),$$

con n il numero di nodi del grafo.

Gli autori dell'articolo hanno considerato i 107 personaggi dell'opera come i nodi di un grafo, collegati fra loro da spigoli pesati in base alle relazioni instaurate fra loro. Per determinare il peso hanno controllato quante volte i loro nomi comparissero entro 15 parole l'uno dall'altro. Successivamente hanno cercato la partizione in comunità che massimizzasse Q , trovando sette comunità distinte, di cui una contenente circa il 37% dei personaggi. Questa comunità ben più folta rispetto alle altre è stata nuovamente partizionata come fatto precedentemente ottenendo nuove quattro comunità.

Infine hanno calcolato varie misure di centralità per determinare i nodi principali del grafo, cioè i personaggi più importanti dell'opera. Le misure usate sono state il grado, sia pesato che non, l'eigenvector, la closeness, la betweenness e per finire la PageRank, definita seguendo l'approccio in [1], differente da quello seguito in questa tesi.

Ribeiro et al. [18] si concentrano sull'esame delle proprietà strutturali, come l'assortimento e la transitività, delle comunità nel grafo sociale di *J.R.R. Il Signore degli Anelli di Tolkien* (che includeva quel romanzo integrale, insieme al testo di *Lo Hobbit* e *Il Silmarillion*). La transitività, o coefficiente di clustering, è una misura che quantifica il numero di trinagoli in un grafo. Viene definito il coefficiente di clustering locale come

$$T_i = \frac{2n_i}{k_i(k_i - 1)}$$

dove k_i è il numero di vicini del nodo i e n_i è la quantità di spigoli che ci sono tra i vicini di i . Infine si definisce la transitività media come

$$T = \frac{1}{n} \sum_{i \in V} T_i,$$

con n il numero di nodi. L'assortività è definita come

$$\frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i},$$

dove e_{ij} è la frazione di spigoli che connettono i vertici i e j , $a_i = \sum_j e_{ij}$ e $b_j = \sum_i e_{ij}$. Per approfondimenti sull'assortimento si veda [12], mentre per la transitività [13].

Per formare il grafo dove i nodi rappresentavano i personaggi, gli autori hanno ricercato nel testo i nomi dei personaggi e analizzato le iterazioni fra di loro, che potevano essere o interazioni dirette, allusioni oppure eventi condivisi, per gli spigoli. La struttura trovata è composta da 618 nodi e 19462 spigoli. In questa rete sono stati considerati personaggi anche le divinità, per poi escludere successivamente formando un ulteriore grafo per confrontare i dati quantitativi calcolati, che sono il grado medio, cioè la somma dei gradi dei singoli nodi diviso il numero di nodi, la transitività, l'assortività e la media delle betweenness centrality. Come si può notare in questo caso non si guardano i singoli nodi, ma le proprietà strutturali del grafo.

Infine tramite l'ottimizzazione della modularità sono state individuate cinque comunità, sulle quali sono stati ricalcolati i valori quantitativi precedenti. Si è notato che le comunità mostravano alti valori di transitività, indicando alta densità nei collegamenti, mentre presentavano valori negativi sull'assortività, che denota la presenza di vertici con alta connessione collegati a vertici con connessione bassa.

Agarwal et al. [20] analizzano la rete di *Alice nel paese delle meraviglie*, definita dall'estrazione dei dieci capitoli indipendentemente gli uni dagli altri. Tale analisi può essere importante nel determinare i personaggi con misure di centralità globali basse, ma che sono significativamente importanti per una parte della storia. In questo lavoro si va oltre i limiti dell'analisi di una rete sociale aggiungendo un elemento cruciale: il dinamismo.

Per costruire il grafo Agarwal et al. definiscono due tipi di eventi sociali: l'interazione, in cui le persone sono consapevoli dell'evento sociale, come ad esempio una conversazione, e l'osservazione, nella quale solo una parte è a conoscenza dell'altro e dell'evento, come ad esempio il pensare o parlare di qualcuno. Con questi due tipi di eventi sociali, sono stati generati due grafi, uno indiretto e l'altro diretto. Per analizzarli sono utilizzate il grado, l'in-degree e l'out-degree nel caso diretto, l'hubs e l'authority centrality. Inizialmente sono stati considerati solo tre personaggi e cioè Alice, il Bianconiglio e il Topo e il testo in tutta la sua integrità. In questa prima parte

hanno usato l'hubs e l'authority centrality, trovando che Alice è praticamente sempre prima in classifica, tranne nel grafo diretto con l'authority dove al primo posto si trova il Bianconiglio. Questo significa che si parla molto di questo personaggio o lo si osserva molto, mentre nel grafo indiretto si trova al terzo posto, cioè nelle sue interazioni rimane secondario, rimanendo in primo piano dal punto di vista degli altri personaggi.

Successivamente, utilizzando il grado, l'in-degree e l'out-degree nel caso diretto, e considerando due personaggi in più, cioè la regina e il cappellaio, hanno stilato delle classifiche temporali, mostrando i personaggi chiave dei vari capitoli. Questo lavoro mostra per primo l'importanza della chiave temporale, risolvendo il problema della staticità dei grafi che non permetteva di osservare come alcuni personaggi fossero fondamentali in determinate occasioni

Bonato et al. [19] estrapolano i networks da tre romanzi, *Twilight* di Stephanie Meyer, *The Stand* di Steven King e *Harry Potter and the Goblet of Fire* di J.K. Rowling's, tramite le co-occorrenze dei nomi dei personaggi. Nei dettagli, se il nome di un personaggio e quello di un altro si trovavano a meno di quindici termini di distanza, allora si inserisce uno spigolo fra di loro. Il peso è dato dal numero di occorrenze fra due personaggi. Una volta generato il grafo, gli autori lo hanno analizzato usando varie misure di centralità e suddividendolo in comunità, ottenendo conclusioni letterarie accurate ed identificando con successo i personaggi più influenti e le costellazioni di personaggi minori nei vari libri.

Molti lavori sono poi stati fatti su opere di Shakespeare. Ad esempio V.H.Masias et al. [21] cercano di esplorare e confrontare l'importanza dei personaggi in *Romeo and Juliet*. Per fare ciò utilizzano $\alpha\omega$ -weighted degree, $\alpha\omega$ -weighted betweenness, $\alpha\omega$ -weighted closeness, facendo variare il parametro α ed ottenendo diverse classifiche. Non avendo utilizzato queste misure, rimandiamo a [15] per una loro trattazione.

Varie opere, come *Macbeth*, *Romeo and Juliet*, *Richard III*, *Twelfth night* e *The Winter's tale*, vengono analizzate in [16], da R. Perissinotti. Anche

in questo caso si ricerca l'importanza dei vari personaggi, generando prima dei grafi indiretti, collegando i personaggi che parlano fra di loro. Successivamente ha prodotto dei grafi diretti, uno per ogni personaggio k , con uno spigolo orientato dal nodo i verso quello j se i parla di k a j . Sui grafi non orientati ha calcolato alcune misure di centralità come il grado, la closeness e altre, mentre per quanto riguarda i grafi diretti ha definito due nuove misure di centralità per analizzare non il singolo nodo, ma il grafo nella sua integrità. Viene definita la misura di centralità del gap spettrale ponderata come

$$\lambda := N \frac{\lambda_1 - |\lambda_2|}{\lambda_1},$$

dove N è il numero di nodi del grafo e λ_1 e λ_2 rispettivamente la radice di Perron e il secondo autovalore più grande della matrice di adiacenza associata alla più grande componente connessa del grafo.

Definisce poi la misura di centralità del tempo medio ponderato come

$$MT := \frac{1}{N} \sum_{j,k} \frac{1}{N(N-1)} d(j,k),$$

dove $d(i,j)$ è la distanza dal nodo j al nodo k e N , come prima, il numero di nodi. Per i dettagli vedere [16].

Confrontando le calssifiche dei personaggi prima per i grafi indiretti, poi indiretti, Perissinotti nota come alcuni personaggi siano importanti nei grafi diretti, mentre secondari nei grafi indiretti. Questi personaggi sono molto spettegolati, al centro di discorsi, ma incapaci di cambiare la trama e gli eventi, in modo analogo a quanto visto per il Bianconiglio in [20].

Per concludere Perissinotti partiziona in comunità i grafi, utilizzando un nuovo algoritmo basato sulle celle di Voronoi. Per i dettagli vedere [16].

Capitolo 5

L'Otello di Shakespeare

Come abbiamo visto, in tutti i lavori precedenti si è cercato di trovare le comunità nei grafi, individuare i personaggi chiave vuoi in singoli capitoli, vuoi nell'integrità del testo; semplificando potremmo dire che rispondono alle domande 'Chi sono i personaggi più importanti?' e 'Che gruppi sociali, fazioni, o comunità ci sono nell'opera? '. Il nostro lavoro invece ha come punto centrale non più i personaggi in sé, ma i *topics*, gli argomenti che sono affrontati nell'opera. Di cosa parla questo testo? E' un testo religioso o che tratta il tema amoroso? E' centrale il tema della morte o quello dell'onestà, della virtù? Queste sono le domande a cui cerchiamo di dare una risposta.

Dopo questa breve introduzione che motiva ciò che abbiamo fatto, possiamo descrivere i particolari di questo lavoro.

Abbiamo preso in considerazione una sola opera di Shakespeare e cioè *The Tragedy of Othello, Moor of Venice* [22]. I campi semantici considerati sono quattordici: Body, Beauty, Death, Love, Mind, Money, Nature, Pleasure, Power, Religion, See-Seen, Speak, Truth, Virtue. Per ciascuno sono state scelte una serie di parole chiave che utilizzeremo per capire quando si parla di un determinato argomento.

A questo punto, potremmo suddividere il nostro lavoro in due parti.

5.1 Personaggi importanti per campi semantici

La prima parte consiste nel provare a capire quali siano i personaggi più importanti rispetto ad ogni singolo campo semantico. Per fare questo abbiamo considerato innanzitutto l'intero testo dell'opera, per poi creare un grafo indiretto pesato per ogni *topic*, in cui i nodi rappresentano i personaggi e questi sono collegati se parlano fra di loro usando una delle parole chiave che caratterizzano il *topic*. Come peso per gli spigoli usiamo il numero delle interazioni descritte prima che si hanno fra i due personaggi. Per fare tutto ciò abbiamo utilizzato Python, scrivendo un programma che leggesse il testo dell'opera, individuando le varie interazioni fra i personaggi e che restituisse in output la matrice di adiacenza pesata del grafo. Abbiamo quindi usato le misure di centralità descritte nel capitolo 3 per stilare le classifiche, una per ogni misura. Per quanto riguarda la closeness e la betweenness abbiamo usato il grafo non pesato, mentre per le altre misure il grafo pesato. Abbiamo poi ritenuto opportuno calcolare la closeness usando un grafo con un nuovo peso, che avvicinasse i personaggi che interagivano di più fra loro, ovviamente riguardo ad un determinato *topic*. Considerato p il peso definito prima, cioè il numero delle interazioni, definiamo il nuovo peso $p_2 = \frac{1}{p}$. In questo modo se i e j interagiscono molto, si troveranno a una distanza data da p_2 più piccola rispetto a quella di due nodi che interagiscono meno. In seguito, $closeness_p$ denoterà la misura ottenuta sul grafo pesato con p_2 , mentre $closeness$ quella sul grafo non pesato. Abbiamo quindi fatto la media aritmetica fra le posizioni ottenute da ogni personaggio per ogni misura ottenendo una classifica finale R_{media} . Per il calcolo delle misure di centralità e per produrre le classifiche abbiamo utilizzato matlab.

Abbiamo notato che sono sempre gli stessi personaggi ad occupare le posizioni più alte in ogni campo semantico, in particolare Iago, Othello, Cassio, Desdemona, Roderigo ed Emilia (vedi per esempio tabella 5.1).

A questo punto abbiamo voluto controllare se questi personaggi fossero anche i più importanti senza considerare particolari campi semantici. Quindi abbiamo generato un grafo, dove i nodi rappresentano i personaggi e li

abbiamo collegati con uno spigolo non orientato nel caso in cui ci fosse un'interazione fra loro. Come peso abbiamo scelto il numero totale di interazioni fra i due personaggi. Poi, calcolando le medesime misure di centralità di prima, abbiamo stilato una classifica di importanza generale dei personaggi per ognuna di esse e facendo una media come prima ottenuto una nuova classifica, vedi tabella 5.2. I sei personaggi che occupano le prime posizioni per ogni campo semantico sono proprio i personaggi più importanti secondo la classifica generale.

Herald, che compare in una sola scena facendo un proclamo al popolo, sembrerebbe il personaggio più importante dell'opera per molti *topics* basandosi sulle misure *betweenness* e *closeness*, questo perchè, usando il grafo non pesato e considerando il discorso fatto al popolo ascoltato da tutti, risulta collegato a tutti gli altri nodi. Usando la *closeness_p*, cioè sul grafo pesato, si risolve questo inconveniente ed Herald torna fra i personaggi minori.

Tabella 5.1: Ranking Virtue

Name	$R_{pagerank}$	$R_{eigenvector}$	$R_{betweenness}$	$R_{closeness}$	$R_{closenessp}$	R_{degree}	R_{media}
Iago	1	1	4	3	2	1	1
Othello	2	2	3	3	1	2	2
Cassio	4	4	2	2	3	4	3
Desdemona	3	3	7	6	4	3	4
Emilia	6	5	9	7	6	5	5
Roderigo	7	6	8	5	5	6	6
Montano	8	7	6	7	7	7	7
Brabantio	9	8	10	9	8	8	8
Herald	5	14	1	1	14	11	9
Duke	10	10	11	11	9	9	10
first senator	12	12	5	9	11	13	11
Lodovico	11	9	14	11	10	10	12
Gratiano	13	11	12	11	13	12	13
second senator	14	13	13	14	12	14	14
Bianca	17	15	15	16	15	15	15
first officer	16	16	15	14	16	15	16
messenger	15	18	15	17	17	17	17
Clown	18	17	15	17	18	18	18
third gentleman	19	20	15	19	19	19	19
second gentleman	20	19	15	19	20	19	20
first musician	22	21	15	21	20	21	21
sailor	21	23	15	21	20	21	22
first gentleman	23	22	15	21	20	23	23
fourth gentleman	24	24	15	24	20	24	24
Gentleman	24	25	15	24	20	24	25

Tabella 5.2: Classifica personaggi

Name	$R_{pagerank}$	$R_{eigenvector}$	$R_{betweenness}$	$R_{closeness}$	R_{degree}	R_{media}
Iago	1	1	3	1	1	1
Othello	2	2	5	2	2	2
Cassio	3	5	1	3	4	3
Desdemona	4	3	6	4	3	4
Emilia	5	4	8	5	5	5
Roderigo	6	6	7	6	6	6
Montano	7	8	4	7	8	7
Lodovico	9	7	13	8	7	8
Duke	8	11	9	11	11	9
Gratiano	11	9	16	9	9	10
Brabantio	10	10	15	10	10	11
first senator	12	13	9	13	12	12
second senator	13	16	11	15	14	13
Bianca	14	12	18	12	13	13
second gentleman	17	15	12	14	15	15
Clown	15	14	14	16	15	16
Herald	16	18	1	24	17	17
first officer	19	17	17	21	18	18
third gentleman	22	19	18	17	19	19
first gentleman	21	20	18	18	21	20
first musician	18	23	18	20	19	20
messenger	20	24	18	22	21	22
fourth gentleman	24	21	18	19	24	23
sailor	23	25	18	23	23	24
Gentleman	25	22	18	25	25	25

In questa prima parte del lavoro abbiamo deciso anche di considerare l'aspetto temporale e dinamico, dato dal susseguirsi delle quindici scene nell'opera, ma non come fatto in [20]. Infatti le scene non le consideriamo indipendenti le une dalle altre, ma utilizziamo un metodo cumulativo: analizziamo il testo fino a un punto preciso, cioè fino alla fine di ogni scena, e ripetiamo ciò che abbiamo fatto per il testo integrale descritto a inizio sezione. In concreto consideriamo la prima scena e lì stiliamo la classifica come fatto precedentemente, poi facciamo lo stesso prendendo tutto il testo fino alla fine della seconda scena, successivamente fino alla terza e così fino a considerare il testo in tutta la sua integrità. Per questa parte abbiamo deciso di considerare i sei personaggi più importanti in base alla classifica stilata nella tabella 5.2, cioè Othello, Iago, Roderigo, Desdemona, Cassio ed Emilia. Il loro andamento è simile in quasi tutti i campo semantici, con Othello e Iago che si contendono il primo posto, Cassio e Desdemona che si stabilizzano nelle loro posizioni nel concludersi del terzo atto, mentre Roderigo perde posizioni passando alla quarta o quinta posizione ed Emilia, che sembra 'secondaria' per gran parte dell'opera, scala la classifica dalla fine del quarto atto. Un esempio è il grafico nella figura 5.1 che rappresenta l'importanza dei personaggi considerando il *topic* Virtue, oppure quello in figura 5.2. Nelle ascisse troviamo la lunghezza del testo, mentre nelle ordinate la posizione nella classifica. Gli asterischi quindi rappresentano la posizione di ogni personaggio alla fine delle varie scene, punti in cui abbiamo stilato le classifiche. Notiamo che questi non sono equistanziati, questo perchè ogni scena ha una lunghezza diversa rispetto alle altre; in ultimo, le linee verticali rappresentano la fine di un atto.

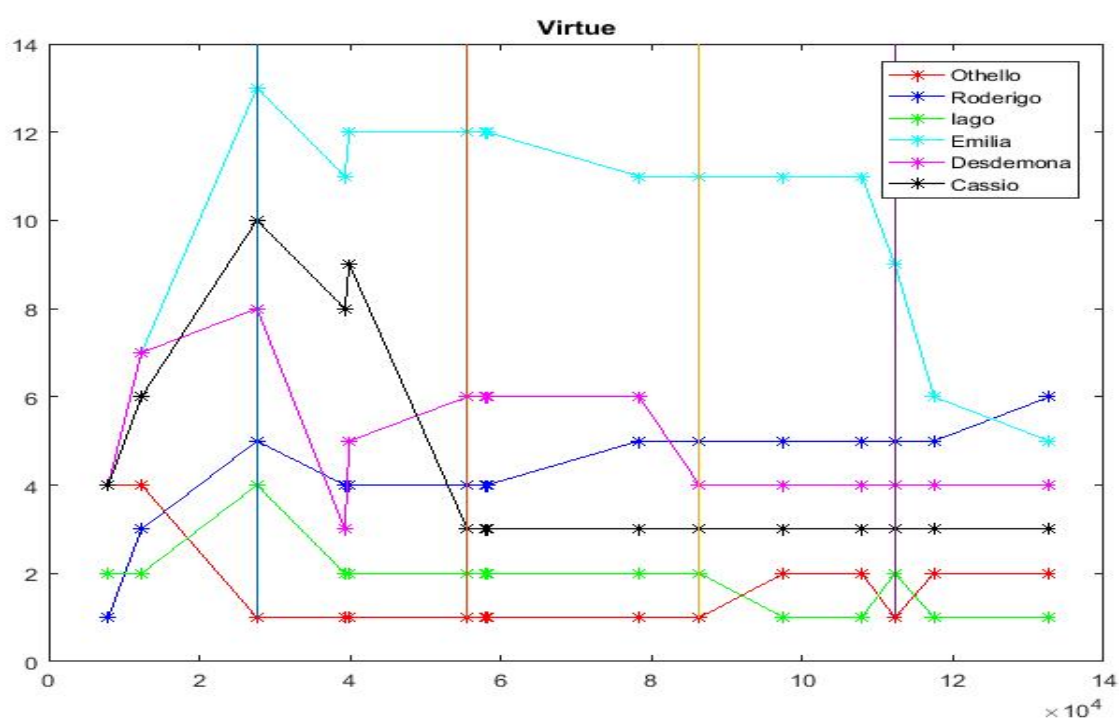


Figura 5.1: Posizioni temporali dei personaggi considerando il campo semantico Virtue

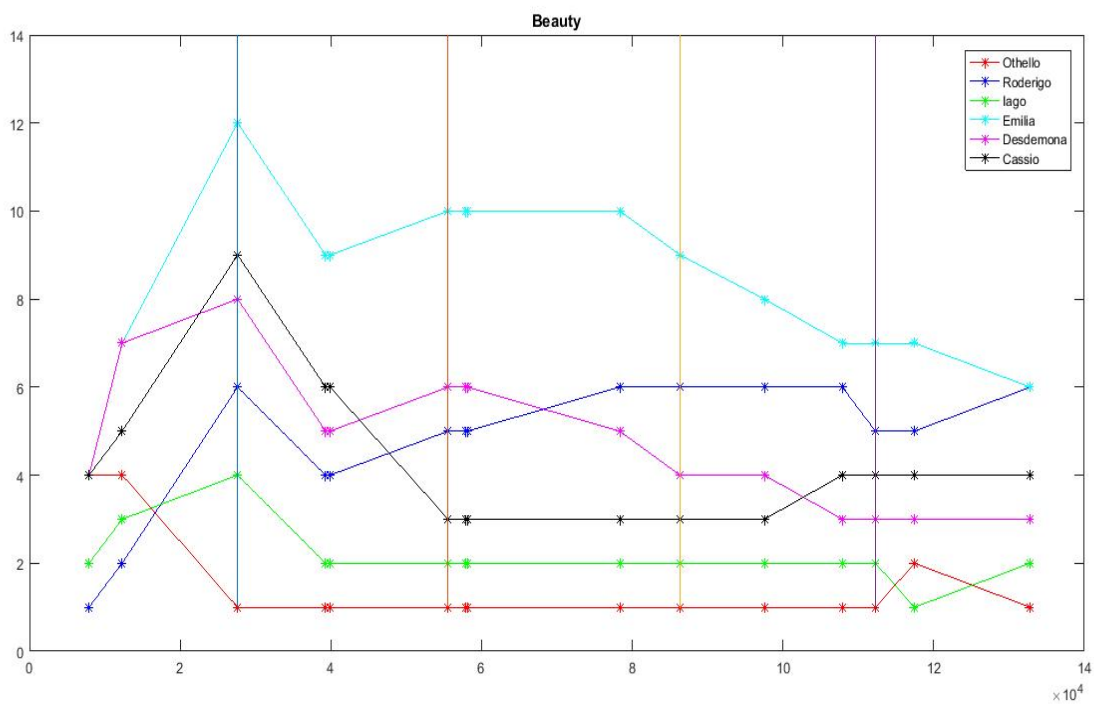


Figura 5.2: Posizioni temporali dei personaggi considerando il campo semantico Beauty

Interessante notare come le posizioni nel *topic* Death siano abbastanza stabili per tutto il testo, mentre nell'ultimo atto si assista a vari cambiamenti, vedi figura 5.3. Effettivamente nelle ultime due scene, che compongono l'atto finale, si assiste a una serie di eventi sanguinosi e la morte, rimasta in secondo piano in tutta l'opera, diventa un argomento di primo piano, con Cassio, vittima di un tentato omicidio, suo maggiore 'esponente'.

Anche per quanto riguarda il campo semantico Truth si ha qualcosa di interessante. Emilia, che non si trova in nessun *topic* nelle prime posizioni, quando si parla di truth diventa una dei tre personaggi principali, vedi figura 5.4. Effettivamente Emilia viene ingannata, raggirata da Iago e lo scopre nell'ultimo atto, motivo per cui diventa predominante nei suoi discorsi la 'verità'.

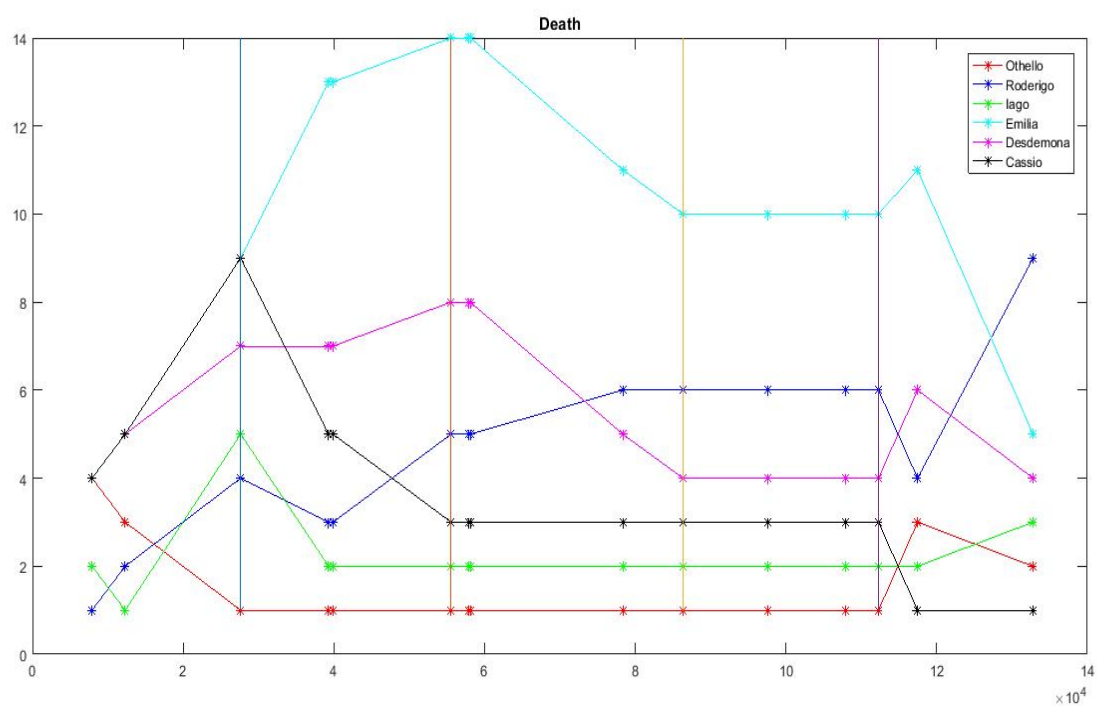


Figura 5.3: Posizioni temporali dei personaggi considerando il campo semantico Death

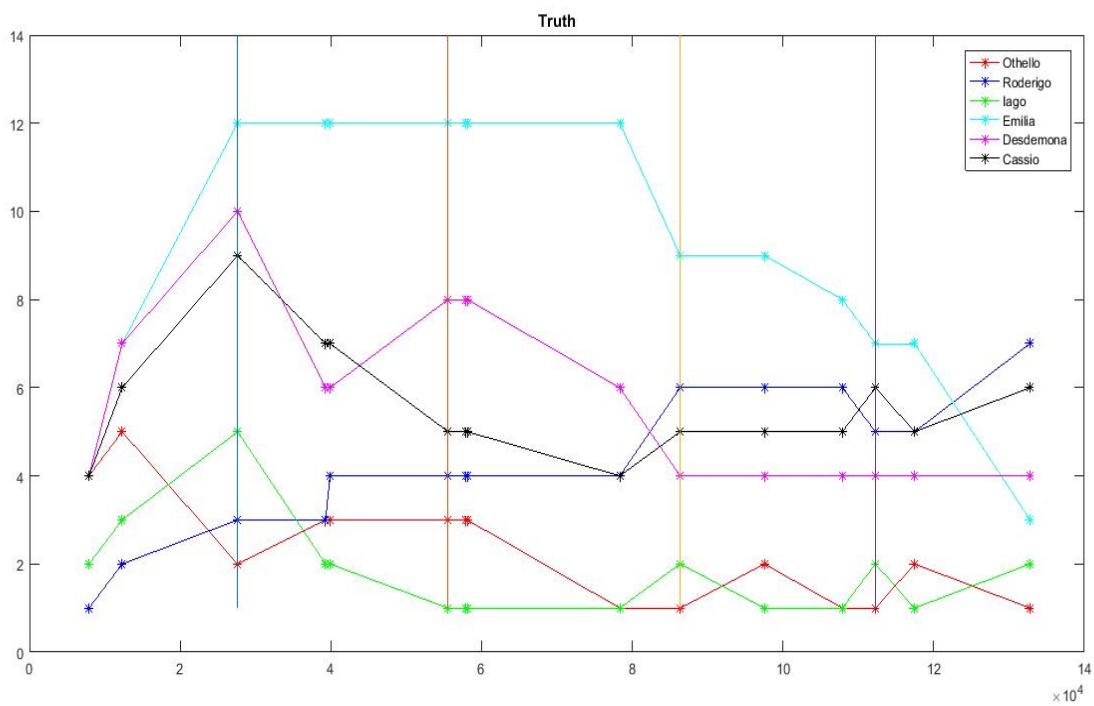


Figura 5.4: Posizioni temporali dei personaggi considerando il campo semantico Truth

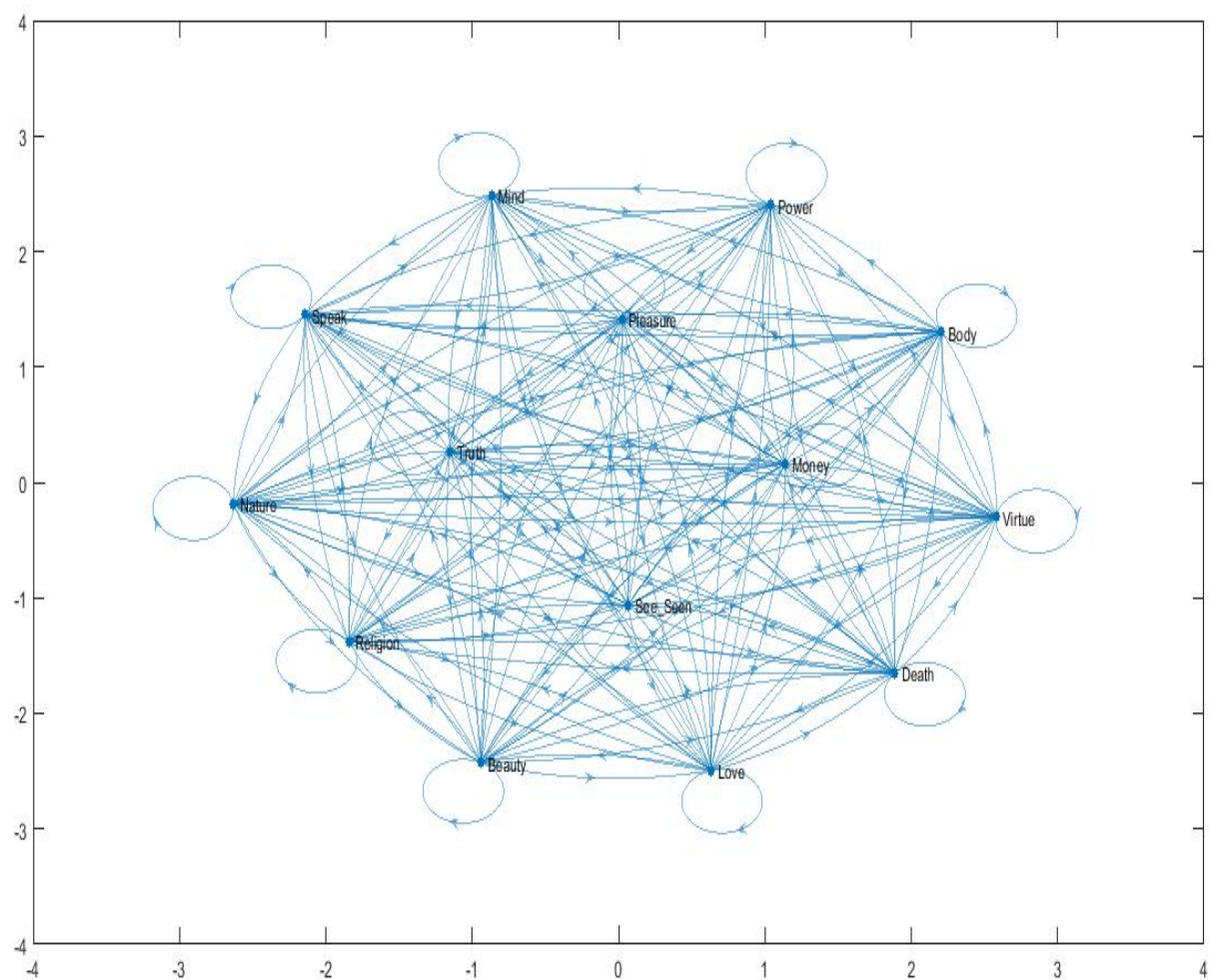
5.2 Centralità dei *topics*

Nella seconda parte abbiamo considerato solo i dialoghi con almeno uno dei sei personaggi più importanti citati precedentemente. Abbiamo quindi cercato di capire quali fossero i campi semantici più importanti in quest'opera, cosa tratta, quali argomenti affronta prevalentemente. In che modo? Abbiamo creato un grafo diretto questa volta con i *topics* come nodi. Un nodo i è collegato al nodo j se individuata una parola chiave di i nel testo se ne trova una di j nei successivi 2000 caratteri. Inoltre abbiamo inserito un peso per fare in modo che collegamenti con distanza ravvicinata avessero più importanza di quelli con distanze più ampie. Sia $dist(p_1, p_2)$ la distanza fra due parole chiave p_1 e p_2 rispettivamente di i e j , intesa come numero di caratteri che le separano, definiamo $P(p_1, p_2) = \frac{1}{\sqrt{dist(p_1, p_2)}}$. Per ogni parola chiave p_1 del campo semantico i trovata nel testo dell'opera, abbiamo calcolato $w(p_1, j) = \sum_{p_2} P(p_1, p_2)$ con p_2 parola chiave del *topic* j che segue p_1 nei limiti caratteriali fissati precedentemente. Infine abbiamo definito il peso dello spigolo diretto dal nodo i al nodo j come la somma dei $w(p_1, j)$ sulle parole p_1 del *topic* i trovate nel testo, cioè $\omega(i, j) = \sum_{p_1} w(p_1, j)$.

La direzione scelta ha senso, anche se avremmo potuto invertirla senza ottenere significativi cambiamenti ai risultati. Infatti, sia $A = \frac{A+A^T}{2} + \frac{A-A^T}{2}$ la matrice di adiacenza del grafo. Le due frazioni rappresentano rispettivamente la parte simmetrica di A , $P_s(A)$, e la parte asimmetrica $P_{as}(A)$. Calcolando le norme in $L^2(\mathbb{R}^{14 \times 14})$ e facendone il rapporto otteniamo che $\frac{\|P_{as}(A)\|}{\|P_s(A)\|} \approx 2 \cdot 10^{-2}$, quindi la parte simmetrica ha molto più peso di quella asimmetrica.

Anche in questa seconda parte del lavoro abbiamo utilizzato Python per leggere il testo e generare le matrici di adiacenza dei grafi.

Notiamo dall'immagine del grafo (Figura 5.5) come ogni *topic* sia collegato con gli altri e con se stesso. Da questo segue che alcune misure di centralità utilizzate nel grafo non pesato, come la betweenness, o la incloseness o ancora la outcloseness perdono di senso.

Figura 5.5: Grafo dei *topics*

Abbiamo quindi calcolato varie misure di centralità, come indegree, l'out-degree, l'hubs, l'authority e pagerank. Proprio come nella prima parte abbiamo stilato per ognuna una classifica e poi tramite una semplice media ottenuto una classifica generale. Anche in questo caso, abbiamo utilizzato Matlab per effettuare queste operazioni.

Tabella 5.3: Ranking *topics*

Name	$R_{indegree}$	$R_{outdegree}$	$R_{pagerank}$	R_{hubs}	$R_{authorities}$	R_{media}
Virtue	1	1	1	1	1	1
Mind	2	2	2	2	2	2
Power	3	3	3	3	3	3
Religion	4	4	4	4	4	4
Body	5	6	5	5	5	5
Love	6	5	6	6	6	6
Truth	7	7	7	7	7	7
Nature	8	8	8	8	8	8
Pleasure	9	9	9	9	9	9
Speak	10	10	10	10	10	10
See_Seen	11	11	11	11	11	11
Beauty	12	12	12	12	12	12
Death	13	13	13	13	13	13
Money	14	14	14	14	14	14

Come si evince dalla tabella 5.3, per qualunque misura usata, la classifica è sempre la medesima. Da questi risultati si potrebbe concludere che *The Tragedy of Othello, Moor of Venice* è un'opera che tratta soprattutto la virtù e di ciò che la mente costruisce e ricostruisce e non dell'amore o della morte. Può essere utile osservare l'istogramma (figura 5.6) della matrice di adiacenza pesata del grafo, perchè fa capire visivamente la differenza fra i *topics* nelle prime posizioni e quelli nelle ultime.

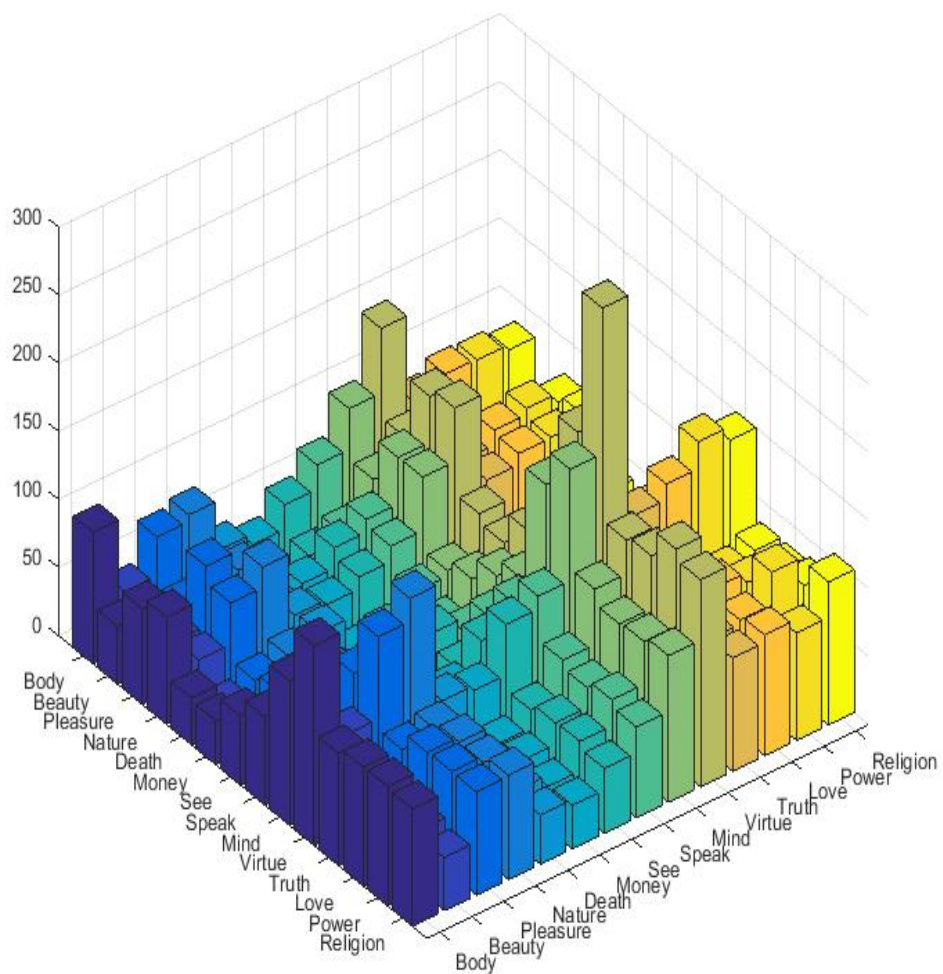


Figura 5.6: Istogramma della matrice di adiacenza del grafo 5.5

Questo è quello che succede se si guarda il testo in tutta la sua integrità, senza considerare l'aspetto temporale ed evolutivo dell'opera. Per questo, proprio come nella prima parte, con il procedimento già descritto ad inizio sezione, abbiamo calcolato alla fine di ogni scena la classifica dei topics. Queste classifiche, se osservate tutte insieme, ci descrivono le posizioni temporali dei vari campi semantici (figura 5.7).

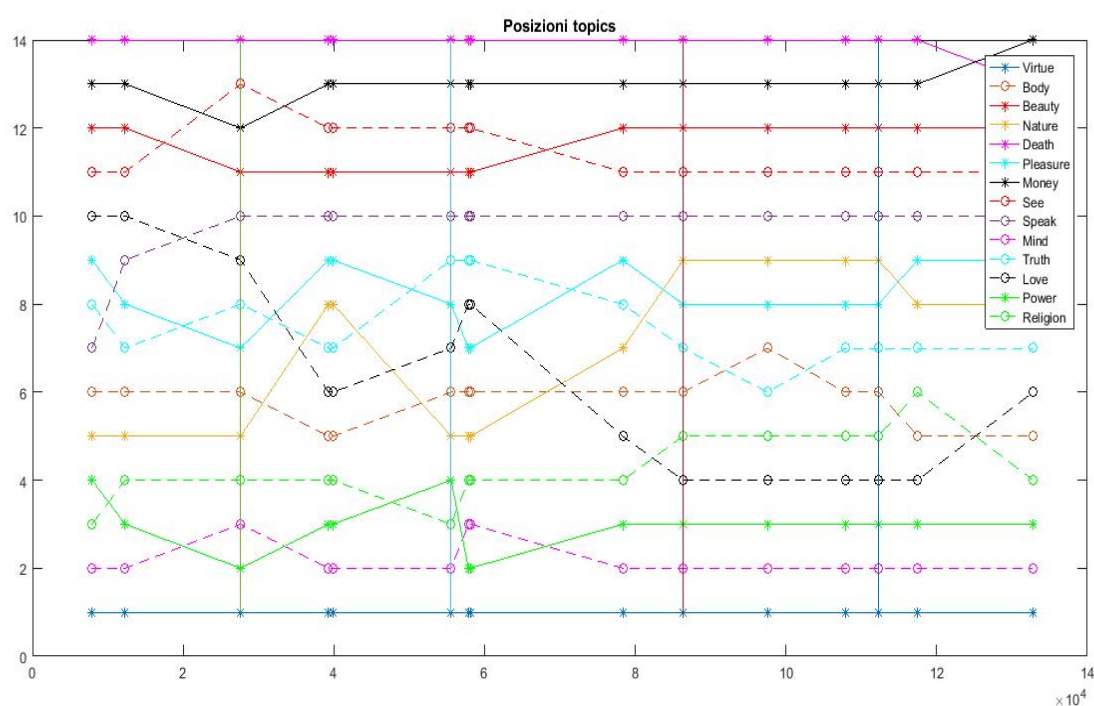


Figura 5.7: Classifica dinamica

Questa classifica che considera la componente temporale, ci fa capire come sia predominante in tutta l'opera il *topic* Virtue. Anche Mind si attesta fin da subito nelle prime posizioni, così come Money, Death, See, Speak si confermano campi semantici proprio poco importanti. I *topics* centrali hanno invece risultati molto simili. Quindi si potrebbe dedurre la presenza di campi semantici fondamentali, come Virtue e Mind, campi semantici importanti, di cui fanno parte love, body e tutti gli altri *topics* che si attestano nel centro

classifica, campi semantici secondari, come See, Speak, Death e Money.

Tutto questo lavoro abbiamo pensato di farlo anche per i sei personaggi principali, già citati nelle precedenti sezioni. Quali sono per loro i *topics* più importanti? Come cambiano nel corso dell'opera?

Abbiamo quindi generato un grafo per ogni personaggio, dove i nodi rappresentano i campi semantici e questi vengono collegati con uno spigolo orientato da i a j se il personaggio, a cui si riferisce il grafo, parla del *topic* j entro i 2000 caratteri successivi a una parola di i . Non ci importa chi abbia parlato di i , ma ci importa di chi parla, e di che cosa, dopo i . Il peso di questo spigolo viene calcolato con la procedura descritta all'inizio della sezione.

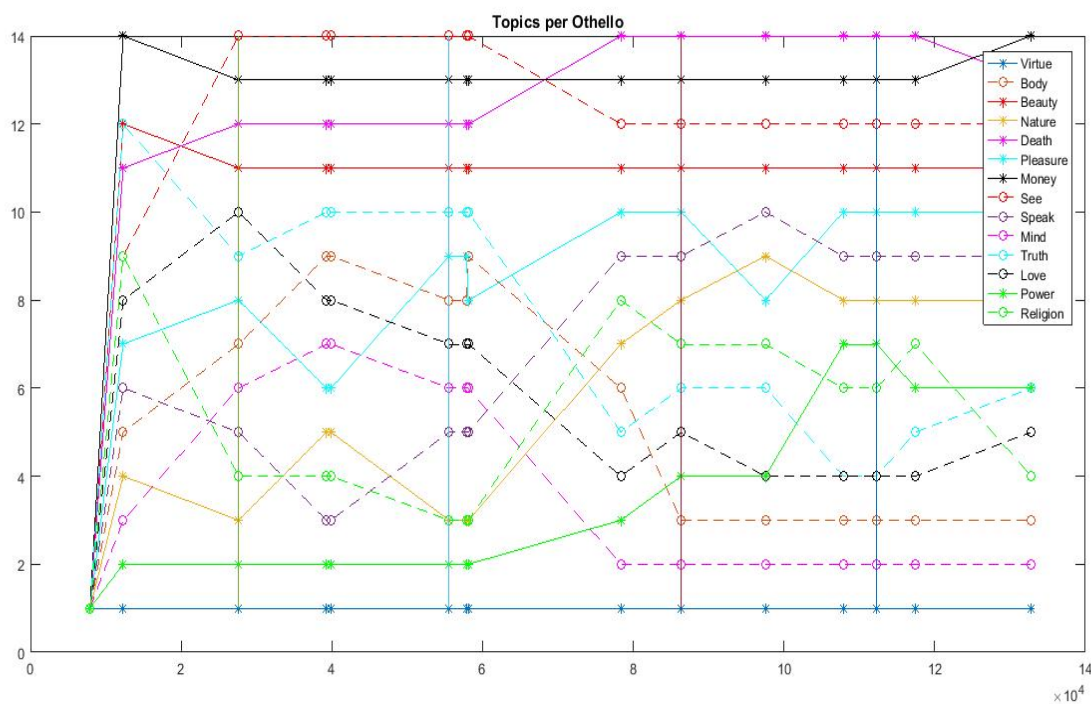


Figura 5.8: Classifica dinamica per Othello

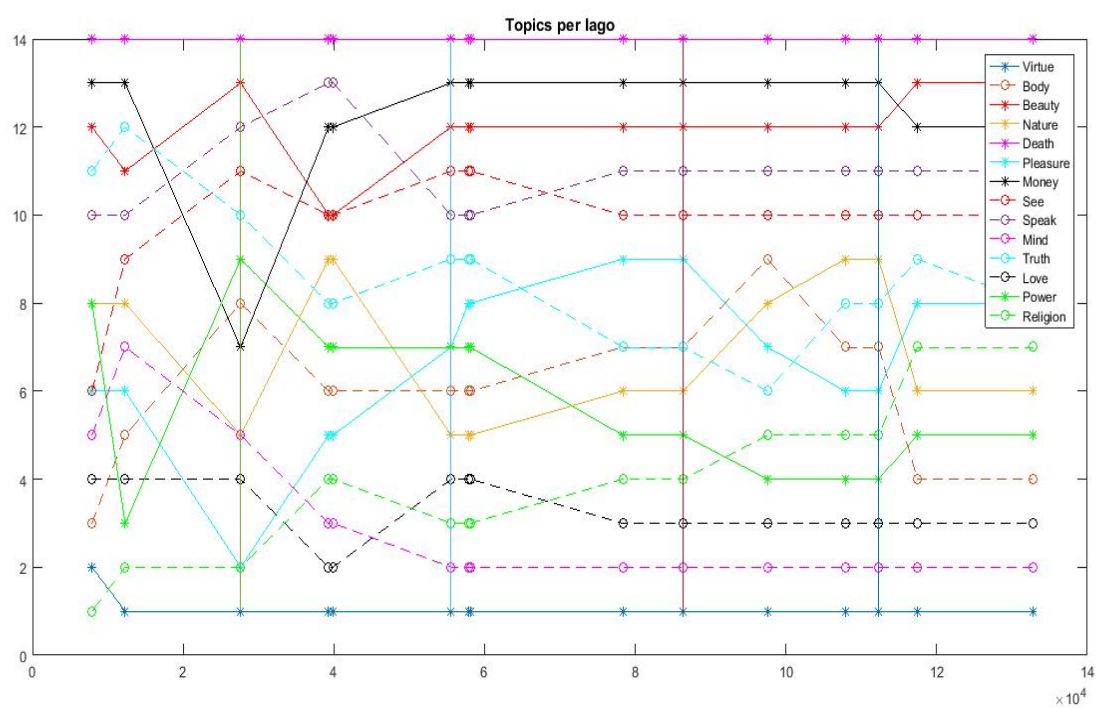


Figura 5.9: Classifica dinamica per Iago

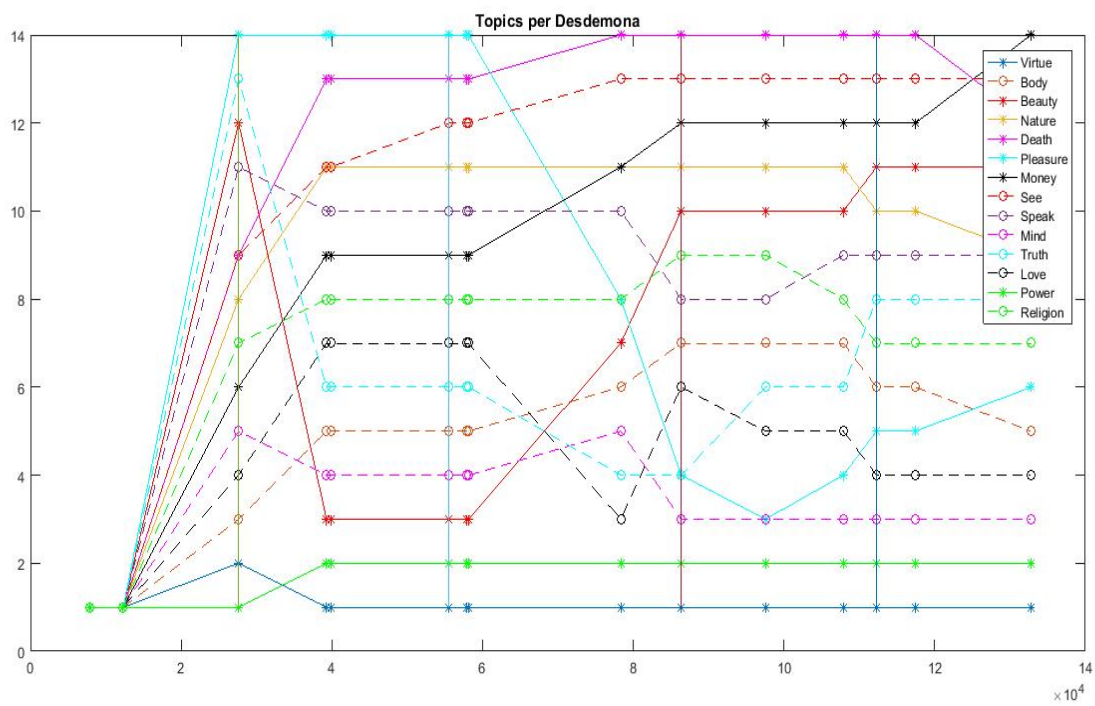


Figura 5.10: Classifica dinamica per Desdemona

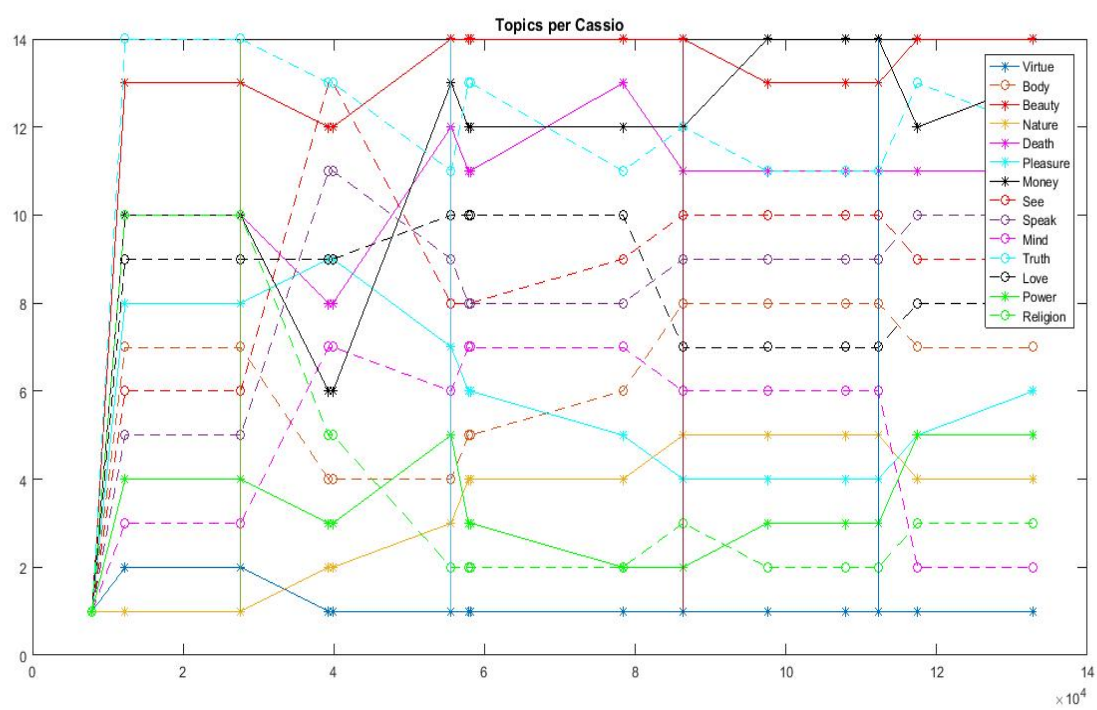


Figura 5.11: Classifica dinamica per Cassio

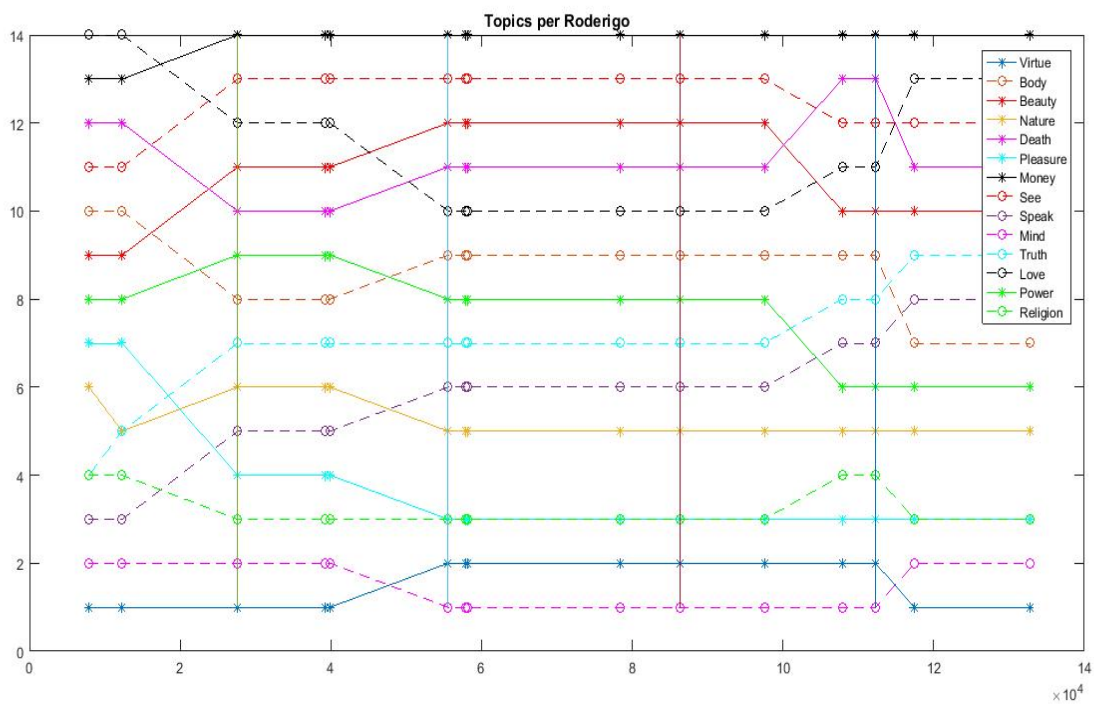


Figura 5.12: Classifica dinamica per Roderigo

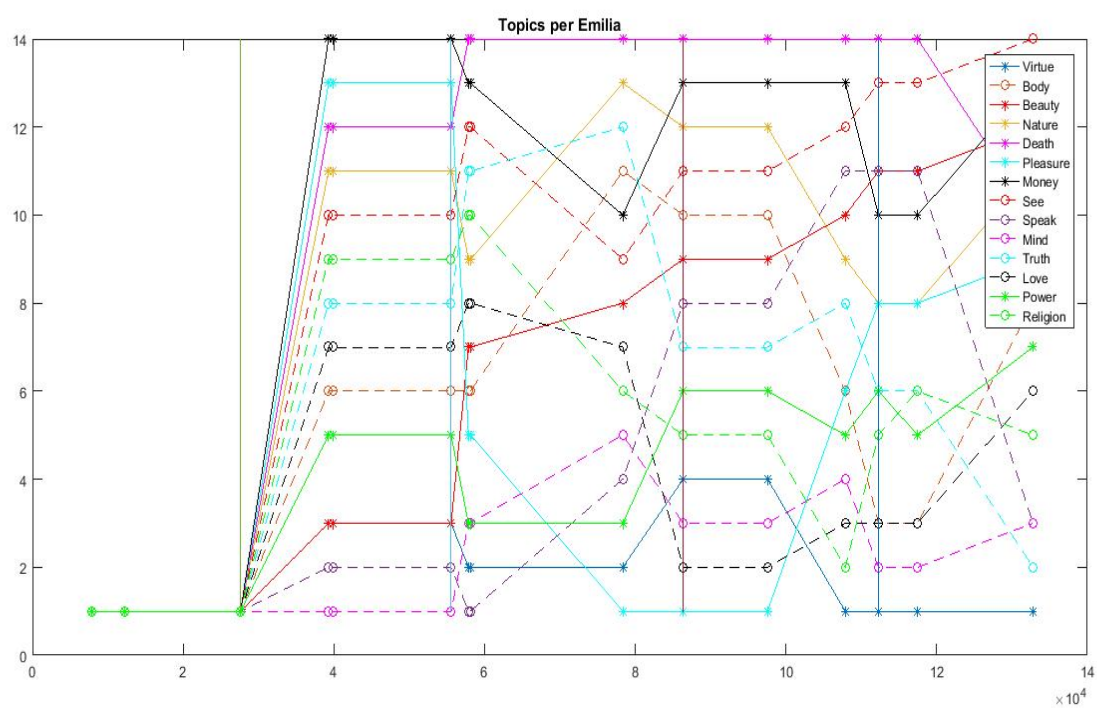


Figura 5.13: Classifica dinamica per Emilia

Vediamo come i *topics* fondamentali lo rimangano anche per i vari personaggi. Da notare come Power sia fondamentale per Desdemona, infatti essa è, potremmo dire, sottomessa ad Othello e nei loro dialoghi è sempre presente un rapporto di potere nel quale Othello è la figura predominante e Desdemona è quella sottomessa.

La classifica non è invece stabile per Emilia, questo perchè lei è il personaggio meno importante dei sei fondamentali e come avevamo già visto solo nelle ultime scene si affermava la sua importanza. Notiamo che proprio nell'ultima scena il *topic* Truth ha un exploit attestandosi al secondo posto, risultato coerente con quanto visto nella sezione precedente, dove Emilia nell'ultimo capitolo scalava le posizioni proprio in quel campo semantico diventando un personaggio fondamentale su questo argomento.

Una domanda lecita da porsi è quanto influiscono i personaggi secondari rispetto a questi sei fondamentali nel porre l'importanza sui vari *topics*.

Per rispondere a questa domanda abbiamo sommato fra loro le matrici di adiacenza di ognuno dei sei personaggi e abbiamo sottratto alla matrice di adiacenza del grafo dei *topics* la matrice somma, cioè $A_{topics} - \sum_{i=1}^6 A_i$ con A_i la matrice di adiacenza del grafo dei *topics* per il personaggio fondamentale i -esimo. La matrice così ottenuta rappresenta il grafo dei *topics* dei personaggi secondari e facendo il rapporto fra le norme in $L^2(\mathbb{R}^{14 \times 14})$ di ques'ultima e di quella dei personaggi fondamentali otteniamo 0.1232, che ci fa capire come gli altri personaggi abbiano molto poco peso.

I nostri risultati per avere la speranza di essere significativi, devono essere il più possibile indipendenti sia dalle scelte personali e arbitrarie dell'esaminatore; sia di tipo semantico che in termini di dettagli matematici. Un esempio che riguarda entrambe le questioni è la cardinalità delle parole chiavi dei diversi campi semantici. Ogni *topic* ha infatti un diverso numero di termini che lo rappresenta, ad esempio Virtue ne ha 142, più di ogni altro *topic*, mentre Death ne ha il numero minore, cioè solo 24. La correttezza di tali liste è un problema linguistico/letterario, non matematico ed esula dallo scopo di questa tesi che è mostrare invece come grazie alla teoria dei grafi si

possano capire, non solo quali siano i personaggi più importanti, ma anche quali siano i temi affrontati in modo predominante e da chi. Ovviamente più si cerca di raffinare le ricerche, più ci si scontra con problemi multidisciplinari, come il precedente. Nonostante questo, possiamo variare i parametri per studiare la stabilità dei risultati ottenuti e rafforzarne la significatività, ad esempio cercando di uniformare la cardinalità delle parole chiave dei diversi campi semantici.

Questa breve parentesi su questo problema ci introduce nell'ultima parte del lavoro effettuato. Per testare, rispetto alla quantità di termini di ogni categoria, i risultati ottenuti si sono effettuate due controprove: la prima consiste nel raggruppare *topics* compatibili in macroaree più grandi, ottenendo liste di parole più omogenee e fedeli alle iniziali. La seconda consiste in una riduzione delle parole chiave, usando solo le prime venti per frequenza nel testo per ogni *topics*. In questo modo si risolve il problema della omogeneità delle liste visto che ogni campo semantico ha lo stesso numero di rappresentanti, anche se con questo metodo si ignorano una moltitudine di parole presenti nel testo.

Per quanto riguarda la prima controprova, il professore Coronato ha ideato sei macroaree: *Nature*, con 122 parole chiave associate, *Body-Life-Death*, con 101 termini relativi, *Power-Money*, con 124 parole chiave; *Beauty-Pleasure-Love*, con 135, *Mind-Truth-Sight-Seeming*, con 134 ed infine *Virtue-Religion* con 146. In questi raggruppamenti non ha trovato spazio il *topic* *Speak*, che abbiamo quindi tralasciato, poichè non accorpabile agli altri campi semantici e con solo una trentina di parole chiave. Si nota come le differenze di cardinalità siano diminuite. Con queste nuove macroaree semantiche, abbiamo ripetuto il proedimento spiegato ad inizio sezione, ottenendo una classifica molto stabile che risalta la divisione fra campi semantici fondamentali e secondari. Osservando infatti i valori numerici nella tabella 5.4 e la classifica temporale nel grafico 5.14, possiamo notare come le macroaree *Mind-Truth-see* e *Virtue-Religion* siano saldi nelle prime due posizioni per tutto il corso dell'opera con i rispettivi valori delle misure di centralità ben

più alti rispetto a quelli degli altri.

La macroarea Beauty-Pleasure-Love è si attesta presto al terzo posto, con molto distacco rispetto a Body-Death e Money-Power, i quali hanno valori numeri praticamente uguali. In fondo alla classifica troviamo Nature, che non è stata accorpata ad altri *topics* e ne risente moltissimo.

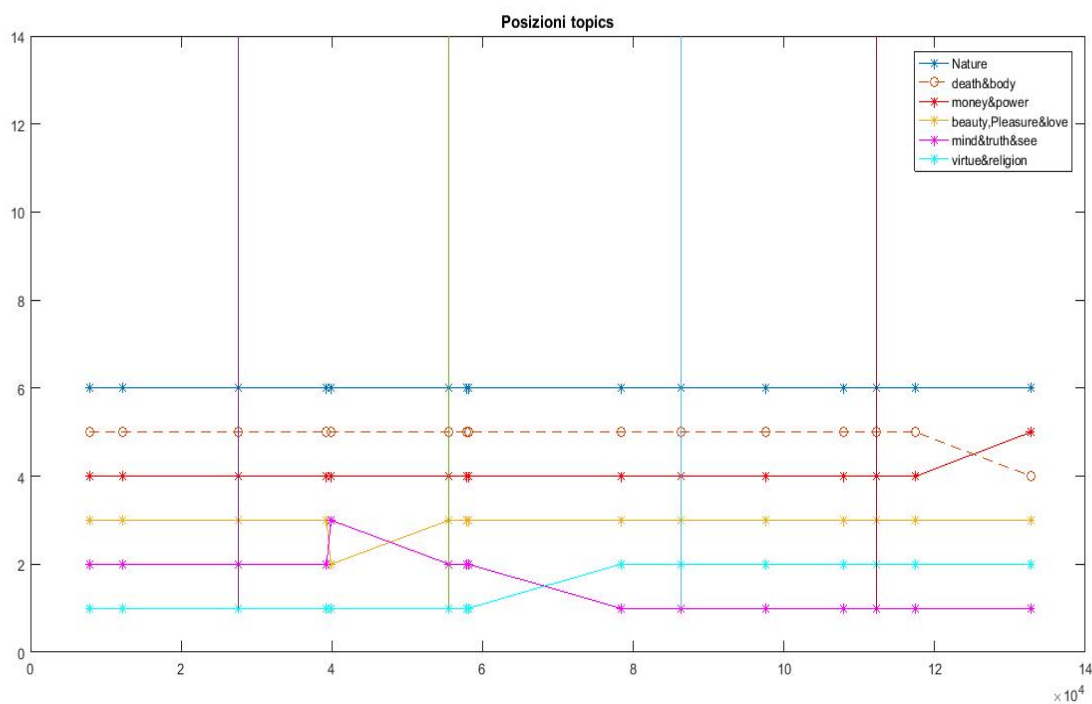


Figura 5.14: Classifica temporale per macroaree

Tabella 5.4: Tabella valori numerici finali della prima controprova

Name	indegree	outdegree	hubs	authoritiess	pagerank
nature	548,6499	535,8060	0,0584	0,0601	0,0770
death_body	1042,8791	1019,2438	0,1110	0,1132	0,1239
money_power	1012,1233	1032,5417	0,1139	0,1119	0,1204
beauty_Pleasure_love	1986,5153	1971,1487	0,2194	0,2208	0,2123
mind_truth_see	2260,3195	2273,6945	0,2550	0,2543	0,2375
virtue_religion	2159,9052	2177,9575	0,2422	0,2397	0,2288

Per quanto riguarda la seconda controprova, i risultati, vedi figura 5.15, non stravolgono la classifica temporale in figura 5.7. Mind ottiene la prima posizione dalla fine del terzo atto a discapito di virtue, che rimane comunque al secondo posto. Questo dato è molto interessante ed è a riprova che non sono le prove visibili o le questioni religiose, ma ciò che la mente costruisce e ricostruisce a essere preminente. Per una tragedia legata al visibile, che include anche il colore della pelle, non è un risultato scontato. Inoltre da questa prova si ha la conferma che i campi semantici fondamentali sono questi ultimi.

Ci sono altri cambi, come Power che scende quarto posto, sorpassato da Religion, oppure Pleasure e Love che hanno dei miglioramenti a discapito ad esempio di Body e di Nature, ma, nonostante ciò, le fasce che avevamo individuato persistono. Questo si capisce bene osservando i valori numerici nella tabella 5.5: Mind e Virtue occupano saldamente le prime due posizioni, mentre Religion, Power e Love hanno valori molto simili, formando una seconda fascia di importanza. A seguire Pleasure, Speak, Body e Truth a formare un'altra fascia con valori molto simili, per terminare con Nature e See davanti all'ultima fascia composta da Death Beauty e Money, con valori estremamente bassi.

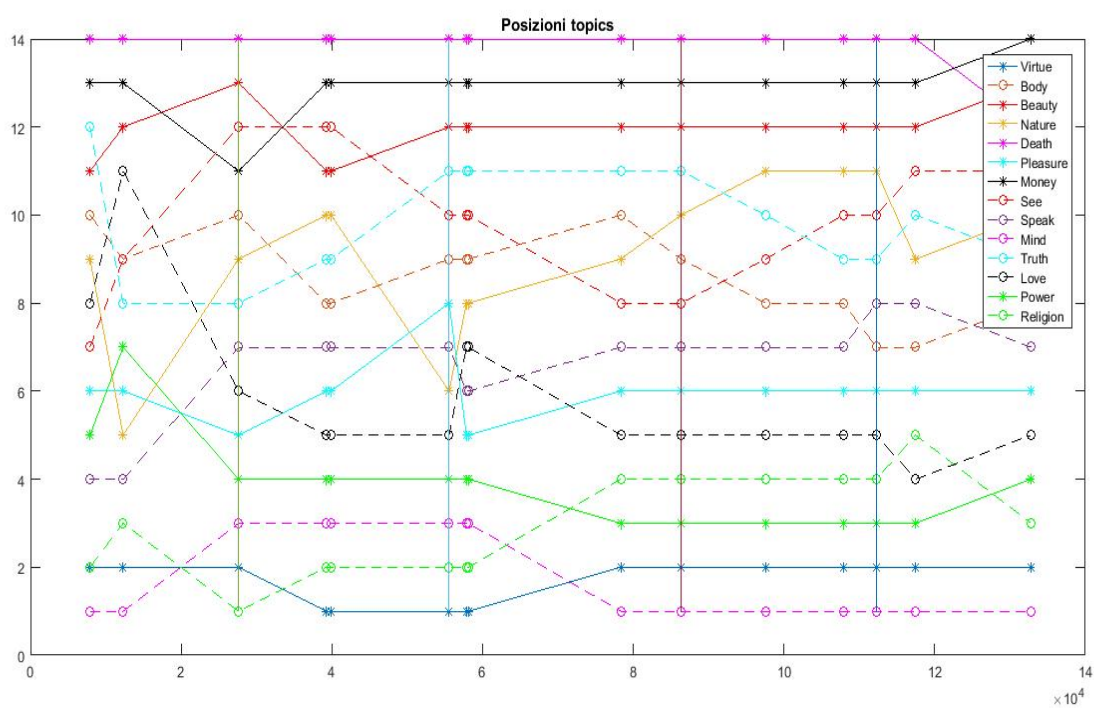


Figura 5.15: Classifica temporale della seconda controprova

Tabella 5.5: Tabella valori numerici finali della seconda controprova

Name	indegree	outdegree	hubs	authoritiess	pagerank
Mind	966,6713	980,9208	0,1093	0,1079	0,1020
Virtue	901,8239	901,3522	0,1007	0,1006	0,0958
Religion	828,1528	842,2015	0,0943	0,0921	0,0889
Power	792,1508	808,6169	0,0900	0,0885	0,0854
Love	783,9236	774,2523	0,0860	0,0869	0,0848
Pleasure	681,3749	677,5800	0,0752	0,0759	0,0750
Speak	668,3452	676,7348	0,0755	0,0740	0,0738
Body	653,5914	633,9827	0,0699	0,0726	0,0724
Truth	616,8397	617,9070	0,0682	0,0685	0,0689
Nature	547,2782	544,3812	0,0603	0,0610	0,0624
See_Seen	522,0067	508,7559	0,0568	0,0580	0,0600
Death	379,5955	367,5438	0,0399	0,0411	0,0468
Beauty	353,6871	349,2670	0,0389	0,0391	0,0442
Money	306,1900	318,1350	0,0352	0,0338	0,0397

Quello che si riesce a distinguere in modo netto e in tutte le prove che abbiamo fatto sono delle fasce di importanza. All'interno di esse i valori sono molto simili e suscettibili a rumore o cambiamenti, per cui si deve diffidare da una classifica minuziosa, mentre ci si può basare su queste fasce perchè paiono stabili e con differenze di valori più accentuate.

Un altro parametro da cui dipendono tutti i risultati precedente è la lunghezza dell'intervallo nel quale si cercavano le co-occorrenze di parole. Avevamo fissato questo valore a 2000 caratteri. Per rafforzare ulteriormente i risultati e renderli più consistenti, abbiamo deciso di variare anche questo parametro, considerando i campi semantici della seconda controprova, cioè ognuno con 20 parole chiave.

Abbiamo quindi diminuito i caratteri a 200, poi a 100 ed infine a 50. In questa riduzione speravamo di trovare un grafo diverso dal grafo 5.5, quello dove ogni nodo si ritrova collegato a tutti gli altri, ma anche usando 50 caratteri, prima dell'inizio del secondo atto, ci si ritrovava con il medesimo grafo.

Nell'applicare questi tagli si osserva che le fasce rimangono ben distinte, anche se all'interno di esse diminuisce la distanza dei valori numerici dei singoli *topics*. Riportiamo la classifica, figura 5.16, e i valori numerici, tabella 5.6, solo del caso dell'intervallo lungo 100 caratteri a titolo d'esempio, mentre per gli altri rimandiamo all'appendice.

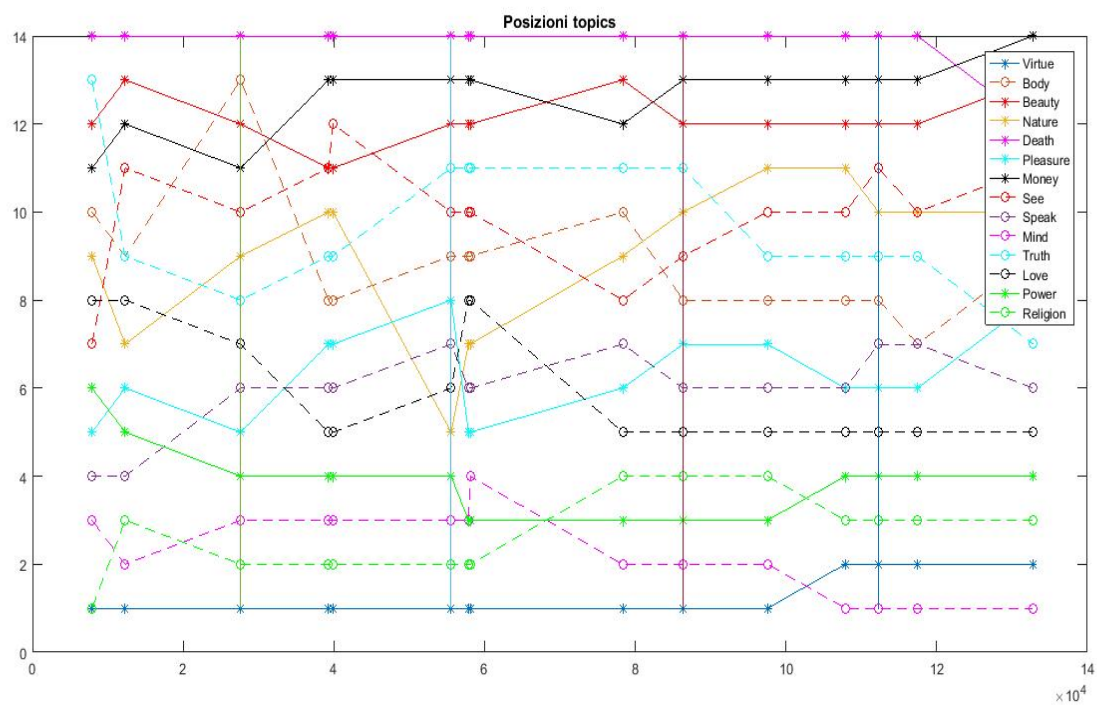


Figura 5.16: Grafico dopo la riduzione a 100 caratteri

Tabella 5.6: Valori numerici con riduzione a 100 caratteri

Name	indegree	outdegree	hubs	authoritiess	pagerank
Money	57,2742	55,6715	0,0321	0,0319	0,0392
Beauty	67,6000	66,7291	0,0388	0,0371	0,0445
Death	75,8274	69,8710	0,0375	0,0413	0,0485
See_Seen	99,6361	91,0788	0,0518	0,0566	0,0601
Nature	101,9734	96,5502	0,0552	0,0596	0,0609
Body	119,5425	127,0450	0,0710	0,0688	0,0691
Truth	123,3440	122,4090	0,0706	0,0718	0,0707
Pleasure	122,4113	123,9016	0,0697	0,0711	0,0708
Speak	127,2104	142,2599	0,0845	0,0731	0,0727
Love	147,4245	142,0971	0,0805	0,0836	0,0833
Power	153,2250	156,0529	0,0904	0,0908	0,0853
Religion	158,9225	166,9484	0,0985	0,0930	0,0886
Virtue	184,4324	180,3717	0,1067	0,1082	0,1010
Mind	192,2066	190,0441	0,1123	0,1130	0,1052

Conclusioni

Con la collaborazione del professore Coronato Rocco dell'Università degli Studi di Padova si è provato ad interpretare i risultati precedenti dal punto di vista letterario e qualitativo. Le riflessioni che ha scritto sono riportate nel seguito fra virgolette in stile *corsivo* da [24].

Gli aspetti più notevoli ed originali di questa tesi sono che, *invece di seguire vie già esperite quali la ricostruzioni delle comunità parlanti*, l'individuazione di personaggi chiave e minori come visto nel capitolo 4, ci si sofferma in particolare sulla costruzione dei campi semantici di *Otello*, con risultati *"in netta controtendenza rispetto al primato solitamente assegnato in sede qualitativa a considerazioni etnologiche, religiose e sociali legate alla differenza di colore fra Otello e Desdemona"*, e si analizza in chiave accumulativa e graduale la creazione delle varie classifiche di centralità. I risultati più interessanti in chiave letteraria sono essenzialmente due:

- *un'idea dinamica dell'evolversi dei campi semantici;*
- *il riconoscimento di una cesura strutturale al Terzo Atto.*

Per quanto riguarda il primo è doveroso ricordare che la ricostruzione delle comunità di parlanti è un aspetto importante, ma spesso rimane il dubbio sugli argomenti di cui parlino prevalentemente quei personaggi. *"Per un qualitativo, e sicuramente anche per un lettore/spettatore, sapere di cosa parlino o non parlino, e chi utilizzi di più alcuni argomenti rispetto ad altri, rimane un interesse fondamentale. Vedere poi non il risultato finito, ma come si costruiscano i primati delle varie misure di centralità e quindi dei campi*

semantic, e in mano (o in bocca) a quali personaggi, è poi fondamentale per lo studio dell'opera letteraria, e in particolare quella teatrale". Poter seguire, di scena in scena, e di personaggio in personaggio, i *topics* prevalenti di cui parlano, anche se non sappiamo ancora che cosa dicano a loro proposito, *"dà un'idea molto suggestiva dell'evolversi della struttura concettuale della tragedia"*. In effetti i dati raccolti non ci permettono ancora di dire con chiarezza cosa in una determinata scena Iago dica ad esempio della virtù, se non andando direttamente al testo, ma comunque ci restituiscono un quadro generale grazie al quale possiamo intuire in che punti diventano predominanti determinati campi semantic e per chi.

Da questo punto di vista il dato più suggestivo è l'incontestabile primato di Virtue e Mind, sia considerati da soli sia mediante riduzione delle *key-words*, rispetto alle altre aree come Death, Money e anche Religion *"di cui ci si aspettava il primato tenendo conto degli orientamenti prevalenti della critica shakespeariana degli ultimi decenni"*. Dal punto di vista qualitativo, *"il risultato fa supporre che per Shakespeare non siano né i fatti o le evidenze, visibili e sensibili, né i sistemi di credenze o pensieri consolidati, quali la religione, a determinare la verità e il bene e i loro rispettivi opposti, ma unicamente la mente che cerca di determinare cosa sia virtuoso, o meglio ancora la mente di un personaggio sovrastante per capacità intellettuali e retoriche, quale Iago, che condiziona quella di Otello"*. Molto interessante, rispetto a questo dato, anche le variazioni: ad esempio il fatto che Iago e Otello si rincorrono dinamicamente scambiandosi di posto, come si vede ad esempio nelle figure A.24, A.17, A.15 ed 5.1, o che Emilia nelle ultime scene porti Truth così in alto, vedi figura 5.4 e 5.13.

Per contrasto, si consideri come Power rimanga fondamentale per Desdemona, figura 5.10: *"non proprio sorprendente, per un'epoca in cui la moglie veniva considerata, anche legalmente, una proprietà del marito, al pari dei beni fisici"*.

Molto confortante che i risultati non cambino di molto variando l'intervallo di co-occorrenza lessicale, vedi figura 5.16, A.25 e A.26, aumentandone

la significatività.

In questo caso, dunque, grazie ai grafi si vede una cosa che non è evidente dal punto di vista qualitativo, ed inoltre *"aiutano un giudizio di genere sull'opera, che diventa così più prossima al teatro filosofico che a quello delle contrapposizioni etniche e sessuali. Lo stesso Shakespeare, del resto, fa dire a Hamlet che "there's nothing good or bad, but thinking makes it so". In Othello lo mette in pratica. Dal punto di vista qualitativo, ciò pone Shakespeare ancora più vicino ad autori influenzati dallo scetticismo pirroniano, quali Montaigne, di cui conosceva le opere"*.

L'altro risultato interessante dei grafi è che invece confermano, ma con nuova evidenza, *"un dato strutturale già noto, ovvero la tipica presenza, soprattutto in una tragedia, di un momento di cesura al Terzo Atto, in Othello durante e soprattutto dopo la lunghissima temptation scene"*. Il punto non è casuale: *"il Terzo Atto è nella teoria classica il momento della complicazione e dell'azione che poi porterà allo scioglimento del Quinto Atto. Shakespeare non amava molto le strutture in generale, e soprattutto quelle (neo)classiche, ma è vero che in Othello, dopo la temptation scene, l'azione è finalmente libera di scorrere, e che questa è una cosa tipica dello Shakespeare tragico (compare anche in Hamlet e in Macbeth)"*. Risale proprio al punto centrale della tragedia il primato di Mind a discapito di Virtue come si evince ad esempio dal grafico in figura 5.15, a riprova che *"non sono le prove visibili o le questioni religiose, ma ciò che la mente costruisce e ricostruisce a essere preminente. Per una tragedia legata al visibile, che include anche considerazioni tipiche dell'epoca circa il colore diverso della pelle di Otello e tutto il corredo di credenze entiche, religiose e sociali, non è un risultato scontato"*.

Un altro effetto interessante è che i grafi confermano *"il principio tipicamente elisabettiano della prevalenza di pochi personaggi, principio che in Shakespeare diventa imponente"*. Da questo punto di vista, il fatto che nelle ultime scene, le distanze sembrano assottigliarsi, con vari cambiamenti nelle classifiche, come ad esempio lo scambio di posizione fra Otello e Iago in molti *topics*, o il primato di Cassio in Death, *"punta ancora a un'intuizione*

qualitativa sul mestiere teatrale di Shakespeare, che spesso opera distribuzioni nascoste dei pesi attoriali coinvolgendo anche personaggi minori".

Un punto che si dovrà considerare forse in futuro è come interpretare gli improvvisi spostamenti in classifica: che Death diventi più prominente nel finale è ovvio, molto meno per altri fenomeni. Un punto dolente su cui lavorare riguarda il grafo dei *topics* in figura 5.5. Il fatto che ogni nodo sia connesso con gli altri, sia in entrata che in uscita, senza ruoli gerarchici chiari, *"non favorisce nuove ipotesi"*, soprattutto per un qualitativo che predilige *"grafi che offrano anche graficamente dei risultati suggestivi e interessanti"*.

Un altro sviluppo su cui puntare nel futuro è tentare di appaiare la ricostruzione delle comunità dei parlanti e la struttura concettuale dei campi semantici, *"secondo una multidimensionalità che sarebbe prossima all'esperienza della lettura di un'opera shakespeariana"*.

Appendice A

Di seguito riportiamo le tabelle e i grafici ottenuti:

A.1 Grafi personaggi per campo semantico

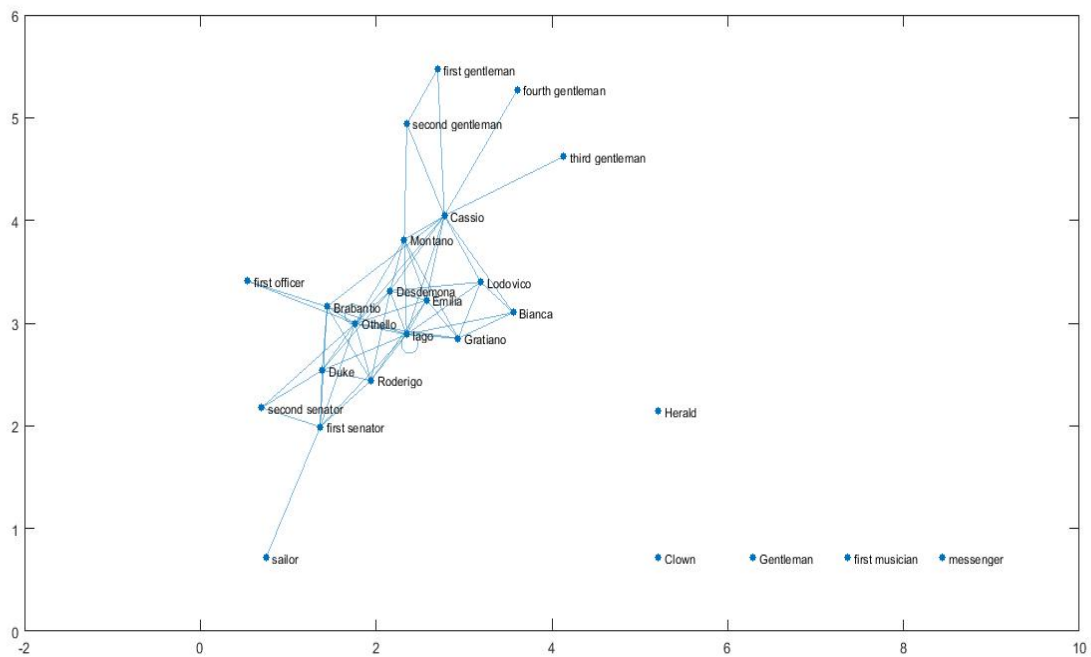


Figura A.1: Grafo Beauty

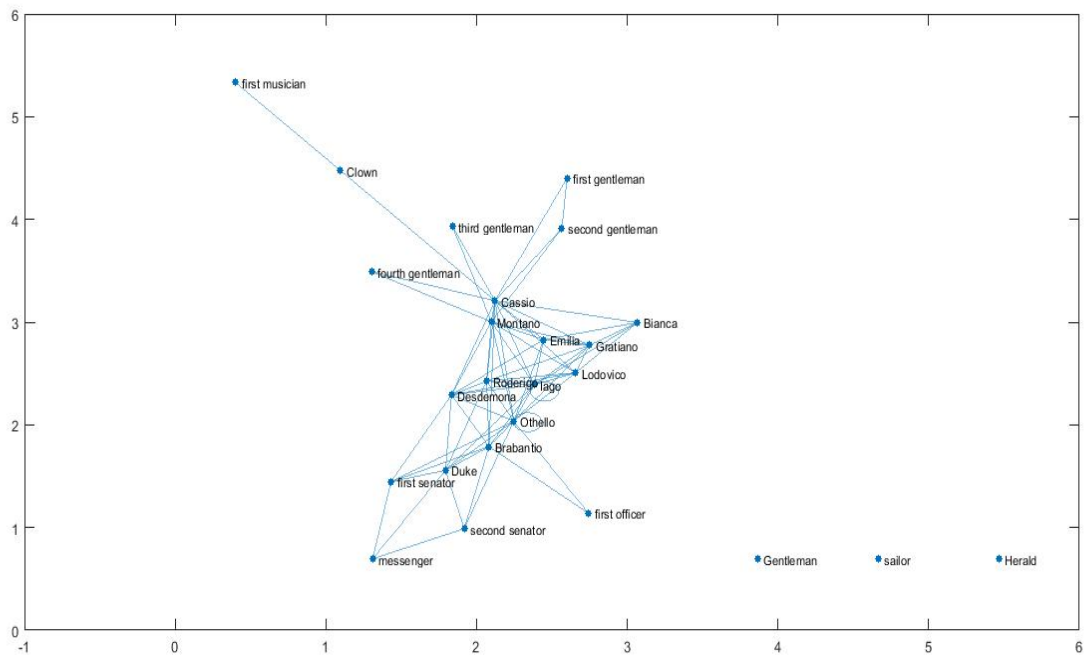


Figura A.2: Grafo Body

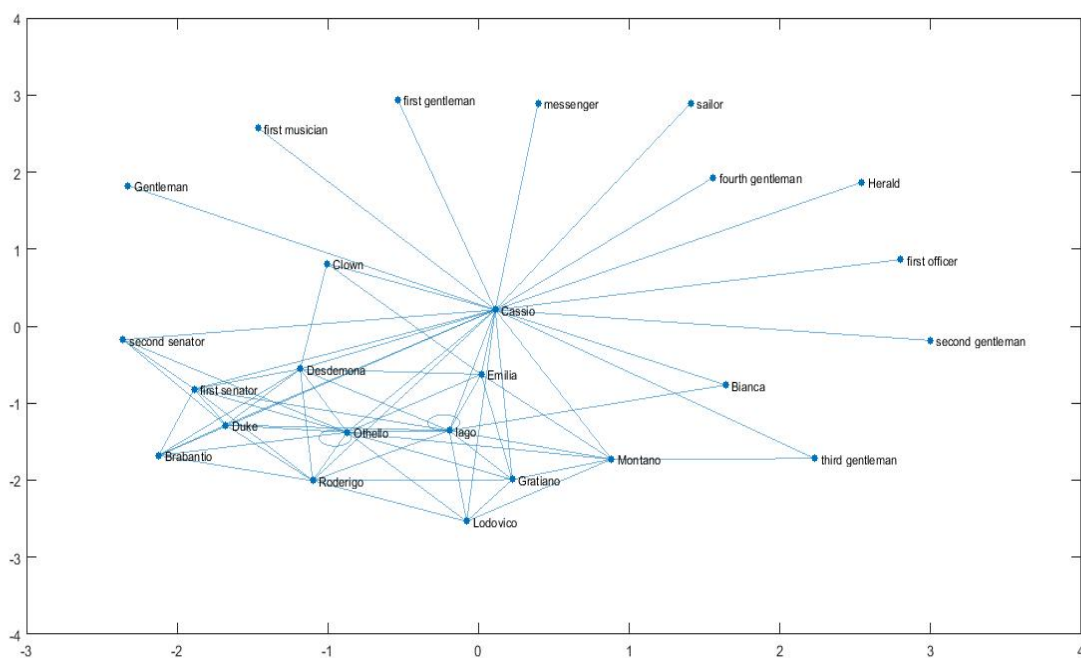


Figura A.3: Grafo Death

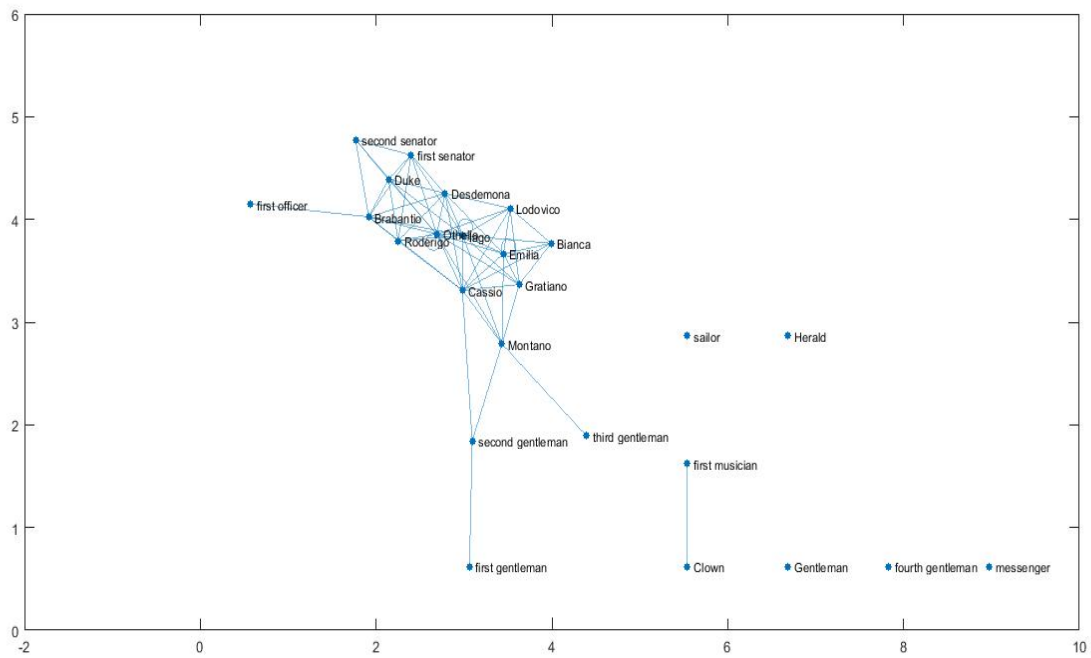


Figura A.4: Grafo Love

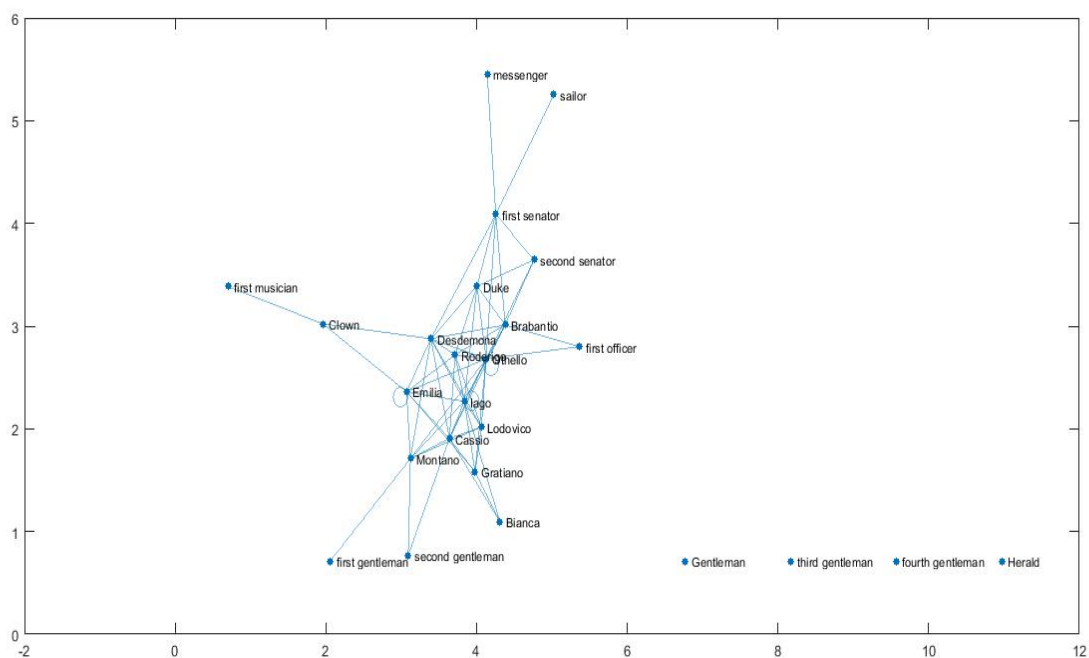


Figura A.5: Grafo Mind

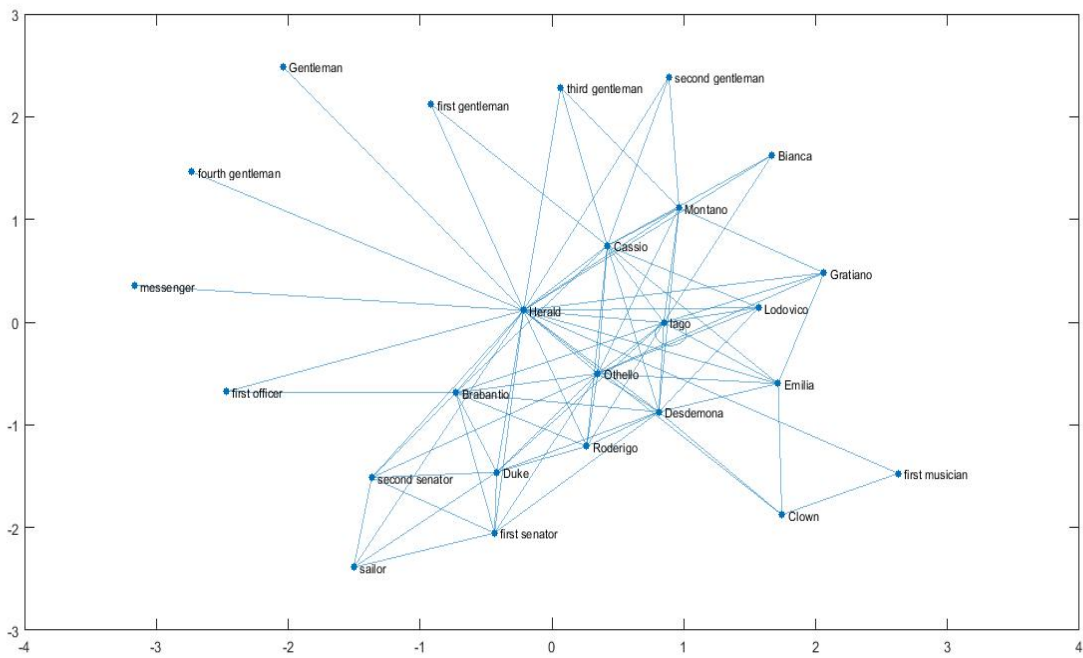


Figura A.6: Grafo Money

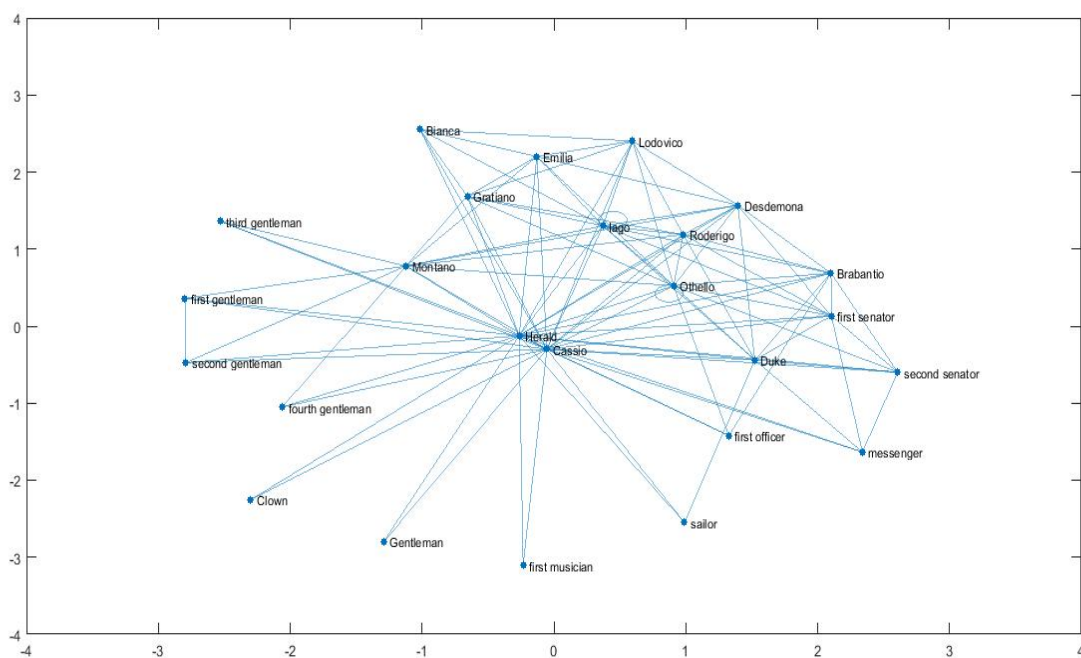


Figura A.7: Grafo Nature

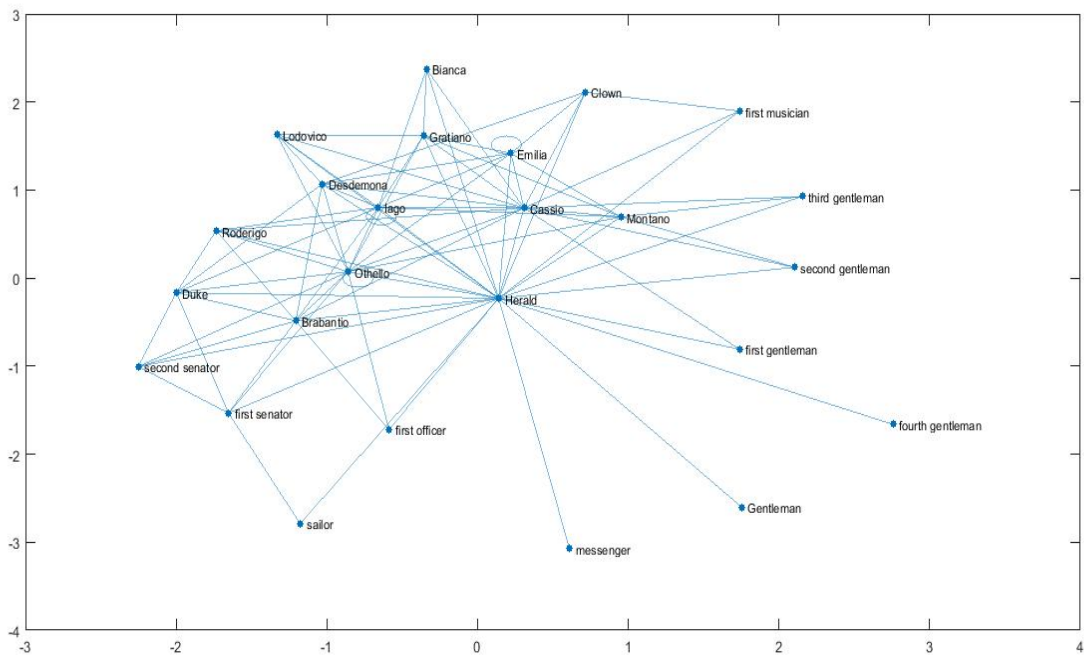


Figura A.8: Grafo Pleasure

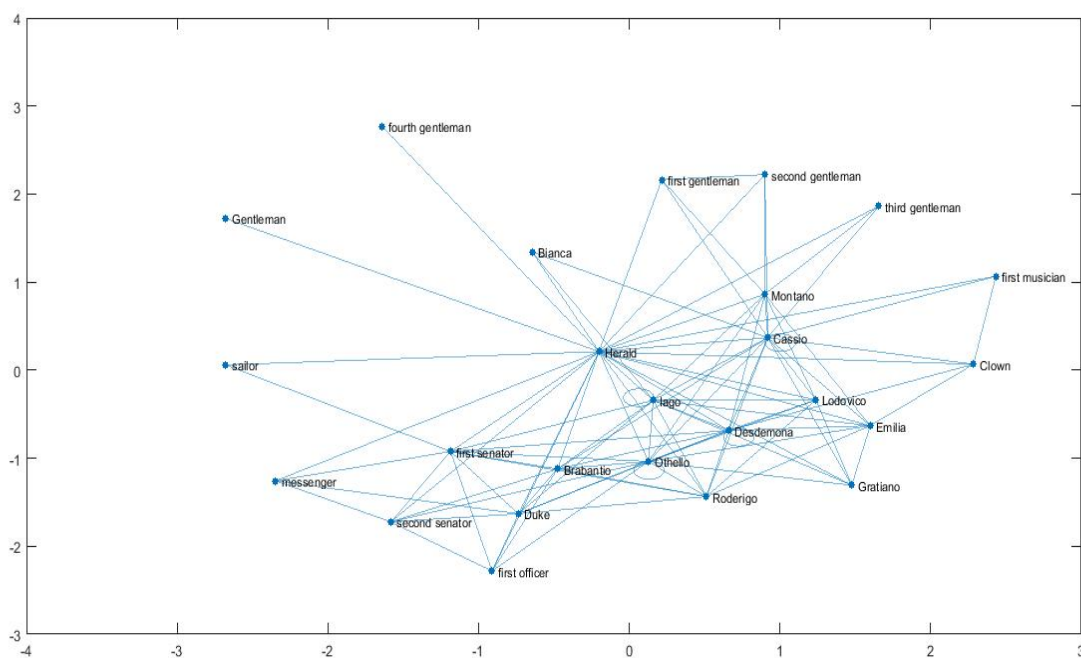


Figura A.9: Grafo Power

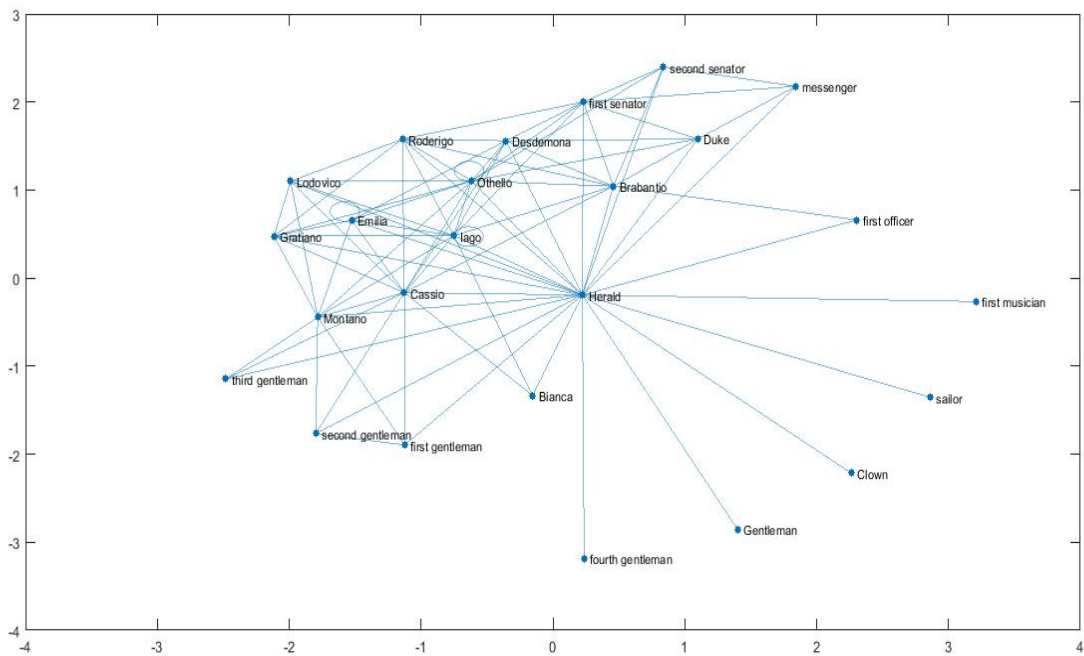


Figura A.10: Grafo Religion

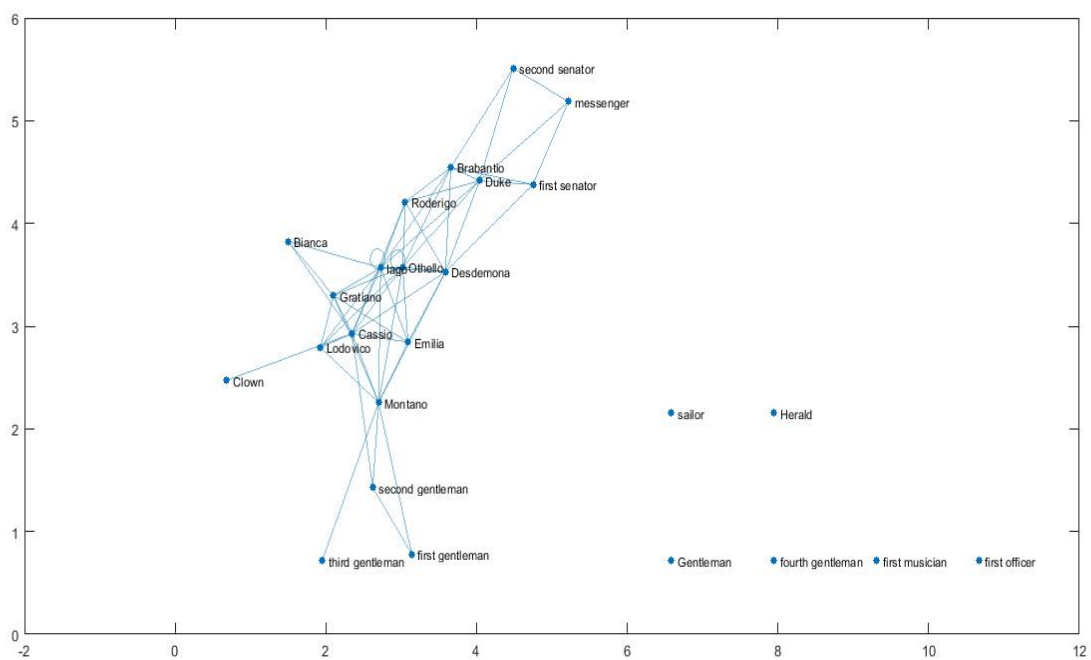


Figura A.11: Grafo See

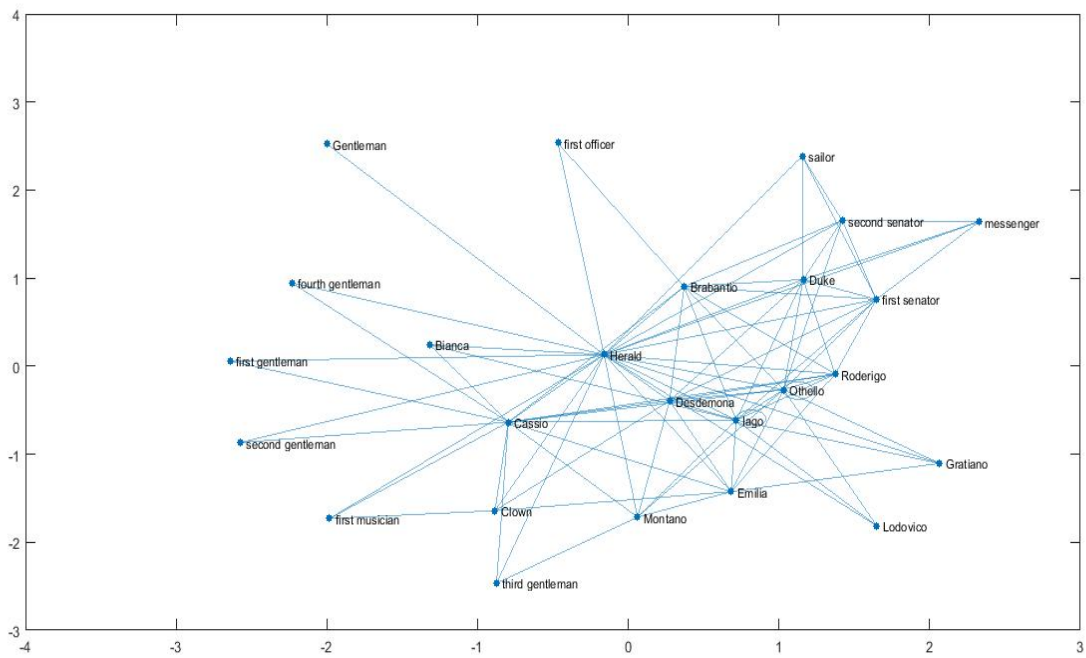


Figura A.12: Grafo Speak

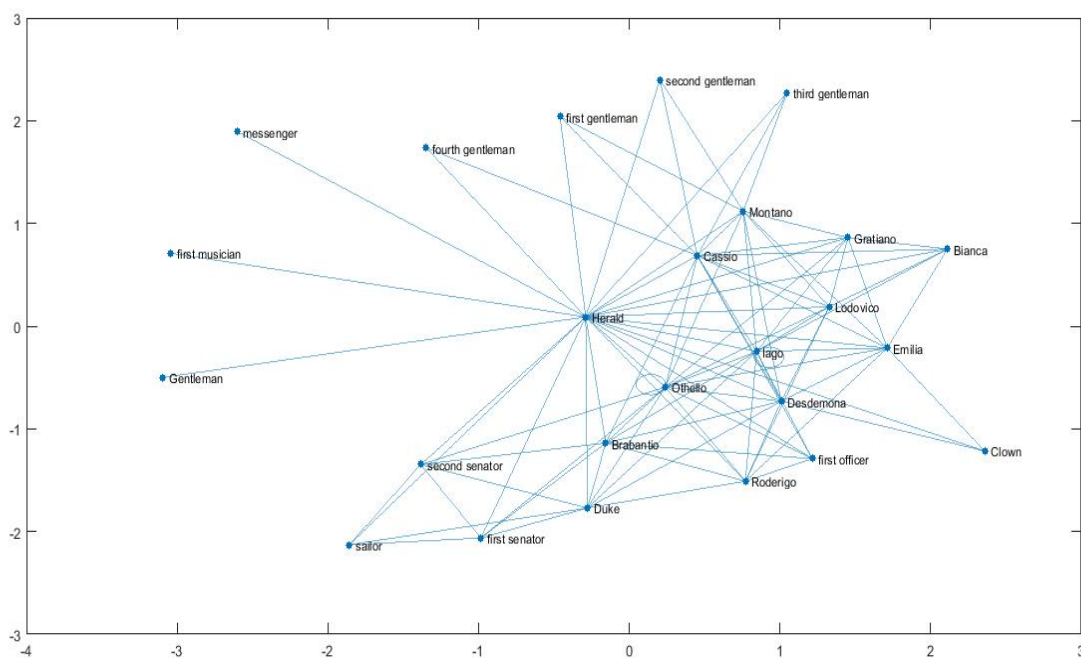


Figura A.13: Grafo Truth

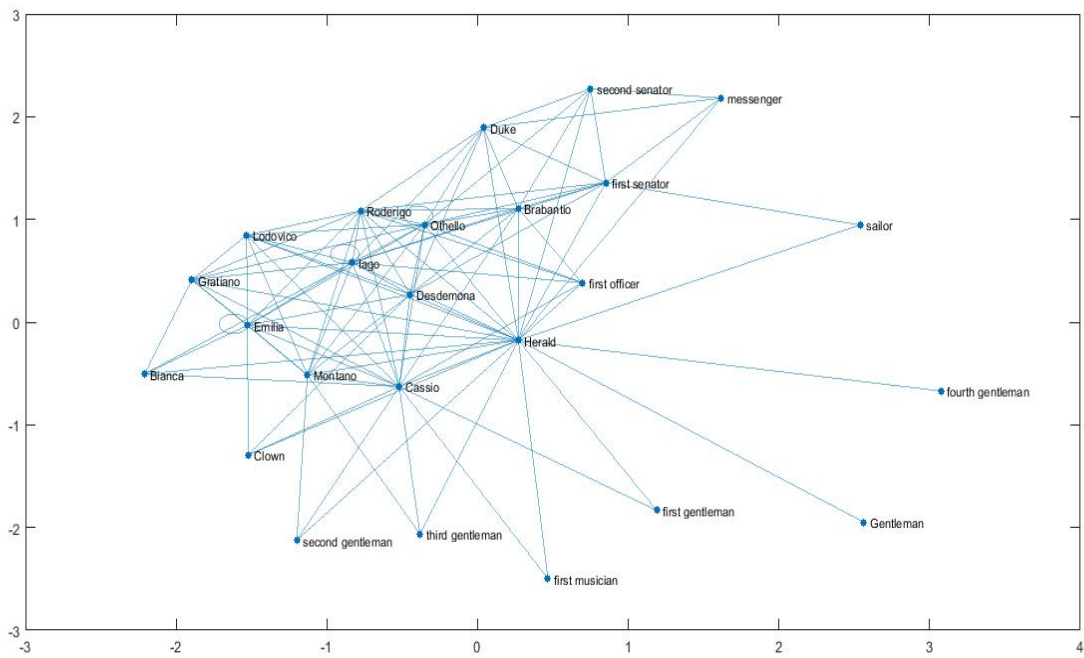


Figura A.14: Grafo Virtue

A.2 Classifiche personaggi per campo semantico

Tabella A.1: Ranking Beauty

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Othello	1	1	2	1	2	1	1
Iago	2	2	3	1	1	2	2
Desdemona	3	3	7	4	4	3	3
Cassio	4	7	1	3	3	7	4
Duke	5	6	8	8	6	4	5
Roderigo	7	4	11	8	5	6	6
Emilia	6	5	10	6	7	5	6
Montano	8	8	6	6	8	8	8
Brabantio	9	9	5	4	9	9	9
first senator	10	10	4	8	10	10	10
second senator	11	11	15	14	11	11	11
Gratiano	13	12	9	11	13	13	12
Bianca	12	13	13	13	12	12	13
Lodovico	15	14	12	11	14	14	14
second gentleman	14	16	13	15	16	15	15
first officer	17	15	15	16	15	15	16
first gentleman	16	17	15	17	17	17	17
fourth gentleman	18	18	15	18	18	18	18
third gentleman	18	19	15	18	18	18	19
sailor	20	20	15	20	20	18	20
Clown	21	21	15	21	21	21	21
Gentleman	21	21	15	21	21	21	21
first musician	21	21	15	21	21	21	21
messenger	21	21	15	21	21	21	21
Herald	21	21	15	21	21	21	21

Tabella A.2: Ranking Body

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree
Othello	2	1	2	2	2	1
Iago	1	2	8	3	1	2
Desdemona	3	3	7	4	4	3
Cassio	4	5	1	1	3	4
Emilia	6	4	14	10	6	5
Brabantio	5	7	5	6	7	6
Roderigo	7	6	10	6	5	7
Montano	8	8	4	4	8	8
Duke	9	10	6	11	10	9
Lodovico	10	9	12	8	9	10
Gratiano	11	11	13	8	11	11
first senator	13	12	9	13	12	12
second senator	15	13	11	14	12	13
Bianca	17	14	17	12	14	14
second gentleman	16	15	15	14	15	15
fourth gentleman	19	17	18	17	17	16
messenger	18	19	16	21	19	16
first officer	20	16	18	20	16	18
third gentleman	21	18	18	17	18	19
Clown	12	21	3	14	20	20
first gentleman	22	20	18	17	21	21
first musician	14	22	18	22	22	21
Gentleman	23	23	18	23	23	23
sailor	23	23	18	23	23	23
Herald	23	23	18	23	23	23

Tabella A.3: Ranking Death

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Cassio	1	3	1	1	1	3	1
Othello	2	1	2	2	3	1	1
Iago	3	2	3	3	2	2	3
Desdemona	5	5	4	4	4	5	4
Emilia	4	4	7	8	5	4	5
Montano	6	6	5	8	6	6	6
Gratiano	7	7	10	8	8	7	7
Lodovico	8	8	11	11	7	8	8
Roderigo	10	9	6	4	9	9	9
Duke	9	10	8	6	10	9	10
first senator	11	11	8	6	12	11	11
Brabantio	12	12	12	11	13	12	12
second senator	13	13	12	13	13	13	13
second gentleman	14	14	12	17	11	14	14
third gentleman	15	15	12	15	15	14	15
first gentleman	16	16	12	17	15	16	16
Clown	17	17	12	14	17	16	17
Bianca	18	18	12	15	17	18	18
Herald	19	19	12	17	17	19	19
first officer	19	20	12	17	17	19	20
fourth gentleman	19	21	12	17	17	19	21
first musician	19	22	12	17	17	19	22
Gentleman	19	23	12	17	17	19	23
messenger	19	24	12	17	17	19	24
sailor	19	25	12	17	17	19	25

Tabella A.4: Ranking Love

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Othello	1	1	3	1	1	1	1
Iago	2	2	5	1	2	2	2
Desdemona	3	3	7	4	3	3	3
Cassio	5	6	1	1	4	5	4
Emilia	4	4	8	6	6	4	5
Brabantio	6	7	4	4	7	6	6
Roderigo	9	5	10	8	5	7	7
Montano	10	8	2	6	8	8	7
Duke	11	9	11	11	9	8	9
first senator	12	10	11	11	10	10	10
Gratiano	13	12	9	8	14	11	11
Lodovico	14	11	13	10	12	11	12
second gentleman	15	15	6	14	15	15	13
second senator	16	13	14	14	11	13	14
Bianca	17	14	14	13	13	14	15
Clown	7	19	14	19	19	17	16
first musician	7	19	14	19	19	17	16
first gentleman	18	18	14	18	16	16	18
first officer	20	16	14	16	17	17	19
third gentleman	19	17	14	17	18	17	20
Gentleman	21	19	14	21	21	21	21
fourth gentleman	21	19	14	21	21	21	21
messenger	21	19	14	21	21	21	21
sailor	21	19	14	21	21	21	21
Herald	21	19	14	21	21	21	21

Tabella A.5: Ranking Mind

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Othello	2	2	1	1	1	2	1
Iago	1	1	9	3	2	1	2
Desdemona	3	3	3	2	5	3	3
Cassio	4	5	5	3	4	4	4
Emilia	5	6	8	6	6	5	5
Roderigo	7	4	12	8	3	6	6
Brabantio	6	7	7	5	8	7	7
Montano	8	8	4	7	7	8	7
first senator	9	12	2	11	11	11	9
Duke	10	11	10	8	9	10	10
Lodovico	11	9	13	8	10	9	11
Gratiano	13	10	11	12	12	12	12
second senator	12	13	14	13	13	13	13
Clown	14	16	6	14	16	16	14
Bianca	15	14	14	16	14	14	15
first officer	16	15	14	15	15	15	16
sailor	17	19	14	19	17	17	17
second gentleman	19	17	14	17	18	17	18
messenger	20	20	14	19	19	19	19
first gentleman	21	18	14	18	20	20	20
first musician	18	21	14	21	21	20	21
Gentleman	22	22	14	22	22	22	22
third gentleman	22	22	14	22	22	22	22
fourth gentleman	22	22	14	22	22	22	22
Herald	22	22	14	22	22	22	22

Tabella A.6: Ranking Money

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Iago	1	1	5	4	2	1	1
Othello	2	2	3	2	1	2	1
Desdemona	3	3	4	4	3	3	3
Cassio	5	5	2	2	5	5	4
Roderigo	6	4	13	9	4	4	5
Duke	7	7	8	7	6	6	6
Emilia	8	6	9	9	7	7	7
Brabantio	9	8	6	6	8	8	8
Herald	4	14	1	1	15	9	9
Montano	11	9	7	8	9	11	10
first senator	10	11	10	9	10	10	11
second senator	12	13	12	12	11	12	12
Lodovico	13	10	15	13	12	13	13
Gratiano	14	12	13	13	12	14	14
second gentleman	17	16	15	17	14	16	15
Clown	16	17	11	15	18	17	16
sailor	15	18	15	15	16	15	16
Bianca	18	15	15	17	17	17	18
third gentleman	20	19	15	17	19	19	19
first gentleman	21	20	15	20	20	20	20
first musician	19	22	15	20	22	20	21
first officer	22	21	15	20	21	20	22
Gentleman	23	23	15	23	23	23	23
messenger	23	24	15	23	23	23	24
fourth gentleman	23	25	15	23	23	23	25

Tabella A.7: Ranking Nature

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Iago	2	1	5	4	1	1	1
Cassio	3	3	1	1	2	3	2
Othello	1	2	4	3	3	2	2
Desdemona	5	5	8	6	5	4	4
Brabantio	4	6	7	9	6	5	5
Roderigo	7	4	8	6	4	6	6
Montano	9	8	3	5	9	8	7
Duke	10	9	6	6	7	8	8
Emilia	8	7	11	11	8	7	9
Herald	6	14	1	1	16	10	10
first senator	11	12	10	9	10	12	11
Lodovico	12	10	11	11	12	11	12
Gratiano	13	11	11	11	13	13	13
second senator	14	13	14	14	11	14	14
Bianca	16	15	15	15	15	15	15
second gentleman	15	16	15	17	14	16	15
first gentleman	17	18	15	17	17	17	17
third gentleman	18	19	15	20	18	18	18
first officer	19	17	15	17	19	18	18
fourth gentleman	21	20	15	20	20	21	20
messenger	20	22	15	16	22	20	21
Clown	22	21	15	23	20	22	22
sailor	23	23	15	20	22	22	23
Gentleman	24	24	15	23	22	24	24
first musician	24	25	15	23	22	24	25

Tabella A.8: Ranking Pleasure

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Othello	1	2	3	3	1	1	1
Iago	2	1	4	4	2	2	2
Desdemona	4	3	7	6	3	3	3
Cassio	5	6	2	2	5	6	4
Emilia	6	5	10	7	6	4	5
Roderigo	7	4	13	11	4	5	6
Brabantio	8	7	5	5	7	7	7
Herald	3	12	1	1	11	8	8
Montano	9	8	6	7	10	9	9
Duke	10	10	9	7	8	9	10
Lodovico	12	9	13	11	9	12	11
Gratiano	11	11	11	7	12	11	12
first senator	13	13	8	11	13	13	13
second senator	14	14	15	14	13	14	14
Clown	15	15	12	14	15	15	15
Bianca	17	16	15	16	16	16	16
first officer	18	17	15	17	17	17	17
first musician	16	19	15	17	19	17	18
second gentleman	19	18	15	17	18	19	19
third gentleman	20	20	15	17	20	20	20
first gentleman	22	21	15	21	21	21	21
sailor	21	22	15	21	21	21	21
fourth gentleman	23	23	15	23	21	23	23
Gentleman	23	24	15	23	21	23	24
messenger	23	25	15	23	21	23	25

Tabella A.9: Ranking Power

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Othello	1	1	3	3	1	1	1
Iago	2	2	4	4	2	2	2
Desdemona	3	3	7	5	3	3	3
Cassio	4	6	2	2	4	4	4
Roderigo	6	4	11	7	5	6	5
Brabantio	5	7	9	7	7	5	6
Emilia	7	5	10	11	6	7	7
Duke	8	8	8	7	8	8	8
Montano	9	9	6	7	9	9	9
first senator	11	11	5	6	10	10	10
Lodovico	12	10	14	12	11	10	11
Herald	10	14	1	1	22	13	12
second senator	13	12	12	13	12	12	13
Gratiano	15	13	15	14	13	14	14
Clown	14	17	13	16	18	15	15
second gentleman	16	18	16	17	14	17	16
first officer	18	15	16	14	16	16	16
Bianca	19	16	16	20	15	18	18
third gentleman	20	19	16	20	17	19	19
first musician	17	22	16	20	21	19	20
messenger	22	20	16	17	19	21	21
first gentleman	21	21	16	17	23	22	22
sailor	23	23	16	23	19	23	23
Gentleman	24	24	16	24	24	24	24
fourth gentleman	24	25	16	24	24	24	25

Tabella A.10: Ranking Religion

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Othello	1	2	3	3	1	1	1
Iago	2	1	4	4	2	2	2
Desdemona	5	3	8	7	4	3	3
Brabantio	4	6	5	5	5	4	4
Cassio	6	7	2	2	3	6	4
Roderigo	8	5	9	7	6	7	6
Emilia	7	4	13	11	7	5	7
Herald	3	14	1	1	10	9	8
Montano	9	8	6	5	8	8	9
first senator	10	9	7	7	9	10	10
Gratiano	11	10	12	10	13	11	11
Duke	13	12	10	13	12	13	12
second senator	14	13	11	14	11	14	13
Lodovico	12	11	15	11	14	12	14
second gentleman	15	16	16	15	15	15	15
third gentleman	16	17	16	18	16	16	16
Bianca	18	15	16	18	17	17	17
first gentleman	17	18	16	15	19	18	18
messenger	19	20	14	15	20	18	19
first officer	20	19	16	20	18	20	20
Gentleman	21	21	16	21	21	21	21
fourth gentleman	21	22	16	21	21	21	22
Clown	21	23	16	21	21	21	23
first musician	21	24	16	21	21	21	24
sailor	21	25	16	21	21	21	25

Tabella A.11: Ranking See

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Iago	1	1	3	1	1	1	1
Othello	2	2	6	2	2	2	2
Cassio	3	3	2	3	4	3	3
Desdemona	4	5	5	3	7	4	4
Roderigo	7	4	9	6	3	6	5
Brabantio	5	7	7	10	5	5	5
Montano	6	9	1	5	6	8	7
Duke	8	11	4	6	11	9	8
Emilia	9	6	14	6	9	7	9
Gratiano	10	10	10	6	10	10	10
Lodovico	11	8	15	11	8	10	11
first senator	13	13	8	13	13	12	12
second senator	14	14	12	18	13	13	13
Bianca	15	12	15	12	12	14	14
second gentleman	12	15	11	14	15	15	14
third gentleman	17	16	15	17	16	16	16
first gentleman	16	18	15	15	17	16	17
messenger	18	19	12	18	19	18	18
Clown	19	17	15	15	18	19	19
Gentleman	20	20	15	20	20	20	20
fourth gentleman	20	20	15	20	20	20	20
first musician	20	20	15	20	20	20	20
first officer	20	20	15	20	20	20	20
sailor	20	20	15	20	20	20	20
Herald	20	20	15	20	20	20	20

Tabella A.12: Ranking Speak

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Othello	1	1	4	3	1	1	1
Iago	2	2	3	3	2	2	2
Cassio	5	5	2	2	3	5	3
Emilia	3	3	9	9	4	3	4
Desdemona	4	4	6	5	5	4	5
Brabantio	7	7	5	6	8	7	6
Roderigo	6	6	12	9	6	6	7
Duke	8	9	7	6	7	8	8
Montano	10	8	10	11	10	9	9
first senator	11	10	7	6	9	10	9
second senator	12	12	11	11	11	11	11
Herald	9	15	1	1	20	13	12
Gratiano	14	11	14	14	12	12	13
Clown	13	14	13	13	14	14	14
Lodovico	15	13	14	17	13	15	15
Bianca	17	16	14	17	15	16	16
first musician	16	21	14	17	17	17	17
first officer	20	17	14	21	16	19	18
sailor	18	19	14	14	19	18	19
second gentleman	21	18	14	21	18	21	20
third gentleman	22	20	14	17	21	21	21
messenger	19	22	14	14	24	19	21
first gentleman	23	23	14	21	22	23	23
fourth gentleman	23	24	14	21	22	23	24
Gentleman	25	25	14	25	25	25	25

Tabella A.13: Ranking Truth

Name	pagerank	eigenvector	betweenness	closeness	closenessp	degree	media
Othello	1	2	3	3	1	1	1
Iago	2	1	5	4	2	2	2
Emilia	3	3	7	7	3	3	3
Brabantio	4	5	8	7	4	4	4
Desdemona	5	4	6	5	5	5	4
Cassio	6	7	2	2	7	7	6
Roderigo	7	6	10	9	6	6	7
Montano	8	8	4	5	8	8	8
Duke	10	9	9	9	9	9	9
Gratiano	11	10	14	12	12	10	10
Lodovico	12	11	11	9	13	11	11
first senator	13	12	12	13	10	12	12
Herald	9	16	1	1	21	14	13
second senator	14	13	12	13	11	13	14
first officer	15	14	15	13	14	15	15
Clown	16	15	17	18	15	16	16
Bianca	17	17	16	13	16	17	17
second gentleman	19	18	17	18	17	18	18
third gentleman	20	19	17	18	18	20	19
sailor	18	21	17	17	20	18	19
first gentleman	20	20	17	18	18	20	21
fourth gentleman	22	22	17	22	22	22	22
Gentleman	23	23	17	23	23	23	23
first musician	23	24	17	23	23	23	24
messenger	23	25	17	23	23	23	25

A.3 Grafici personaggi per campo semantico

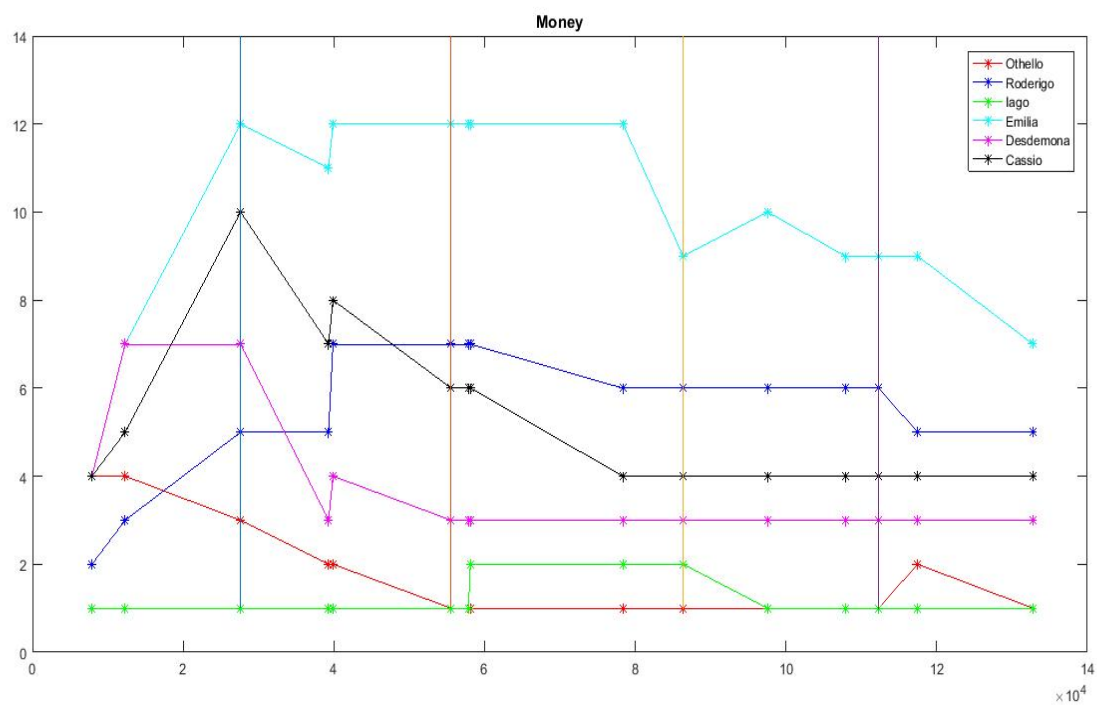


Figura A.15: Posizioni temporali dei personaggi considerando il campo semantico Money

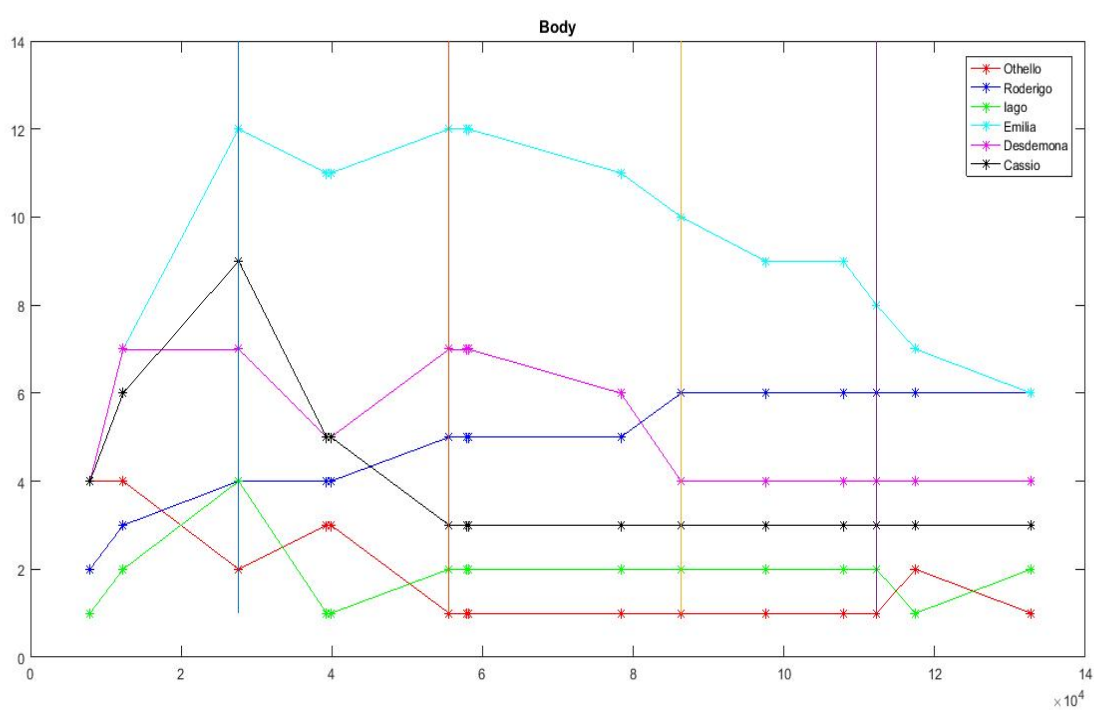


Figura A.16: Posizioni temporali dei personaggi considerando il campo semantico Body

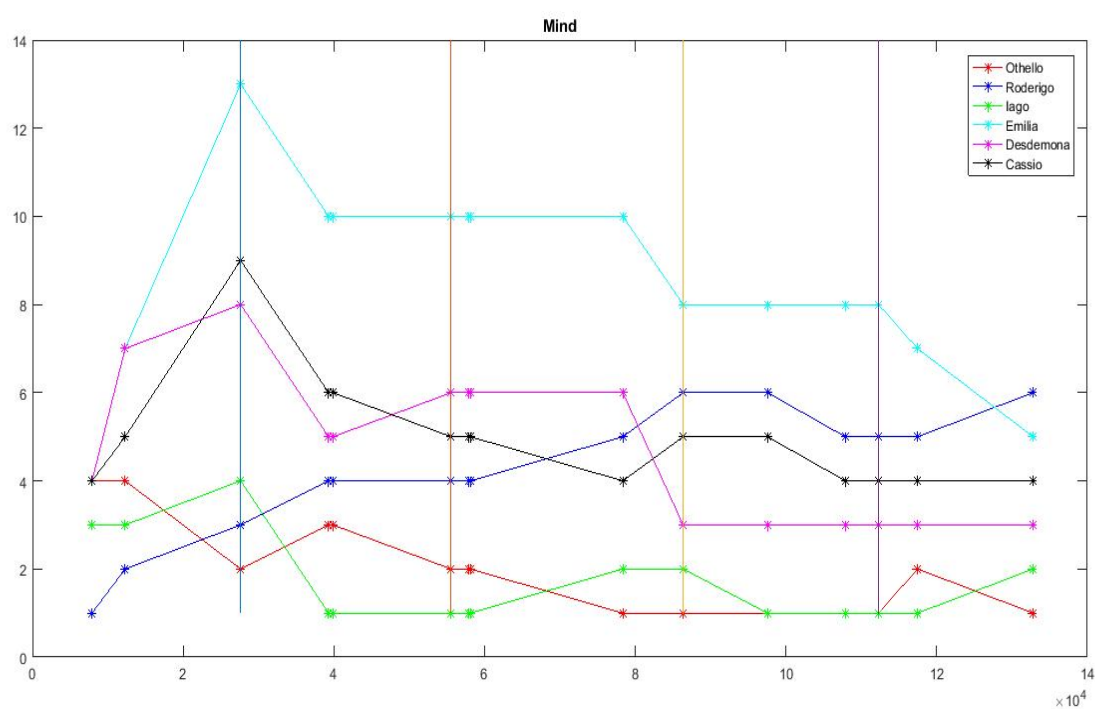


Figura A.17: Posizioni temporali dei personaggi considerando il campo semantico Mind

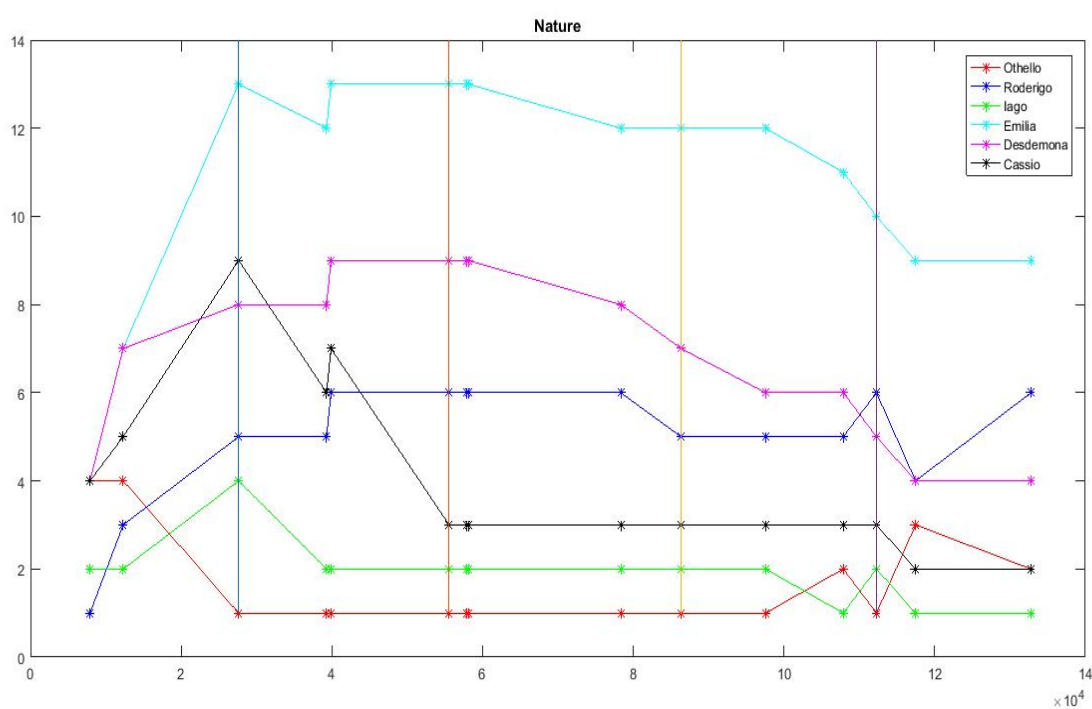


Figura A.18: Posizioni temporali dei personaggi considerando il campo semantico Nature

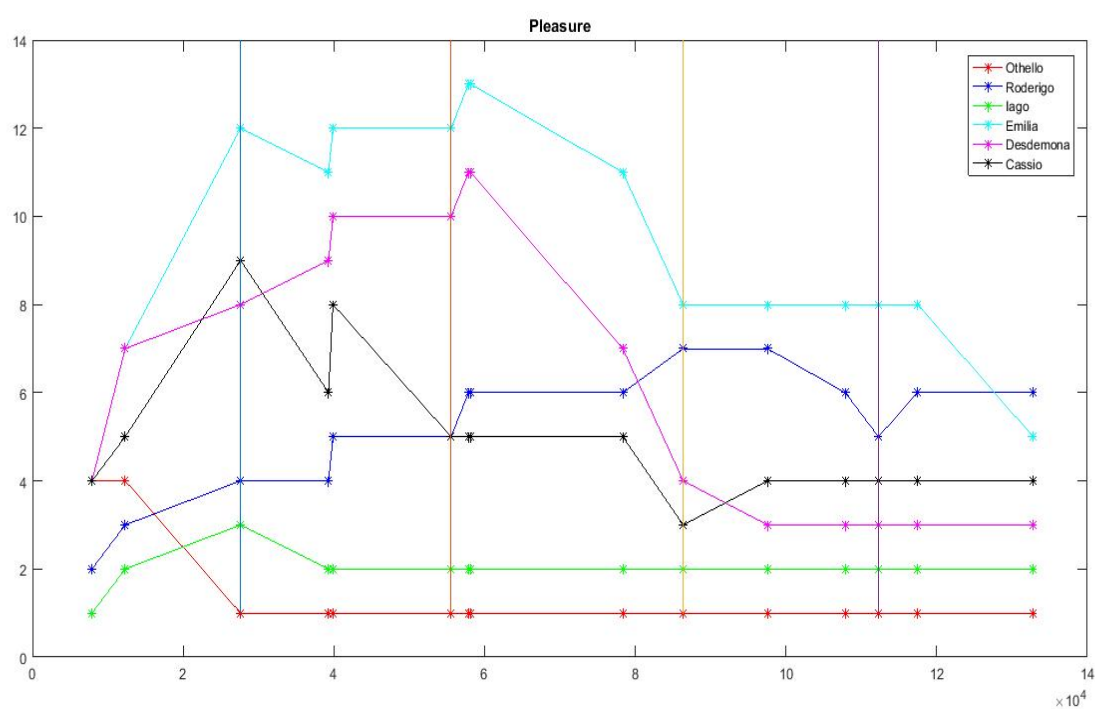


Figura A.19: Posizioni temporali dei personaggi considerando il campo semantico Pleasure

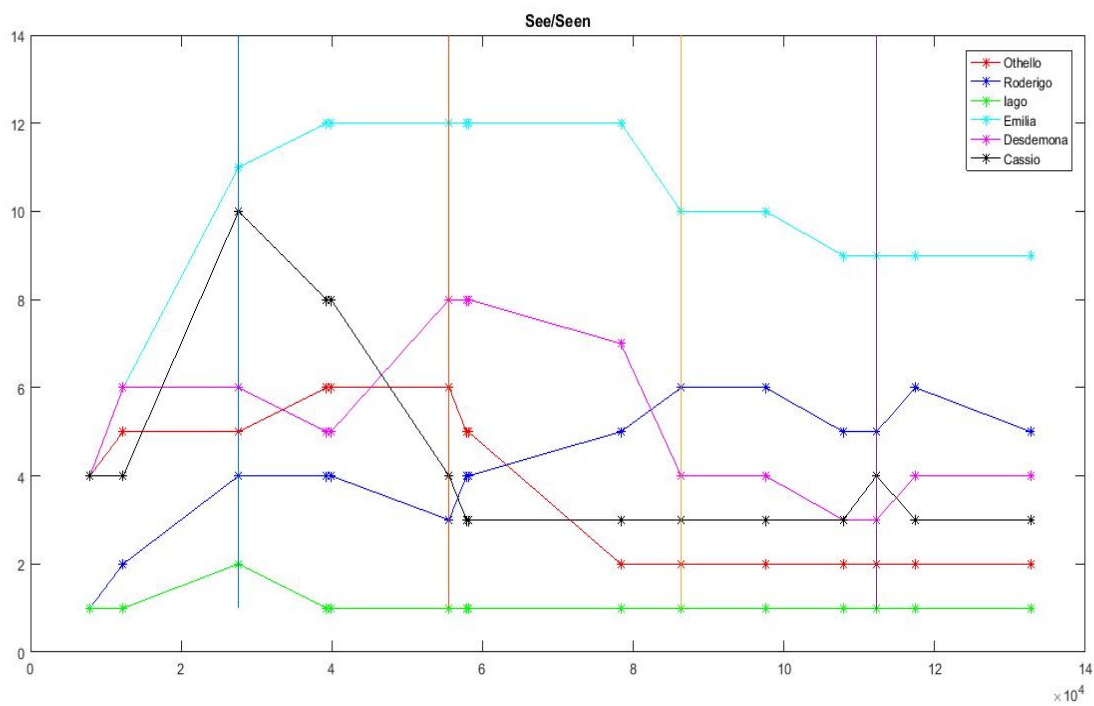


Figura A.20: Posizioni temporali dei personaggi considerando il campo semantico See

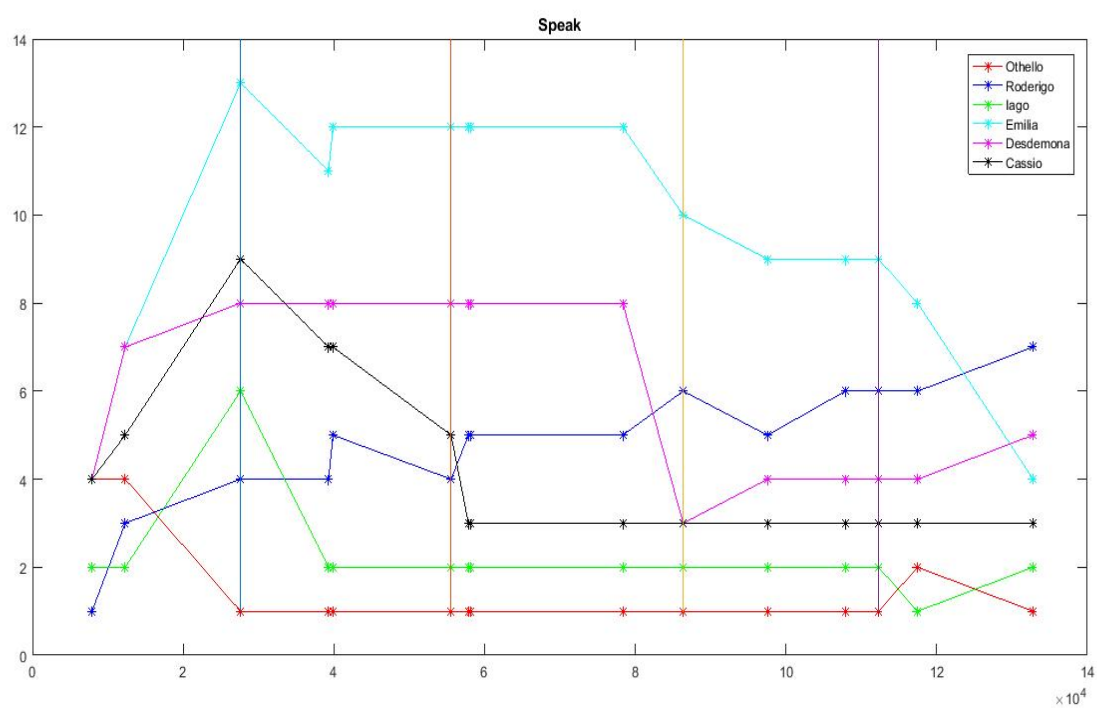


Figura A.21: Posizioni temporali dei personaggi considerando il campo semantico Speak

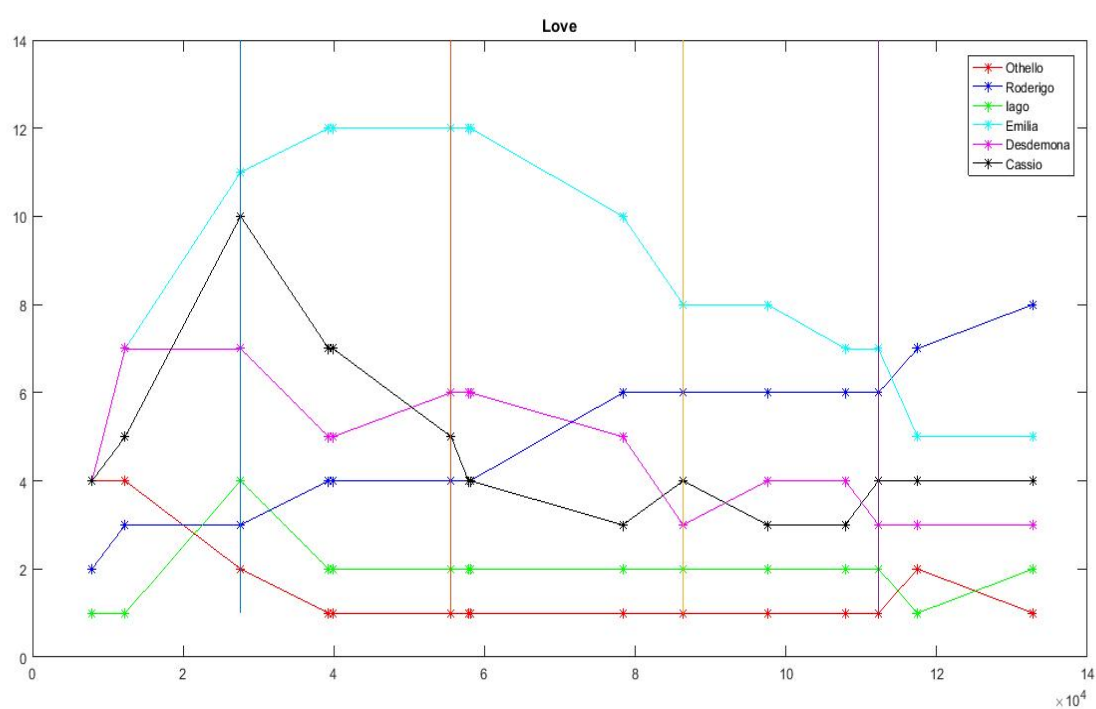


Figura A.22: Posizioni temporali dei personaggi considerando il campo semantico Love

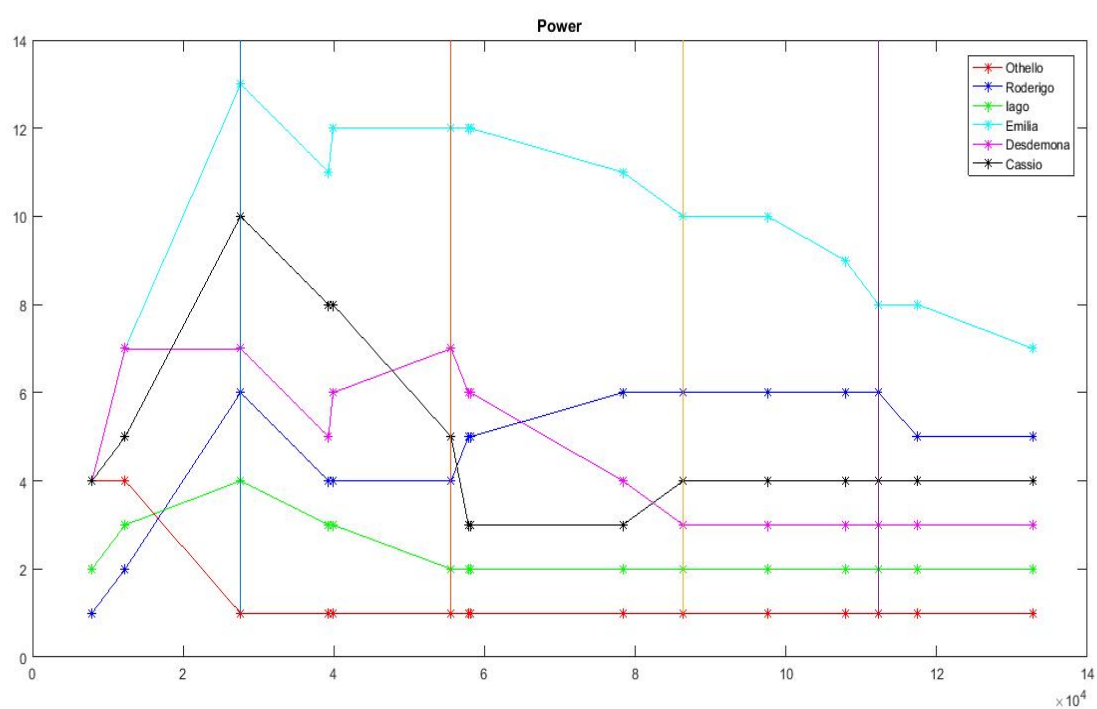


Figura A.23: Posizioni temporali dei personaggi considerando il campo semantico Power

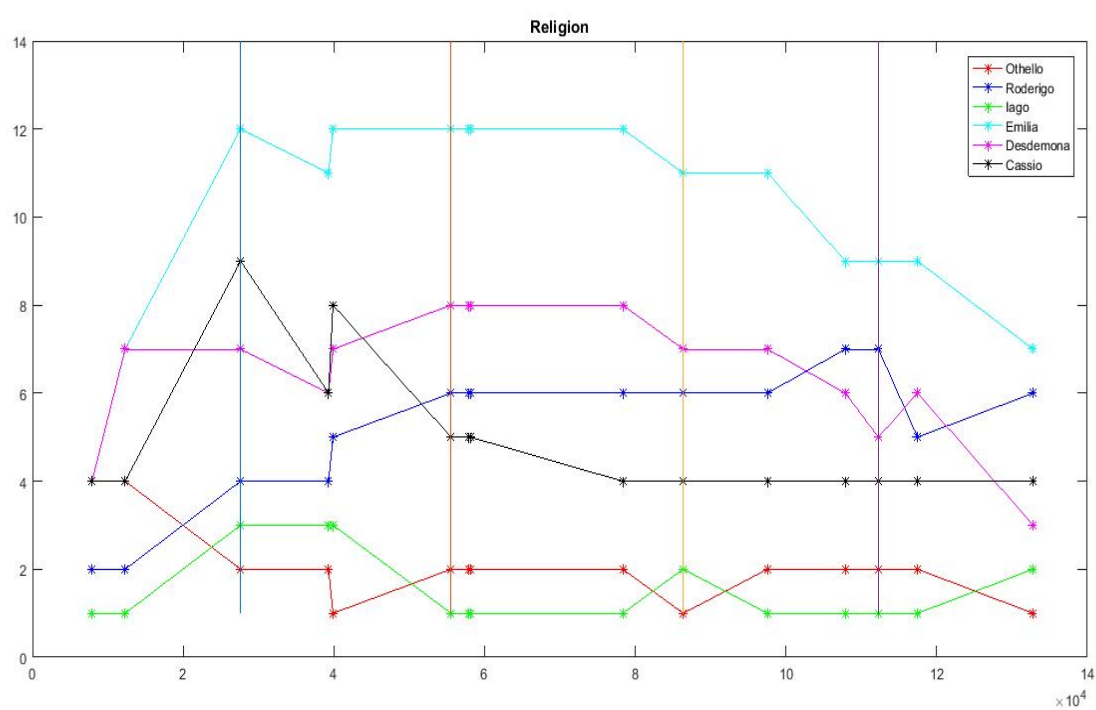


Figura A.24: Posizioni temporali dei personaggi considerando il campo semantico Religion

A.4 Tabelle campi semantici

Tabella A.14: Valori numerici con riduzione a 200 caratteri

Name	indegree	outdegree	hubs	authoritiess	pagerank
Money	86,5649	87,8175	0,0336	0,0321	0,0389
Beauty	104,0658	103,7807	0,0395	0,0387	0,0447
Death	113,7846	108,1595	0,0393	0,0415	0,0480
See_Seen	151,2968	139,7312	0,0526	0,0571	0,0600
Nature	156,2883	150,9501	0,0570	0,0597	0,0613
Truth	185,5501	182,9371	0,0695	0,0708	0,0704
Body	188,7149	187,1402	0,0693	0,0716	0,0717
Pleasure	190,6913	191,7648	0,0721	0,0730	0,0723
Speak	193,2016	206,7317	0,0799	0,0734	0,0730
Love	230,0048	223,7492	0,0844	0,0868	0,0854
Power	232,9158	237,7635	0,0909	0,0902	0,0857
Religion	240,3873	254,9346	0,0983	0,0921	0,0882
Virtue	268,1860	267,6705	0,1037	0,1031	0,0973
Mind	285,6835	284,2051	0,1097	0,1099	0,1031

Tabella A.15: Valori numerici con riduzione a 50 caratteri

Name	indegree	outdegree	hubs	authoritiess	pagerank
Mind	122,9310	123,1952	0,1149	0,1129	0,1056
Virtue	119,085	120,3083	0,1110	0,1102	0,1011
Religion	103,6702	107,5580	0,10023	0,0963	0,0899
Power	96,0797	97,4394	0,0880	0,0904	0,0833
Love	91,5117	86,3121	0,0763	0,0799	0,0817
Speak	82,0940	95,9217	0,0902	0,0730	0,0733
Truth	81,2371	80,7226	0,0733	0,0745	0,0722
Pleasure	76,4513	76,3427	0,0665	0,0697	0,0694
Body	73,5142	82,1862	0,0706	0,0665	0,0662
See_Seen	64,7175	56,9234	0,0503	0,0569	0,0612
Nature	65,2809	60,5202	0,0537	0,0594	0,0611
Death	49,0542	43,2629	0,0354	0,0411	0,0492
Beauty	43,3501	42,3056	0,0389	0,0364	0,0449
Money	38,2894	34,2681	0,0306	0,0328	0,0409

A.5 Grafici campi semantici

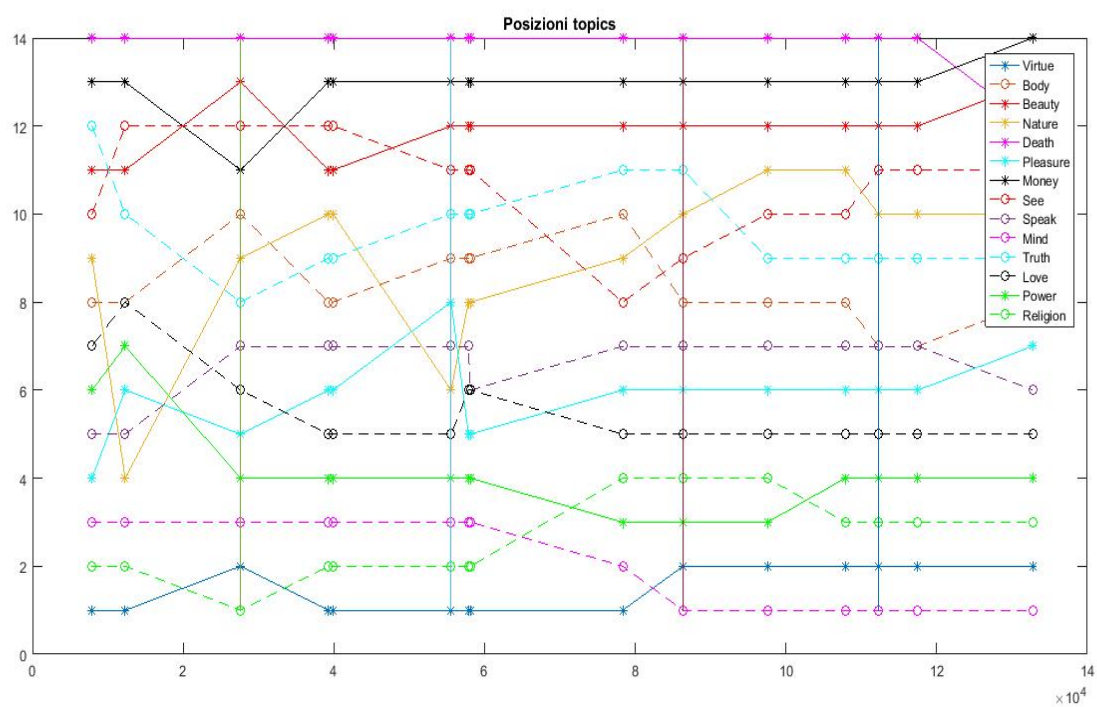


Figura A.25: Classifica temporale dopo la riduzione a 200 caratteri

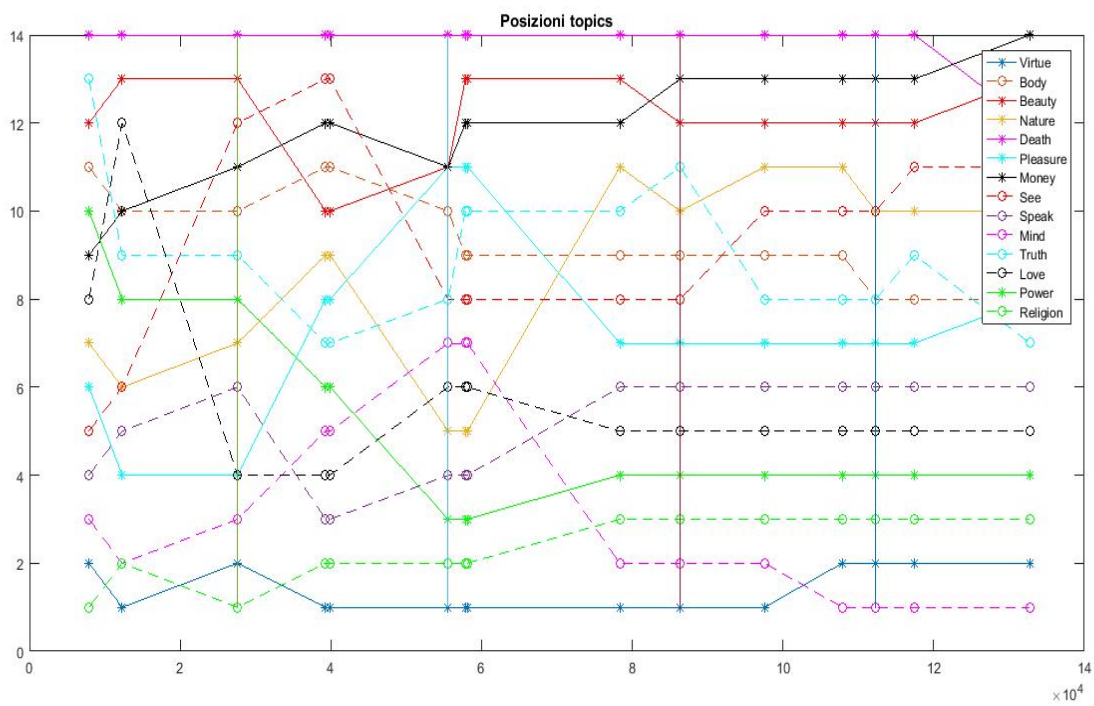


Figura A.26: Classifica temporale dopo la riduzione a 50 caratteri

Bibliografia

- [1] Mark Newman, *Networks An Introduction*, Oxford university press, 2010
- [2] Mark Newman, *The mathematics of network*, Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109-1040
- [3] M. Benzi, Christine Klymko, *On the limiting behavior of parameterdependent network centrality measures*, 2015
- [4] Page, Larry, *PageRank: Bringing Order to the Web* (1999) Engine Rankings, Princeton University Press, Princeton, NJ, 2006
- [5] Reinhard Diestel, *Graph Theory*, 4th Electronic Edition, 2010
- [6] Jonathan Kelner, *18.409 Topics in Theoretical Computer Science: An Algorithmist's Toolkit*. Fall 2009. Massachusetts Institute of Technology: MIT OpenCourseWare, <https://ocw.mit.edu>. License: Creative Commons BY-NC-SA.
- [7] Fan Chiung, *Four proofs for the Cheeger inequality and graph partition algorithms*, Paper presented at the Fourth International Congress of Chinese Mathematicians (ICCM), Hangzhou, China, December 17–22, 2007.
- [8] Merris Russell, *Laplacian matrices of graphs: a survey* Linear Algebra and its Applications 197 (1994): 143–176.
- [9] Andries E. Brouwer, Willem H. Haemers, *Spectra of graphs*, Springer, New York, 2011.

-
- [10] . Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian*, in Problems in Analysis (R. C. Gunnig, Ed.), pp. 195-199, Princeton Univ. Press, Princeton, NJ, 1970.
- [11] Fan R. K. Chung, *Spectral graph theory*, American Mathematical Society, 1997.
- [12] M. E. J. Newman, *Mixing patterns in networks*. Physical Review E. 67 (2): 026126, 2003
- [13] Stanley Wasserman, Kathrine Faust, *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press, 1994.
- [14] M. E. J. Newman, *Modularity and community structure in networks*, in PNAS, vol. 103, n° 23, 2006.
- [15] P. Bonacich, P. Lloyd. *Eigenvector-like measures of centrality for asymmetric relations*, Social Networks 23, Department of Sociology, University of California at Los Angeles, 2001.
- [16] R. Perissinotti, Tesi di Laurea Magistrale in Teoria dei grafi, *Relations among Shakespeare's characters: an analysis in terms of centrality measures and new techniques from graph theory*, Università di Bologna, I sessione 2016-2017.
- [17] A. Beveridge, J. Shan, *Network of Thrones*, Math Horizons Magazine 23, 18–22, 2016
- [18] M.A. Ribeiro, R.A. Vosgerau, M.L.P. Andruchiw, S. Ely de Souza Pinto, *The complex social network from the Lord of the Rings*, Rev. Bras. Ensino Fs. 2016
- [19] A. Bonato, D.R. D'Angelo, E.R. Elenberg, D.F. Gleich, Y. Hou *Mining and Modeling Character Networks*

-
- [20] A. Agarwal, A. Corvalan, J. Jensen, O. Rambow, *Social network analysis of 'Alice in Wonderland'*, Workshop on Computational Linguistics for Literature, 88–96, 2012
- [21] V.H.Masias, P. Baldwin, S. Laengle, A. Vargas, F.A. Crespo *Exploring the prominence of Romeo and Juliet's characters using weighted centrality measures*, Digital Scholarship in the Humanities, Vol. 32, No. 4, 2017
- [22] *The Tragedy of Othello, Moor of Venice*, www.opensourceshakespeare.org
- [23] Carl D. Meyer, *Matrix analysis and applied linear algebra*, SIAM: Society for Industrial and Applied Mathematics; Har/Cdr edition (February 15, 2001)
- [24] Rocco Coronato, *comunicazione privata*, 2018.
- [25] Horn, R. A. and Johnson, C. R., *Matrix Analysis*, Cambridge University Press, 1985.