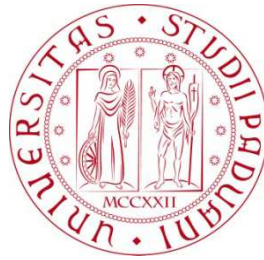


Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Magistrale in  
Scienze Statistiche



Caratteristiche psicologiche delle reti sociali su Facebook

Relatore: Prof. Livio Finos  
Dipartimento di Scienze Statistiche

Correlatore: Dott.ssa Claudia Marino  
Dipartimento di Psicologia

Laureando: Lorenzo Zanotto  
Matricola: 1036759

Anno Accademico 2016 / 2017



## **RINGRAZIAMENTI**

Con la stesura di questa tesi volge al termine il percorso di Laurea Magistrale in Scienze Statistiche.

Ringrazio anzitutto il Prof. *Livio Finos* e la Dott.ssa *Claudia Marino*: senza il loro supporto e la loro guida questa tesi non sarebbe potuta esistere. Mi sono stati forniti tutti gli strumenti utili per intraprendere la strada corretta e poter attraccare la nave in porto.

Un ringraziamento doveroso va a mio Fratello *Francesco* e alla mia Fidanzata *Elena* che con la loro forza mi hanno sostenuto anche nei momenti più bui, siete sempre al mio fianco.

Trovo doveroso ringraziare l'intera famiglia che da sempre mi supporta, i Cugini, gli Zii e i Nonni; inoltre ringrazio la mia Amica *Giulia*, con la quale abbiamo condiviso i migliori momenti dell'esperienza universitaria.

*Infine un ringraziamento speciale a Mamma Licia e Papà Romeo: è grazie a voi se oggi ho potuto raggiungere questo traguardo, intendo dedicarvi la mia tesi, vi voglio bene.*

*Lorenzo Zanotto*



# INTRODUZIONE

La prima cosa che contraddistingue l'essere umano da un robot è la capacità di relazionarsi con suoi simili. Qualità innata del mammifero per eccellenza, che gli ha permesso di fondare la civiltà sulla terra. In primis la capacità di comunicare, attraverso il corpo e le parole, associato al saper ascoltare, utile per poter instaurare una comunicazione di successo. Attività semplici svolte quotidianamente, utili a creare e mantenere attiva la rete delle relazioni personali.

Il grande potere di relazionarsi con gli altri non sta solo nella comunicazione, non sta solo nell'aumentare il proprio bagaglio culturale per via delle informazioni che vengono scambiate, sta nella capacità di modificare il comportamento dei soggetti che vi partecipano. Proprio per questo motivo è divenuta interessante l'analisi dei fenomeni comportamentali, facendo sorgere la *Network Analysis* che introduce un cambio di approccio, spostando l'attenzione dal singolo individuo alla coppia di soggetti, descrivendo la relazione esistente.

Con l'introduzione dei *Social Network*, complesse piattaforme per la gestione delle relazioni, il legame assume una consistenza digitale. Mentre la relazione classica, quella basata sulla classica stretta di mano produce informazione difficile da recuperare, la relazione digitale crea dati che hanno bisogno di essere analizzati; a questo scopo è nata la *Social Network Analysis*.

Ogni soggetto possiede un profilo psicologico ed un profilo digitale, l'obiettivo di questo lavoro è la descrizione di quali sono e come influiscono le variabili psicologiche personali sulla probabilità che esista una relazione tra due soggetti qualsiasi all'interno di una *rete sociale*.

Nel *Capitolo 1* sarà introdotto il concetto di *rete sociale*, con la descrizione dei due principali elementi costitutivi e delle essenziali analisi descrittive per valutare la tipologia di rete e relazioni presenti al suo interno.

Nel *Capitolo 2* sarà introdotta la *Social Network Analysis*, a seguire la *teoria del mondo piccolo* di Milgram (Milgram & Travers, 1969) e infine il cambio di

approccio che suggerisce di spostare l'attenzione dagli attributi del singolo soggetto alla qualità della relazione per una coppia di attori inseriti in un contesto di multiple relazioni tra simili.

Nel *Capitolo 3* vengono presentati i dati grezzi utilizzati per svolgere il lavoro, con la specifica dei file e del pacchetto *myFBr* per R Studio (Finos, et al., 2017) utilizzato durante la costruzione dei dataset.

Nel *Capitolo 4* sarà descritto il *Dataset Utenti* con le informazioni personali di ogni singolo individuo riguardanti le cinque grandi dimensioni della personalità, i cosiddetti *Big Five* (Dallago & Roccato, 2010) e gli indicatori del benessere personale ricavati dal *Social-Emotional Health Surveys* (Furlong, et al., 2013).

Nel *Capitolo 5* sarà introdotta e descritta la *Matrice di Adiacenza*, contenente le relazioni presenti nella rete, specificandone poi la modalità di costruzione e le due principali problematiche incontrate con i dati *Facebook*.

Nei *Capitolo 6* vengono descritte le caratteristiche della rete, proponendo un'interpretazione grafica coadiuvata dai principali indici descritti al termine del *Capitolo 1*.

Nel *Capitolo 7* vengono descritte le caratteristiche di ciascuno dei soggetti, presenti nel *Dataset Utenti*, dando uno sguardo dapprima alle distribuzioni marginali, condizionate e poi verificando la presenza di eventuale correlazione.

Nel *Capitolo 8* viene introdotto il *modello ERGM* e il metodo di stima della probabilità di interesse, cui faranno seguito i tre modelli utilizzati (Handcock, et al., 2017).

# INDICE

RINGRAZIAMENTI.....	III
INTRODUZIONE .....	V
1. RETE SOCIALE.....	1
1.1. Cos'è una Rete Sociale.....	1
1.2. Elementi costitutivi di una Rete Sociale .....	1
I Nodi o Vertici ( <i>Nodes</i> ) .....	1
Le Relazioni o Archi ( <i>Links</i> ) .....	2
1.3. Le tipologie di Rete Sociale .....	2
1.4. Le analisi descrittive di una Rete Sociale .....	3
<i>Density</i> , la densità della Rete Sociale (Network Level) .....	3
<i>Reciprocity</i> , l'indice di reciprocità della Rete Sociale (Network Level).....	4
<i>Transitivity</i> , l'indice di transitività (Node Level) .....	4
<i>Diameter</i> , il diametro della Rete Sociale (Network Level) .....	5
<i>Degree</i> , primo indice di centralità (Node Level) .....	5
<i>Betweenness</i> , secondo indice di centralità (Node Level).....	6
<i>Assortativity</i> , indice di omofilia (Network Level) .....	6
2. LA SOCIAL NETWORK ANALYSIS .....	9
2.1. Cos'è la Social Network Analysis .....	9
2.2. La Teoria del Mondo Piccolo di Milgram.....	9
2.3. L'esperimento di Milgram .....	10
2.4. I sei gradi di separazione.....	10
2.5. Differenze rispetto all'Approccio Classico .....	11

3.	I DATI GREZZI .....	15
3.1.	Il profilo Facebook .....	15
3.2.	Struttura dei dati in cartelle .....	15
	La cartella <i>html</i> .....	16
	La cartella <i>photos</i> .....	17
	La cartella <i>videos</i> .....	18
	Il file <i>index</i> .....	18
3.3.	La libreria myFBr per RStudio .....	19
	La funzione <i>getInfo</i> .....	20
	La funzione <i>getFriends</i> .....	20
4.	IL DATASET UTENTI .....	21
4.1.	Dataset Utenti .....	21
4.2.	La scala di <i>Personalità</i> .....	22
4.3.	La scala di <i>Benessere</i> .....	22
5.	LA MATRICE DI ADIACENZA .....	25
5.1.	Tipologia .....	25
5.2.	Modalità di costruzione .....	25
5.3.	Le problematiche incontrate .....	26
	<i>La diagonale</i> .....	27
	<i>L'amico del suo non amico</i> .....	27
6.	ANALISI DESCRITTIVE DELLA RETE .....	29
7.	ANALISI DESCRITTIVE DELLE VARIABILI .....	37
8.	IL MODELLO ERGM .....	57
8.1.	Il modello ERGM – parte teorica .....	57
8.2.	Il modello più semplice .....	58
8.3.	Il modello con covariata quantitativa .....	59



8.4. Il modello con covariata qualitativa ( <i>Homophily</i> ) .....	60
8.5. Il modello completo .....	62
Modello 1 .....	63
Modello 2 .....	64
Modello 3 .....	65
CONCLUSIONI .....	67
INDICE DELLE FIGURE.....	69
BIBLIOGRAFIA .....	71



# 1. RETE SOCIALE

## 1.1. Cos'è una Rete Sociale

Con il termine rete in ambito sociale si intende un insieme di individui collegati tra di loro, a coppie di due o più, tramite un insieme di relazioni. Di base non è importante il motivo del collegamento tra di essi, piuttosto interessante è il fatto che esso sia presente o meno.

Il concetto di rete sociale rimane generico fino quando non viene specificato il motivo, il luogo, il contesto di formazione della stessa.

## 1.2. Elementi costitutivi di una Rete Sociale

Per fissare nella mente il concetto di Rete Sociale, partiamo dall'immaginare una semplice rete da pesca, formata da tanti *nodi* collegati tra di loro dai *fili* della rete. Più fili sono presenti nella rete da pesca, più essa è robusta, stessa logica si può declinare alle reti sociali. L'esempio ha lo scopo di fissare nella mente del lettore l'immagine di rete che verrà utilizzata d'ora in avanti.

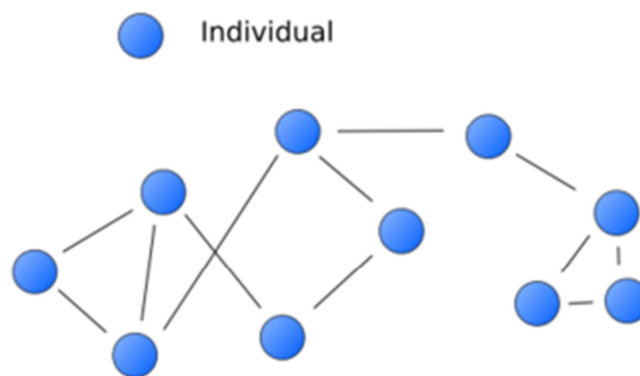
### **I Nodi o Vertici (*Nodes*)**

Ha senso parlare di rete sociale proprio per la presenza di elementi che sviluppano relazioni tra loro di qualsiasi natura esse siano. Soggetti fondamentali sono infatti i *nodi*, coloro che appartengono alla rete, coloro che interagiscono, possono essere anche persone fisiche con un nome e un cognome. Il ruolo dei nodi è fondamentale, senza nodi infatti la rete cesserebbe di esistere, dato che i legami non sono unità indipendenti ma vengono creati dai nodi stessi.

### **Le Relazioni o Archi (*Links*)**

I nodi della rete sono interconnessi tra loro tramite legami che si caratterizzano in monodirezionali (dal punto A al punto B), bidirezionali (dal punto A al punto B, e viceversa), multidirezionali (dal punto A ai punti B, C, e viceversa), questi elementi sono chiamati *link*.

In sintesi ciò che verrà spontaneo immaginare è proprio un fitto reticolo formato da tanti nodi che vengono uniti tra di loro, come indicato nella Figura 1.



**Figura 1 - Esempio di RETE SOCIALE**

### **1.3. Le tipologie di Rete Sociale**

Esistono diverse tipologie di rete sociale, si differenziano per il contesto in cui nascono, per la tipologia di nodi presenti all'interno, per l'estensione, per la difficoltà di accesso e per la capacità di condivisione.

Fin da piccoli, siamo stati immediatamente inseriti all'interno della più semplice rete sociale a disposizione: la *famiglia*.

Certo, nella famiglia, i nodi sono i parenti, l'estensione è limitata, la difficoltà di accesso è elevata e la capacità di condivisione è la più alta tra tutte le reti con un'ampia varietà di argomenti.

La rete sociale formata dai *compagni di scuola* e la rete formata dai *colleghi di lavoro* sicuramente non saranno di facile accesso, non saranno illimitate e sicuramente le unità presenti al loro interno condivideranno principalmente aspetti rispettivamente della vita studentesca e lavorativa.

La rete sociale digitale, *social network*, nella quale siamo entrati per nostra scelta, accesso libero, ha invece estensione illimitata e capacità di consivisione molto elevata soprattutto per la varietà dei temi trattati. È quindi certo come le reti sociali svolgano un ruolo fondamentale nella gestione delle organizzazioni di persone e nella risoluzione di problemi complessi, dove per complessi si intende sia la difficoltà tecnica del problema sia la vastità data dal numero di soggetti coinvolti. Nei capitoli che seguono saranno analizzati i dati provenienti da *Facebook*, il social network più diffuso tra quelli presenti sul web.

#### **1.4. Le analisi descrittive di una Rete Sociale**

Esplorare la rete significa ricavare gli indicatori che ci forniscono la possibilità di interpretare l'oggetto su cui si sta lavorando. Alcuni indici descrittivi sono riferiti al singolo nodo della rete mentre altri si riferiscono all'intera rete, per questo è stata usata la dicitura "*Node Level*" che sta per "a livello di nodo" e "*Network Level*" che sta per "a livello di rete".

##### ***Density*, la densità della Rete Sociale (Network Level)**

La densità della matrice che rappresenta la Rete Sociale è il rapporto tra il numero di collegamenti presenti nella rete e il numero massimo di collegamenti possibili.

$$\Delta = \frac{\sum_{i,j=1}^N y_{i,j}}{N(N-1)}$$

### **Reciprocity, l'indice di reciprocità della Rete Sociale (Network Level)**

La reciprocità è la proporzione dei legami corrispondenti esistenti tra i nodi della Rete Sociale, ricopre un ruolo fondamentale nelle reti in cui i legami *non* sono reciproci per natura dei dati (non in questo). In determinate condizioni è utile quindi verificare il grado di reciprocità. Può essere utilizzato ugualmente nelle reti non orientate per verificare la completa corrispondenza biunivoca tra i nodi.

### **Transitivity, l'indice di transitività (Node Level)**

La transitività all'interno di una relazione è basata sul concetto di *triade*, sono infatti coinvolti 3 attori per poter parlare di transitività come indicato in Figura 2. Esiste transitività nella triade quando, dati tre attori indicizzati da  $i$ ,  $j$  e  $h$ , si verifica che:

- c'è un legame tra  $i$  e  $j$ ;
- c'è un legame tra  $j$  e  $h$ ;
- allora c'è un legame anche tra  $h$  e  $i$ .

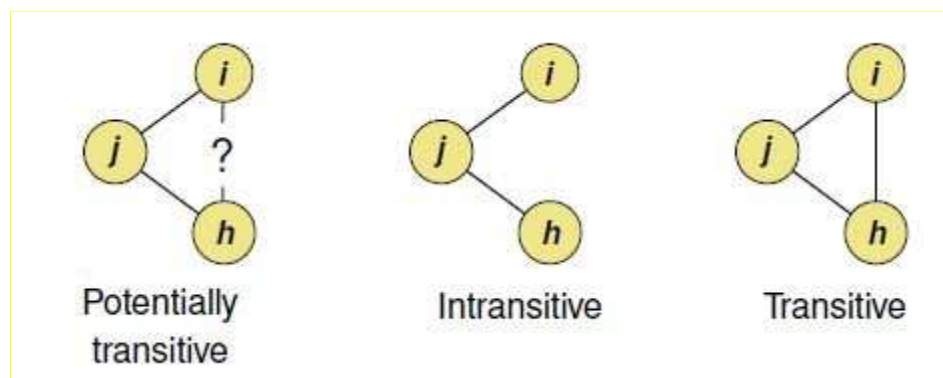


Figura 2 - Esempio di TRANSITIVITY nella triade  $i, j, h$

$$\text{Indice di Transitività} = \frac{N \text{ di triadi transitive} - \text{reali}}{N \text{ di triadi transitive} - \text{potenziali}}$$

L'indice varia da 0 a 1, dove 1 sta ad indicare che il grafo è completamente transitivo. Nelle reti sociali solitamente l'indice di transitività varia tra 0.3 e 0.6 .

### **Diameter, il diametro della Rete Sociale (Network Level)**

Un diametro di rete è la *distanza geodetica*<sup>1</sup> più lunga nella rete, ciò significa che il diametro è la lunghezza del percorso più lungo che collega coppie di nodi.

Per capire praticamente cos'è il diametro di una rete sociale, immaginiamo di essere a Venezia e immaginiamo che per trasferire una lettera tra i due punti più distanti della città sia necessario attraversare 9 ponti differenti che collegano le 10 zone da oltrepassare per arrivare a destinazione.

### **Degree, primo indice di centralità (Node Level)**

Per misurare la centralità della Rete Sociale solitamente sono utilizzati due indicatori principali, *Degree* e *Betweenness*.

$$d_i = \sum_{j=1}^v Y_{ij}$$

Il grado  $d_i$  dell'utente  $i$  si può identificare come il numero dei collegamenti presenti. All'interno di reti asimmetriche si suddivide in *Outdegree* e *Indegree*, rispettivamente il grado in entrata e il grado in uscita. Nel caso di Facebook, per

---

<sup>1</sup> In matematica e, più precisamente in geometria differenziale, una geodetica è la curva più breve che congiunge due punti di uno spazio.

calcolare il grado del soggetto *i-esimo* si risponde alla domanda: “Nella lista di quanti amici compare il nome del soggetto *i-esimo*?”.

### ***Betweenness*, secondo indice di centralità (Node Level)**

Betweenness invece è la media del numero delle volte in cui un attore si trova all'interno di un percorso che congiunge altri 2 nodi.

Misura il controllo che, il nodo *i*, può avere sull'informazione che parte dal nodo *u* e arriva al nodo *v*, a patto che il nodo *i* si trovi all'interno del percorso.

$$d_i = \sum_{u \neq i \neq v} \frac{n_{uv}(i)}{n_{uv}}$$

L'indice di Betweenness del nodo *i* è la somma (eseguita su tutte le coppie (*u,v*) di nodi con  $u \neq v \neq i$ ) dei rapporti tra:

- il numero dei percorsi più brevi tra *u* e *v* che passano attraverso *i*, indicato al *numeratore*;
- il numero dei percorsi più brevi tra *u* e *v*, indicato al *denominatore*.

### ***Assortativity*, indice di omofilia (Network Level)**

L'Omofilia è la tendenza del genere umano a formare legami con persone demograficamente (età, sesso, razza) e psicologicamente (intelligenza, attitudini, aspirazioni, educazione) simili. Anche all'interno delle Reti Sociali è possibile fornire una misura dell'Omofilia, tramite *l'Assortative Mixing* della rete (Newman, 2002) secondo cui i nodi tendono ad avere legami con altri nodi che presentano caratteristiche simili. L'Assortative Mixing può essere di due tipi:



- *Discreto*, i nodi sono classificati in base a *caratteristiche intrinseche* (tipo);
- *Scalare*, i nodi sono classificati in base a *caratteristiche scalari* (grado).

Dopo aver scelto la caratteristica dei nodi che li divide in gruppi, viene creata una tabella contenente le proporzioni di edge distribuiti tra le possibili combinazioni.

	<b>Gruppo 1</b>	<b>Gruppo 2</b>	<b>Gruppo 3</b>
<b>Gruppo 1</b>	$e_{11}$	$e_{12}$	$e_{13}$
<b>Gruppo 2</b>	$e_{21}$	$e_{22}$	$e_{23}$
<b>Gruppo 3</b>	$e_{31}$	$e_{32}$	$e_{33}$
<b>Marginali</b>	$a_1$	$a_2$	$a_3$

Tabella 1 - Divisione in gruppi per il calcolo di ASSORTATIVITY

- $e_{kk}$  è la frazione *reale* di collegamenti tra soggetti appartenenti allo stesso gruppo;
- $a_k$  è il valore *previsto* di collegamenti tra soggetti appartenenti allo stesso gruppo, in una rete con connessioni casuali.

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

L'indice di Assortative Mixing è un valore reale compreso tra -1 e 1; nel caso di  $R=1$  la rete è perfettamente assortita, i collegamenti sono eterogenei, i nodi non sono collegati solo con soggetti simili;

- $R = 1$ , Assortative Mixing perfetto;
- $R = 0$  nessun Assortative Mixing;
- $R < 0$  Disassortative Mixing.

## **2. LA SOCIAL NETWORK ANALYSIS**

### **2.1. Cos'è la Social Network Analysis**

Prima di procedere è doveroso affermare che la popolazione presente sulla terra è organizzata in network, non in gruppi. Il concetto di network sposta l'attenzione dal singolo individuo alla rete di cui l'individuo ne è l'elemento caratterizzante, la sottile differenza si basa sul fatto che oltre ad esprimere le qualità personali nella rete emergono qualità provenienti dall'interazione con altri individui presenti.

La Social Network Analysis<sup>2</sup> rappresenta un insieme di moderne tecniche per l'analisi reti sociali, analizza la rete nel suo complesso dai nodi alle relazioni, spostando il focus sulla visione d'insieme con l'obiettivo di spiegare come le relazioni possano influire sui comportamenti delle unità che vi sono coinvolte.

Negli anni '80 la Social Network Analysis è arrivata anche in Italia, comincia a diffondersi il concetto di rete sociale per spiegare i fenomeni di vita sociale.

Dagli anni '90 l'applicazione delle tecniche SNA si estende a diversi campi, dalle imprese alla politica, dalla famiglia alle migrazioni, per poi approdare nei Social Network, argomento che sarà oggetto di questa tesi. In ognuno di questi casi, l'obiettivo è analizzare il capitale sociale che li caratterizza.

### **2.2. La Teoria del Mondo Piccolo di Milgram**

La *Teoria del Mondo Piccolo* ideata dallo statunitense psicologo Stanley Milgram sostiene che tutte le reti complesse presenti in natura siano tali che qualunque coppia di nodi possa essere collegata tramite un percorso costituito da un numero relativamente piccolo di legami.

---

<sup>2</sup> Anche indicata come SNA, acronimo di Social Network Analysis.

A sostegno di ciò, un grande passo avanti avvenne nel 1967 quando Milgram stesso decise di condurre l'esperimento che avrebbe confermato quelle che erano fino ad allora solamente ipotesi.

### **2.3. L'esperimento di Milgram**

Lo psicologo estrasse in modo totalmente casuale un gruppo di americani residenti nella zona del Midwest degli Stati Uniti e chiese loro di spedire un pacchetto ad un estraneo abitante nel Massachusetts, lontano migliaia di chilometri. Ad ognuno dei selezionati, Milgram comunicò solamente il nome del destinatario, il suo impiego e la zona in cui risiedeva ma non l'indirizzo preciso di residenza. Come indicazione ai partecipanti fu indicato di inviare il pacchetto ad una persona, da loro conosciuta, che a loro intuito avrebbe potuto conoscere con maggior probabilità il destinatario finale. Per arrivare al destinatario la regola prevedeva che ognuno avrebbe selezionato a sua volta un contatto con alta probabilità di conoscere la persona cui il pacco era indirizzato. All'inizio dello studio si pensava servissero almeno cinquanta passaggi, la media invece diceva tutt'altro: da cinque a sette persone avrebbero preso in mano il pacchetto prima di poterlo consegnare al vero proprietario (Milgram & Travers, 1969).

L'esperimento fu poi riproposto nel 2001 con la posta elettronica, e i risultati furono confermati: *effettivamente con 6 passaggi è possibile consegnare il messaggio al destinatario sconosciuto.*

### **2.4. I sei gradi di separazione**

Dalla teoria di Milgram, confermata nell'esperimento precedentemente descritto, prende vita il concetto dei sei gradi di separazione.

L'esperimento ha infatti dimostrato che tra due nodi della rete, A e B, esistono una serie di attori che fungono da distributori utili al collegamento tra i due.

Da questo concetto ne deriva l'idea che ognuno di noi è ha porta aperta sul mondo, in media 6 persone ci distanziano da qualunque altra persona presente sul pianeta, come indicato nella Figura 3.

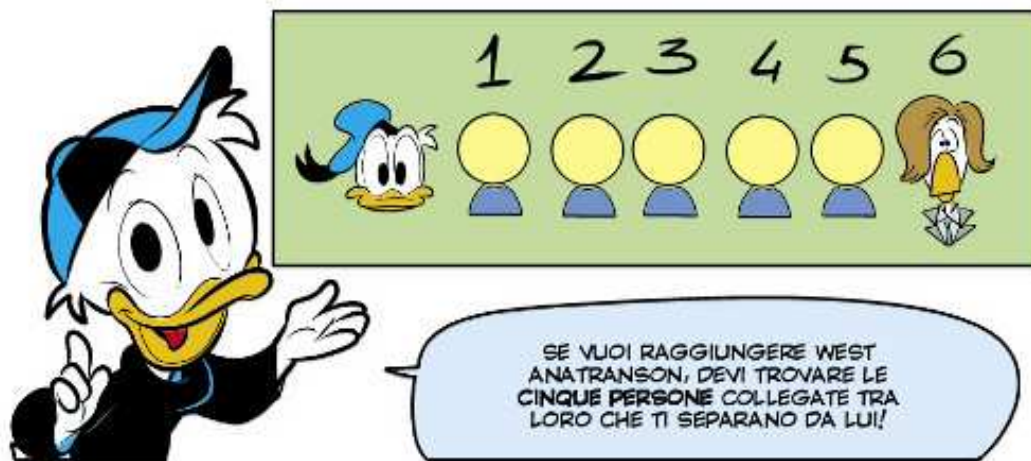


Figura 3 - I sei gradi di separazione

## 2.5. Differenze rispetto all'Approccio Classico

Nel *Classico Approccio* della ricerca sociale, l'attenzione è rivolta alle caratteristiche delle singole unità.

Raccogliendo questo tipo di dati si arriva solitamente ad una matrice rettangolare, nelle *righe* sono posizionate le unità osservate e le *colonne* rappresentano le variabili raccolte.

I dati sono quindi disposti in record, e vanno quindi letti per riga, come indicato nella Tabella 2.

---

*...possiede le seguenti proprietà.*

---

<i>L'attore nella riga...</i>	<b>Età</b>	<b>Sesso</b>	<b>Altezza</b>	<b>X<sub>1</sub></b>
<b>Marco</b>	15	M	185	X <sub>11</sub>
<b>Silvio</b>	26	M	164	X <sub>21</sub>
<b>Laura</b>	54	F	175	X <sub>31</sub>
<b>Dario</b>	37	M	173	X <sub>41</sub>

---

**Tabella 2 - Approccio CLASSICO**

La Social Network Analysis rivolge invece l'attenzione sulle relazioni che coinvolgono le singole unità. Ciò porta alla costruzione di una matrice quadrata, le unità sono posizionate sia sulle *righe* sia sulle *colonne*.

Gli unici valori che contiene la matrice sono 0 e 1, se le due unità sono rispettivamente in relazione o no.

---

*...si relaziona con l'attore della colonna.*

---

<i>L'attore nella riga...</i>	<b>Marco</b>	<b>Silvio</b>	<b>Laura</b>	<b>Dario</b>
<b>Marco</b>	--	1	1	1
<b>Silvio</b>	0	--	0	1
<b>Laura</b>	1	0	--	1
<b>Dario</b>	1	0	0	--

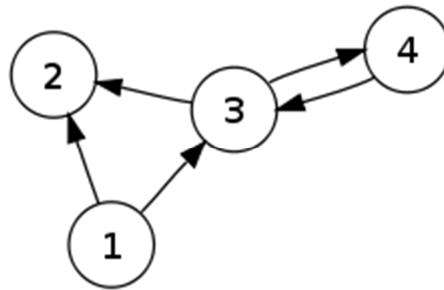
---

**Tabella 3 - Approccio SNA (Matrice di Adiacenza)**

Si noti che la diagonale non contiene valori ciò deriva dalla convenzione che un individuo non può interagire con se stesso.

La matrice nella Tabella 3 è detta *Matrice di Adiacenza*, o anche *Matrice delle Connessioni* utilizzata per la rappresentazione di un grafo che rappresenta la rete come visto in Figura 1.

Per fissare graficamente il concetto, si faccia riferimento al grafo in Figura 4.



**Figura 4 - GRAFO orientato con 4 nodi**

La corrispondente rappresentazione matriciale tramite la matrice di adiacenza è indicata di seguito.

$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$





## 3. I DATI GREZZI

### 3.1. Il profilo Facebook

Oramai quasi tutti dispongono di un profilo Facebook, il più noto social network tra quelli che il mercato offre, si veda Figura 5.

Facebook raccoglie e salva ogni nostra operazione digitale immagazzinando per esempio tutte le fotografie che vengono caricate, tutti i messaggi che inviamo e riceviamo, tutte le richieste di amicizia inviate, ricevute, accettate e rifiutate.

Queste semplici informazioni sono sufficienti per delineare il profilo social della persona che lo utilizza.

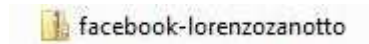


Figura 5 - Pagina iniziale FACEBOOK

### 3.2. Struttura dei dati in cartelle

Il download dei dati avviene direttamente dal profilo Facebook, nell'apposita sezione all'interno della voce *Impostazioni*.

Il motore gestionale utilizzato da Facebook crea un file compresso *.zip*, come indicato nella Figura 6.



**Figura 6 - Cartella.zip**

All'interno del file sono presenti 4 elementi:

- la cartella *html*;
- la cartella *photos*;
- la cartella *videos*;
- il file *index*.



**Figura 7 - Il contenuto del file Facebook**

### **La cartella *html***

La cartella *html* contiene i file *.html*<sup>3</sup> presenti nella Figura 8, i quali al loro interno hanno gli effettivi dati grezzi da analizzare.

---

<sup>3</sup> Acronimo di HyperText Markup Language, comunemente chiamate pagine web.

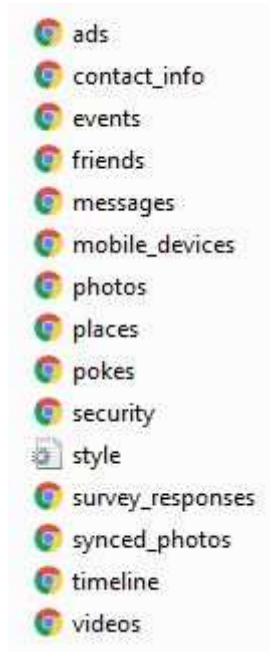


Figura 8 - La cartella HTML

### La cartella *photos*

La cartella *photos* contiene altre sottocartelle in cui sono organizzate tutte le foto che sono state caricate nel profilo dalla data della sua creazione alla data del suo download. Inoltre è presente il file *profile* che è l'ultima foto profilo utilizzata prima di scaricare i dati come è possibile osservare nella Figura 9.

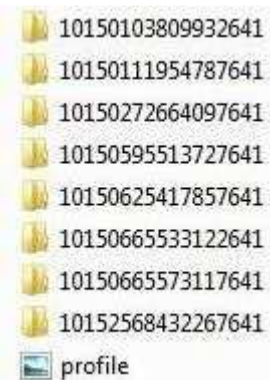
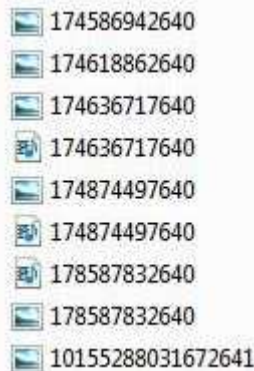


Figura 9 - La cartella PHOTOS

### **La cartella *videos***

Anche la cartella *videos* contiene la lista dei video che sono stati caricati nel profilo dalla data della sua creazione alla data del suo download, ben organizzati in cartelle, come si vede nella Figura 10.



**Figura 10 - La cartella VIDEOS**

### **Il file *index***

Il file *index* è file di indice da utilizzare come mappa per accedere alle informazioni contenute nelle cartelle viste precedentemente. In termini pratici il file *index* è la maschera che vediamo quando eseguiamo l'accesso al nostro profilo Facebook alla voce *Il mio Profilo*. Il file *index* è strutturato come si può vedere nella Figura 11.



## Lorenzo Zanotto

### Profilo

Info di contatto

Diario

Foto

Foto sincronizzate

Video

Amici

Messaggi

Poke

Eventi

Sicurezza

Inserzioni

Dispositivi mobili

Profilo	<a href="http://www.facebook.com/lorenzo.zanotto">http://www.facebook.com/lorenzo.zanotto</a>
E-mail	pollo_lollo_@hotmail.it
Data di registrazione	Sabato 6 settembre 2008 alle ore 14:23 UTC+02
Data di nascita	11/25/1989
Genere	Uomo
Città in cui ti trovi adesso	Padova
Città natale	Rovigo
Situazione sentimentale	Impegnato con Elena Previato dal 8 ottobre 2011
Familiari	Francesco Zanotto (Fratello) Enrico Montagnana (Cugino) Romeo Zanotto (Padre)
Istruzione	itis F.Viola - Rovigo '08 2008 ITIS "F.Viola" Rovigo 2008 Scienze statistiche University of Padua

Figura 11 - La pagina INDEX

### 3.3. La libreria myFBr per RStudio

Per l'analisi dei dati è stato utilizzato il software statistico *R Studio*, per poter fare ciò si sono dovute leggere le informazioni dai file html e strutturarle all'interno di dataframe Rdata. Per l'importazione dei dati in R Studio è stata utilizzata la libreria *myFBr* creata dal *Prof. Finos* in collaborazione con la *Dott.ssa Marino*, i colleghi *Davide Meneghetti* e *Simone Righetto* (Finos, et al., 2017). MyFBr dispone di alcune funzioni create ad-hoc per la lettura dei file .html prodotti da Facebook, elencate di seguito. All'importazione dei dati segue il processo di creazione dei dataframe *Dataset Utenti* e *Matrice di Adiacenza*, per fare ciò sono state create apposite funzioni interfacciate alla libreria *myFBr*.

### **La funzione *getInfo***

Permette di estrarre le informazioni di base contenute nel profilo, ossia:

- Id profilo Facebook;
- Email;
- Data di registrazione a Facebook;
- Data di scaricamento dei dati Facebook;
- Sesso;
- Data di nascita;
- Luogo di nascita;
- Situazione sentimentale.

### **La funzione *getFriends***

Permette di estrarre le informazioni riguardanti le amicizie, ossia:

- Numero degli amici accettati;
- Numero delle richieste di amicizia effettuate;
- Numero delle richieste di amicizia ricevute;
- Numero degli amici rimossi.

## 4. IL DATASET UTENTI

### 4.1. Dataset Utenti

Il dataset utenti è un insieme di dati strutturati in forma relazionale, può essere vista come una grande tabella in cui:

- ogni *colonna* rappresenta una variabile;
- ogni *riga* rappresenta un'osservazione.

Va specificato che, in questo caso, con il termine “osservazione” si intende uno dei 359 profili Facebook raccolti. In questo tipo di dataset il numero delle colonne solitamente è inferiore al numero delle righe, la quantità di *variabili* raccolte non dipende dal numero di *osservazioni* raccolte. Le variabili previste per il dataset utenti sono elencate nell'elenco puntato di seguito.

- Chiave univoca<sup>4</sup>, per mettere in relazione i diversi dataset;
- Sesso del soggetto;
- Il numero della Wave in cui è stato raccolto il profilo;
- I cinque elementi che compongono la scala di *Personalità*;
- I cinque elementi che compongono la scala di *Benessere*;
- Il punteggio di *Autostima*;
- Il *Degree* e l'indice *Betweenness*;

---

<sup>4</sup> Anche citata come *Primary Key*, serve a identificare univocamente le righe del dataset in cui è definita. Può essere specificata una sola volta per ogni dataset e due righe distinte non possono avere lo stesso valore sul campo scelto come chiave primaria.

## 4.2. La scala di *Personalità*

Descrivere e codificare la personalità individuale risulta essere di notevole importanza per la spiegazione delle relazioni che intercorrono tra due o più soggetti, come nel caso di questo lavoro. A partire da appositi studi psicolessicali somministrati a campione, tramite il famoso *Big Five Questionnaire*<sup>5</sup> in cui è stato possibile codificare tutte le esperienze raccolte e ritenute significative per la descrizione della personalità umana, sono stati ricavati i cosiddetti *Big Five*<sup>6</sup>, cinque fattori principali che individuano le *cinque macro-categorie* utilizzate più di frequente dai soggetti coinvolti nello studio per descrivere le diversità tra individui (Dallago & Roccato, 2010).

Cinque dimensioni, elencate di seguito, che rappresentano quindi una descrizione completa e parsimoniosa della personalità, applicabile alle diverse culture che prese sempre più consensi dagli anni sessanta in poi.

- *Estroversione* vs *Introversione*;
- *Gradevolezza* vs *Ostilità*;
- *Consapevolezza* vs *Inconsapevolezza*;
- *Stabilità Emotiva* vs *Nevroticismo*;
- *Apertura Mentale* vs *Chiusura Mentale*;

## 4.3. La scala di *Benessere*

Tramite lo studio svolto in collaborazione con la *University of California, Santa Barbara*, sono stati raccolti una serie di dodici indicatori specifici che formano in totale *quattro sezioni generali* del primo ordine, elencate di seguito e specificate nella Figura 12 (Furlong, et al., 2013).

---

<sup>5</sup> Anche indicato come BFQ, acronimo di Big Five Questionnaire.

<sup>6</sup> Anche indicato con FFA, acronimo di Five Factor Approach oppure FFM, acronimo di Five Factor Model.



- *Belief-In-Self*, credere in se stessi;
- *Belief-in-Others*, credere negli altri;
- *Emotional Competence*, competenza emotiva;
- *Engaged Living*, impegno.

All'interno del SEHS<sup>7</sup> si ruota attorno al concetto di *Covitality*, termine utilizzato per descrivere una serie di comportamenti e aspetti della persona che rendono la vita migliore, producono benessere necessario alla sopravvivenza a scuola, sul posto di lavoro e nella vita di tutti i giorni.



Figura 12 - Scala di Benessere

---

<sup>7</sup> È l'acronimo di Social-Emotional Health Surveys, l'indagine sulla salute sociale ed emotiva.



## 5. LA MATRICE DI ADIACENZA

### 5.1. Tipologia

Il secondo dataset costruito è la *Matrice di Adiacenza*, vista al *Paragrafo 2.5* composta da  $n$  righe e  $n$  colonne. Proprio per il fatto che i dati provenienti da una rete sociale formata da  $n$  soggetti indicano la relazione esistente tra tutte le possibili coppie  $i,j$  con  $i,j=1\dots n$ , utilizzeremo la terminologia  $y_{i,j}$  per indicare la relazione presente tra il soggetto  $v_i$  e il soggetto  $v_j$ . La matrice di adiacenza considera il caso di *presenza* o *assenza* della relazione tra i soggetti che lo compongono. La codifica utilizzata è la seguente:

- 0, assenza di relazione tra soggetti  $i$  e  $j$  con  $i \neq j$ ;
- 1, presenza di relazione tra soggetti  $i$  e  $j$  con  $i \neq j$ .

### 5.2. Modalità di costruzione

La costruzione della matrice di adiacenza è avvenuta da sinistra verso destra, dall'altro verso il basso, quindi con compilazione per righe. La logica di costruzione, avendo i dati Facebook, si basa su un semplice quesito:

*Avendo la lista delle amicizie accettate del soggetto  $i$ , il soggetto  $j$  (con  $i \neq j$ ) è contenuto in questa lista?*

- $y_{i,j} = 1$  se la risposta è SI;
- $y_{i,j} = 0$  se la risposta è NO.

E così percorrendo tutte le 359 righe, controllando le amicizie di tutti i 359 profili, con la funzione *Dataset.Data2* indicata di seguito.

```
DATASET.DATA2 <- function(input_DATA2)
{
  matrice_temp <- matrix(ncol = nrow(input_DATA2), nrow =
nrow(input_DATA2))

  for(i in 1:nrow(input_DATA2))
  {
    percorso=paste("D:/Lorenzo/2017 - TESI MAGISTRALE/PROFILI
FACEBOOK/",input_DATA2[i,1],sep = "")
    perE=paste(percorso,"/html/friends.htm",sep = "")
    pg=htmlParse(perE)#lettura intero file
    nomi=getNodeSet(pg,"//div[h2[text()='Amici']]/ul/li/text()")
    lunghezza=length(nomi)

    listanomi_temp <- list()

    for(j in 1:lunghezza)
    {
      listanomi_temp[j]=toString.XMLNode(nomi[[j]])
      listanomi_temp[j]=gsub(" $", "", listanomi_temp[j])
    }

    for(k in 1:nrow(input_DATA2))
    {
      if((input_DATA2[k,3] %in% listanomi_temp) == "TRUE")
      {
        matrice_temp[i,k]=1
      }
      else matrice_temp[i,k]=0
    }
  }

  dimnames(matrice_temp)<-list(NULL,1:nrow(listaCartelle_ID))
  return(matrice_temp)
}
```

### 5.3. Le problematiche incontrate

Data la complessità e le differenze temporali di download tra i profili provenienti da Facebook, sono state introdotte delle regole in fase di costruzione della matrice di adiacenza allo scopo di non incorrere in problematiche particolari durante le analisi che seguiranno.

### ***La diagonale***

Subito dopo aver creato la matrice emergeva un dato inaspettato, la diagonale della matrice di adiacenza conteneva in 33 casi valore 1. In relazione a quanto specificato precedentemente, il valore 1 nella cella  $y_{i,j}$  con  $i=j$  significa che il soggetto  $v_i$  e il soggetto  $v_j$ , oltre a coincidere, presentavano una relazione. Ha espresso che tra le amicizie del soggetto  $v_i$  è presente un collegamento che ha il suo stesso nome. In tutti i casi si tratta di omonimi che è facile trovare in Facebook e con i quali è molto probabile stringere amicizia proprio per una simpatica questione di omonimia. Si è deciso quindi di rimediare forzando i valori della diagonale a 0 come indicato nella teoria riguardante la matrice di adiacenza vista precedentemente.

### ***L'amico del suo non amico***

L'altro dato che è balzato all'occhio dopo aver costruito la matrice, riguarda sempre il soggetto  $v_i$  e il soggetto  $v_j$ . In particolare il valore  $y_{i,j}$  risulta essere 1, quindi il soggetto  $j$  è presente nella lista degli amici del soggetto  $i$ . La coppia invece non risulta legata in senso contrario, nel senso che il valore  $y_{j,i}$  (indici invertiti) risulta essere 0, quindi il soggetto  $i$  non è presente nella lista degli amici del soggetto  $j$ . Da notare il fatto che, quando due soggetti diventano amici tramite Facebook, ognuno compare nella lista degli amici accettati dell'altro, è quindi impossibile che si verifichi una situazione di questo tipo a meno di alcune situazioni particolari indicate di seguito.

- I due profili sono stati scaricati in *momenti differenti*, ad esempio il soggetto  $v_i$  ha effettuato il download delle informazioni dopo essere diventati amici mentre il soggetto  $v_j$  prima di diventare amici, quindi realmente i due soggetti hanno una relazione ma nei nostri dati non è presente, o meglio è presente solo per metà;

- Potrebbe essersi verificato un caso analogo a quello della diagonale, il soggetto  $v_i$  è amico del soggetto  $v_j$  che però non è presente nella lista dei 359 profili raccolti, risulta quindi essere un *omonimo* esterno alla lista.

La verità è che in nessuno dei due casi elencati abbiamo la certezza dei fatti, si è quindi deciso di portare a 0 sia il valore di  $y_{i,j}$ , sia il valore di  $y_{j,i}$  per evitare di inciampare in qualsiasi tipo di errore. A questo punto si può affermare che la matrice di adiacenza è *simmetrica*.

## 6. ANALISI DESCRITTIVE DELLA RETE

Come indicato nel *Capitolo 1 al Paragrafo 1.4* esplorare la rete significa esplorare la matrice di adiacenza ricavando gli indicatori che diano la possibilità di interpretare l'oggetto su cui si sta lavorando. Di seguito, in Figura 13 l'immagine della rete nel suo complesso, si notano i nodi di colore verde e giallo mentre i collegamenti tra i nodi di colore grigio. I profili esterni, disposti in posizione circolare non presentano alcun legame con il resto dei profili raccolti.

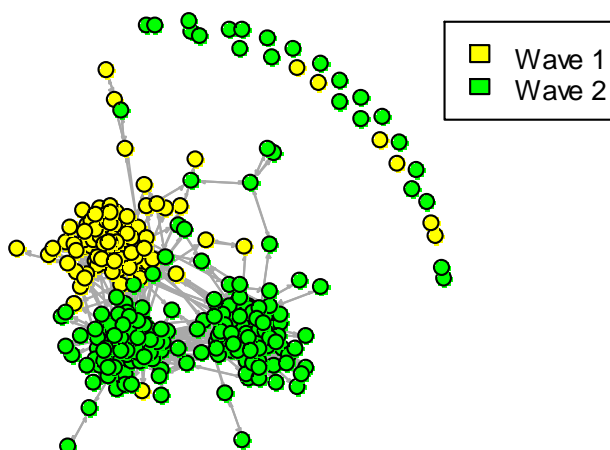


Figura 13 – Il grafico della RETE divisa per WAVE

Per un riassunto delle misure riguardanti le analisi descrittive si veda la Tabella 4 di seguito.

	Node Level	Network Level
<i>Density</i>		0.039
<i>Reciprocity</i>		1.000
<i>Transitivity</i>		0.400
<i>Diameter</i>		9
<i>Degree</i>	Si veda la Figura 15	
<i>(CENTRALITY)</i>		
<i>Betweenness</i>	Si veda la Figura 17	
<i>(CENTRALITY)</i>		
<i>Assortativity</i>		0.645
<i>(HOMOPHILY)</i>		

**Tabella 4 - Analisi descrittive della RETE**

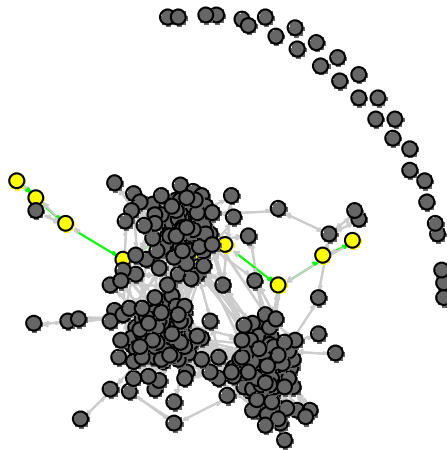
La matrice non è particolarmente densa di collegamenti tra i soggetti partecipanti, questo primo dato risulta molto interessante proprio perché guardando la Figura 13 si potrebbe pensare il contrario, essendo il disegno apparentemente fitto di collegamenti rispetto all'indice *Density* molto inferiore ad 1.00 che è il valore massimo.



Il comando *Reciprocity* restituisce il valore 1.00 come da aspettative, ciò sta ad indicare che la matrice è perfettamente simmetrica e quanto detto al *Paragrafo 5.3* è perfettamente confermato.

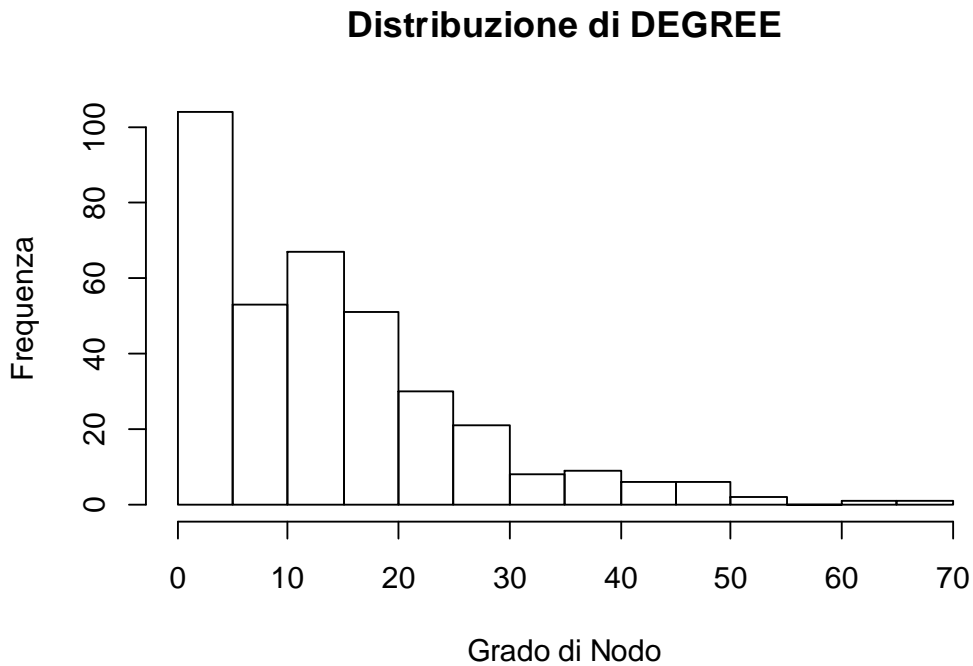
L'indice *Transitivity* varia da 0.00 a 1.00, dove 1.00 sta ad indicare che il grafo è completamente transitivo. Nelle reti sociali orientate solitamente l'indice di transività varia tra 0.30 e 0.60 .

Il *Diameter* indica quanti sono i collegamenti tra i due punti più lontani della rete, inoltre il comando *get\_diameter* indica quali sono i 10 attori coinvolti nel percorso più lungo, specificandone il numero del nodo associato. Nella Figura 14 sono indicati i 10 punti evidenziati in giallo, anche se non si riescono ad individuare tutti perfettamente essendo la figura molto fitta soprattutto nella parte centrale.



**Figura 14 - Nodi coinvolti nel DIAMETRO della RETE**

Il *Degree* di ciascun nodo, anche indicato come grado di notorietà all'interno del network, è uno dei due indici principali per analizzare la centralità della rete sociale. I dati sono riportati in Figura 15. Come si può notare, buona parte dei nodi presentano un grado inferiore a 10, mentre pochissimi utenti presentano un grado superiore al valore di 60 collegamenti.



**Figura 15 - Distribuzione di DEGREE**

C'è la necessità di capire meglio come si è modificato il grafico durante entrambe le *Wave* in cui sono stati raccolti i dati, è stato creato il boxplot diviso per i due periodi. Si faccia riferimento alla Figura 16.

### Valori di DEGREE per WAVE

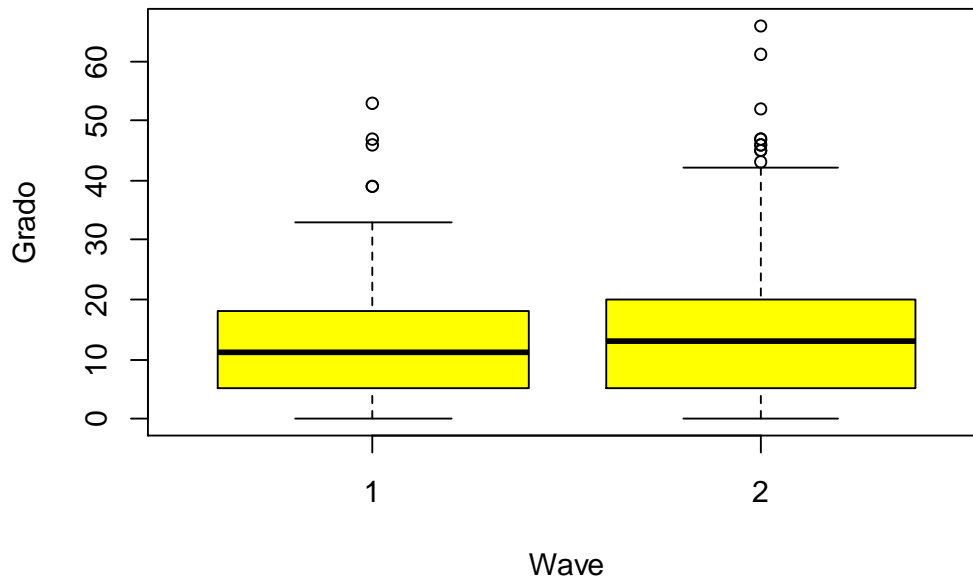


Figura 16 - Valori di GRADO per WAVE

La mediana (valore che sta al 50% percento delle osservazioni) coincide anche se nella *Wave 2* sembra esserci più di qualche soggetto che ha un grado molto elevato (outlayer), oltre 60 collegamenti personali. Da notare che, nella *Wave 2*, più di una osservazione esce dal valore adiacente superiore.

Il secondo indice che da informazioni riguardo la centralità della rete, come detto in precedenza, è *Betweenness* i cui dati sono rappresentati nel Figura 17.

## Distribuzione di BETWEENNESS

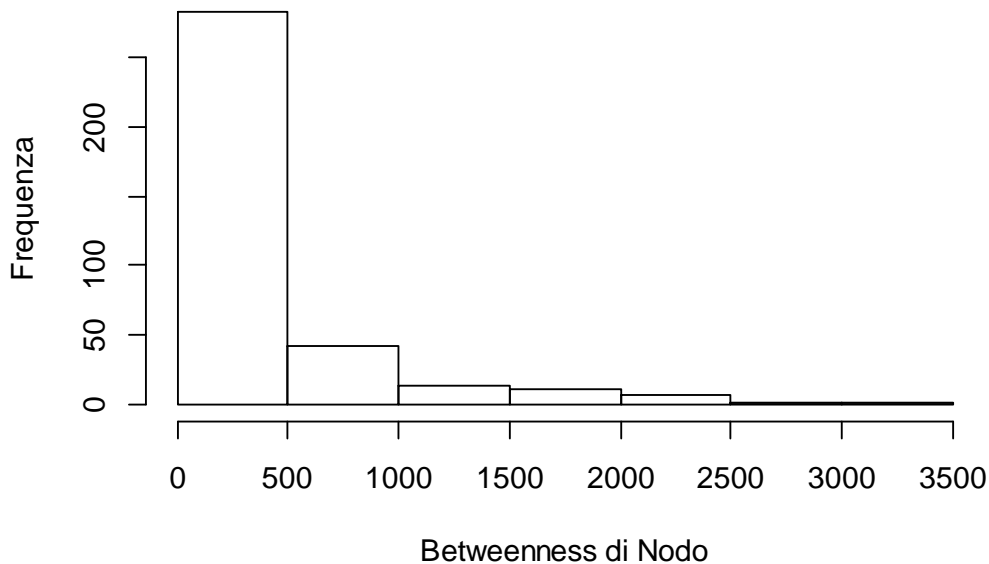


Figura 17 - Distribuzione di BETWEENNESS

Come visto per Degree, la maggior parte degli user presentano un valore di Betweenness inferiore a 1000, mentre solo pochi soggetti possiedono valori che superano il rapporto di 2500.

Significa che sono presenti all'interno della rete delle figure che possiamo definire *centrali* proprio per il ruolo di accentrimento e smistamento delle informazioni a livello di network.

È utile capire come si dividono i valori di Betweenness nei due periodi in cui sono stati raccolti i dati, ciò è indicato nel Figura 18 attraverso il boxplot diviso per *Wave*.

### Valori di BETWEENNESS per WAVE

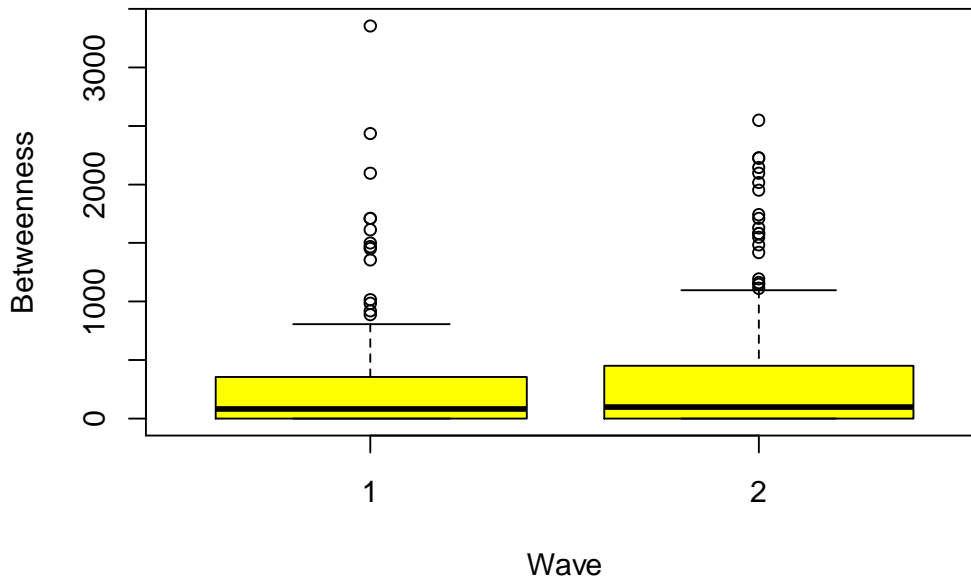


Figura 18 - Valori di BETWEENNESS per WAVE

La distribuzione di Betweenness divisa per *Wave1* e *Wave2* indica valori anomali in entrambi i periodi, inoltre nella *Wave1* viene evidenziato il valore anomalo oltre 3000, il quale sicuramente ricoprirà il ruolo centrale nella rete. Infine l'indice R di *Assortativity Mixing* calcolato utilizzando come variabile di classificazione il *Corso di Studio*, con il valore di 0.645 indica un buon mixing dei collegamenti presenti rispetto al corso di studio frequentato; più precisamente è possibile dire che i nodi all'interno della rete sono collegati eterogeneamente anche con nodi che non frequentano lo stesso corso di studio.



## 7. ANALISI DESCRITTIVE DELLE VARIABILI

Prima di dare inizio all'analisi descrittiva delle variabili esplicative è giusto conoscerle e imparare il loro significato, come indicato nell'elenco che segue.

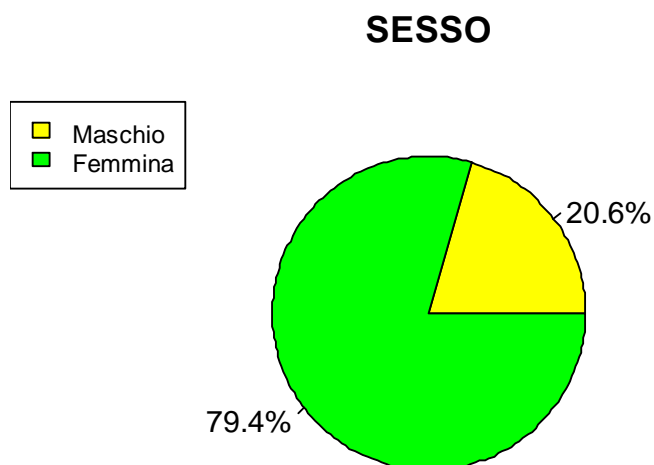
- *Sesso*, maschio o femmina;
- *Wave*, 1 o 2 durante la quale il profilo è stato raccolto;
- *Corso di Studio*, al quale appartiene il soggetto;
- *Degree*, primo indice di centralità (Node Level);
- *Betweenness*, secondo indice di centralità (Node Level);
- *Autostima*, indice di autostima personale;
- *H\_openess*, indice di apertura mentale (BIG5);
- *H\_emotional*, indice di stabilità emotiva (BIG5);
- *H\_continuos*, indice di consapevolezza (BIG5);
- *H\_extraversion*, indice di estroversione (BIG5);
- *H\_agreebleness*, indice di gradevolezza (BIG5);
- *I\_belief\_in\_self*, indice di quanto una persona crede in stessa (SEHS);
- *I\_belief\_in\_others*, indice di quanto una persona crede negli altri (SEHS)
- *I\_emotional\_competence*, indice di quanto una persona riesce a gestire le proprie emozioni (SEHS);
- *I\_engaged\_living*, indice di positività della vita di una persona (SEHS);
- *I\_total*, indice totale dei precedenti 4 indici.

Facendo riferimento all'elenco come traccia operativa per l'analisi descrittiva delle variabili esplicative, si inizia dal Sesso, con i dati in Tabella 5 per le quantità numeriche.

	Femmine	Maschi
Frequenza	285	74

**Tabella 5 - Variabile esplicativa SESSO**

Osservando il Figura 19, è immediato notare che la presenza femminile è superiore a quella maschile di almeno 3 volte.



**Figura 19 - Grafico a torta di SESSO con percentuali**

Si è scelto l'approccio visivo alla rete nel suo complesso insieme di nodi e legami anche per la prossima variabile; per meglio individuare le proporzioni si è scelto



di colorare i vertici con i 3 colori che rappresentano il *Corso di Studio* cui appartiene ogni soggetto. Si faccia riferimento alla Figura 20

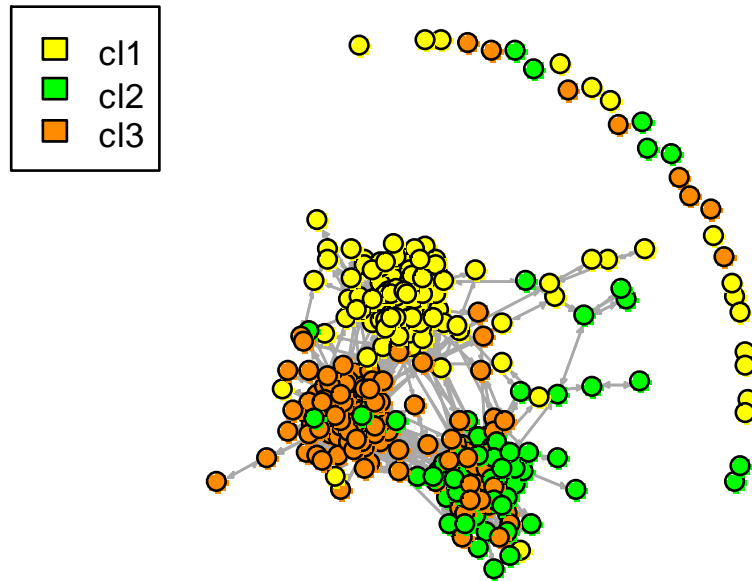


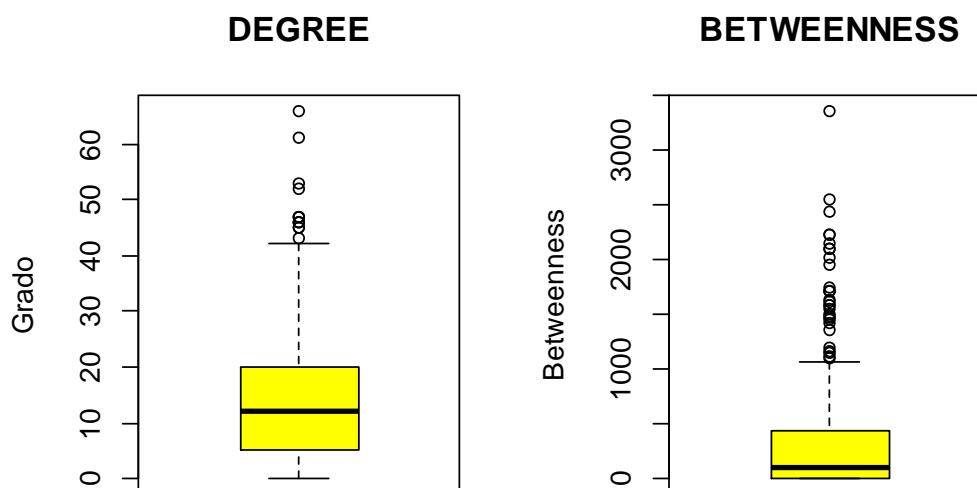
Figura 20 – RETE colorata per CORSO DI STUDIO

Per i valori numerici si faccia riferimento alla Tabella 6.

	cl 1	cl 2	cl 3
<b>Frequenza</b>	120	87	152

**Tabella 6 – Frequenza di CORSO DI STUDIO**

Si prosegue l'analisi descrittiva delle variabili esplicative con i due indici di centralità della rete, *Degree* e *Betweenness*. Si faccia riferimento alla Figura 21 per i boxplot marginali.



**Figura 21 - Boxplot di DEGREE e BETWEENNESS**

Oltre alla visione grafica tramite boxplot si è ritenuto opportuno specificare in Tabella 7 i dati numerici. Da sottolineare la differenza di punteggio *Betweenness* tra il 435 del terzo quartile (75% delle osservazioni) e il 3367 che rappresenta il massimo valore tra quelli calcolati con i dati a disposizione.

	Min	25%	50%	Media	75%	Max
<b>Degree</b>	0	5	12	14	20	66
<b>Betweenness</b>	0	10	102	330	435	3367

Tabella 7 - Quantili di DEGREE e BETWEENNESS

A questo punto, essendoci un'importante disparità di osservazioni tra maschi e femmine è giusto analizzare *Degree* e *Betweenness* in relazione al *Sesso*, come indicato nel Figura 22.

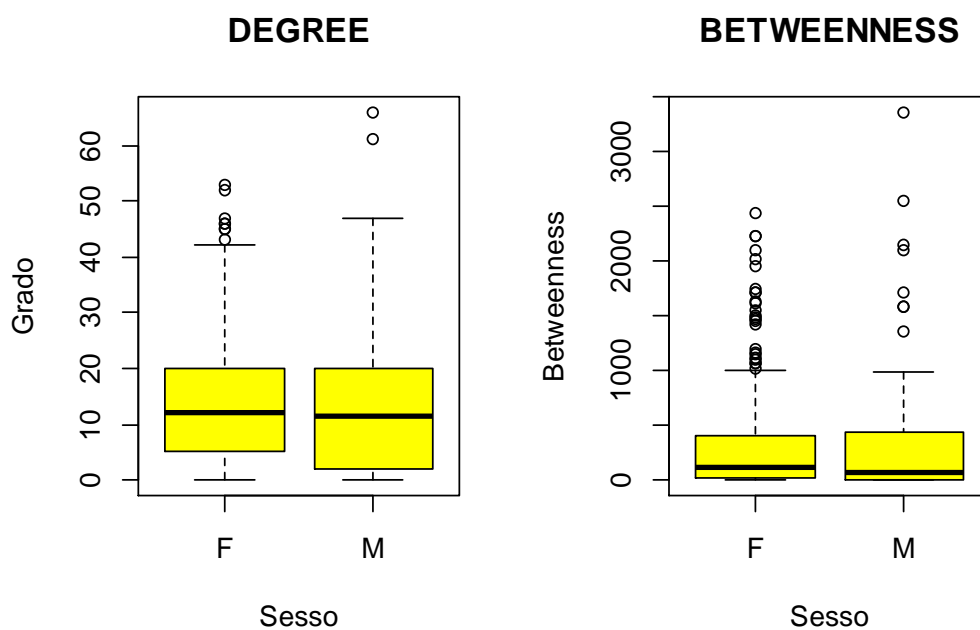


Figura 22 - Boxplot di GRADO e BETWEENNESS separati per SESSO

Dato il boxplot è possibile riconoscere un valore della mediana simile tra maschi e femmine in entrambe le variabili, il primo quantile per i maschi è leggermente

inferiore rispetto alle femmine, ma sostanzialmente le distribuzioni per sesso sono simmetriche alla mediana per Degree con la presenza di alcuni outlier oltre il valore adiacente superiore, che è giusto citare. Per quanto riguarda Betweenness invece le distribuzioni sembrano sbilanciate verso destra, come è possibile vedere dalla Figura 22 la mediana non si posiziona a metà tra il primo e il terzo quartile, è tracciata molto vicino al primo quartile sia per i maschi sia per le femmine. Altra interpretazione di *Degree* e *Betweenness*, utile per l'indice di *Omofilia* calcolato al *Capitolo 6*, prevede la creazione di un boxplot per ogni *Corso di Studio* come indicato in Figura 23 e in Figura 24.

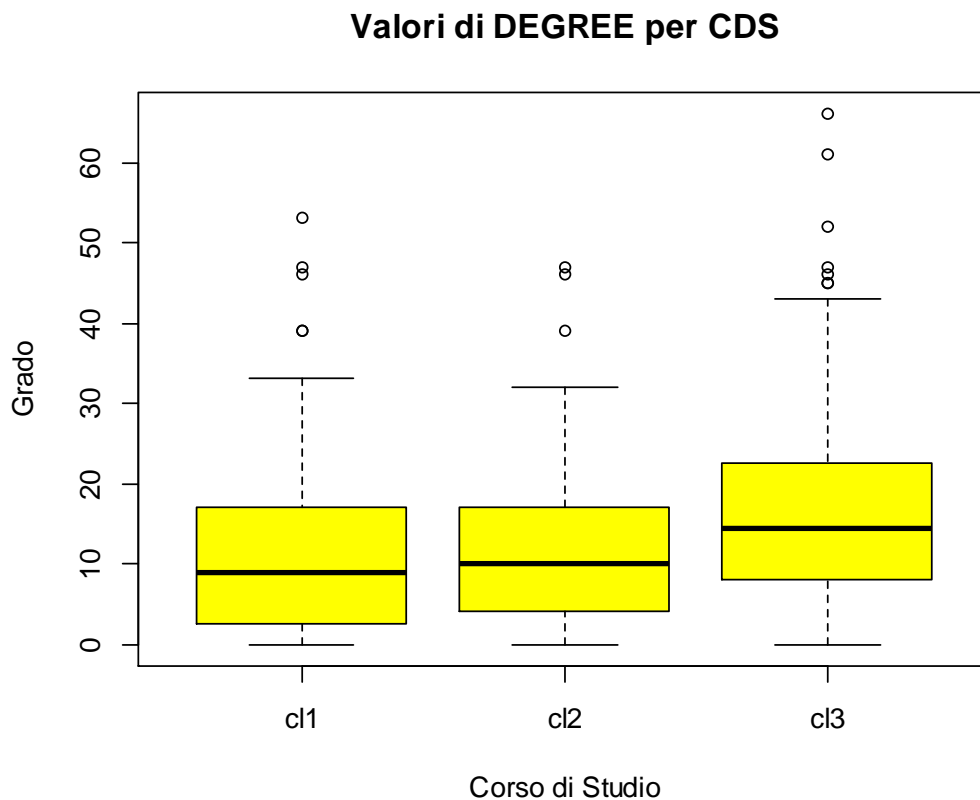
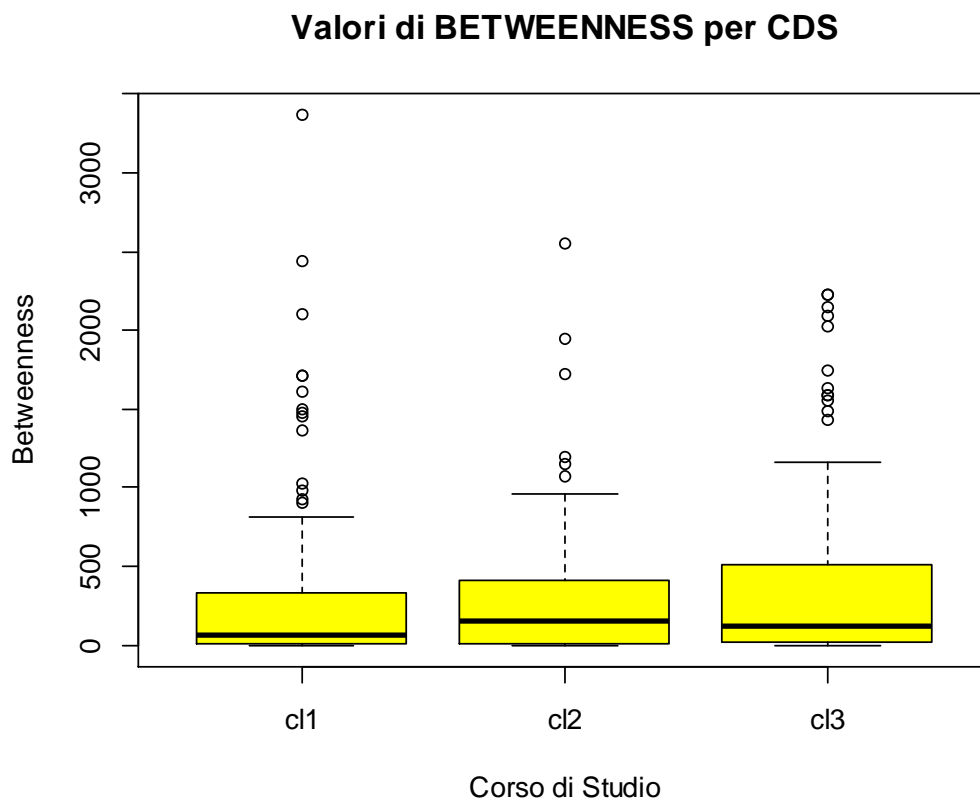


Figura 23 - Boxplot di DEGREE separati per CORSO DI STUDIO

La distribuzione del grado nei 3 corsi di studio è simmetrica rispetto alla mediana, per i primi due corsi è molto simile mentre per il “cl3” i valori di primo e terzo quartile sono traslati più in alto anche se la distribuzione è praticamente identica a meno di outlier posizionati oltre il valore adiacente superiore.



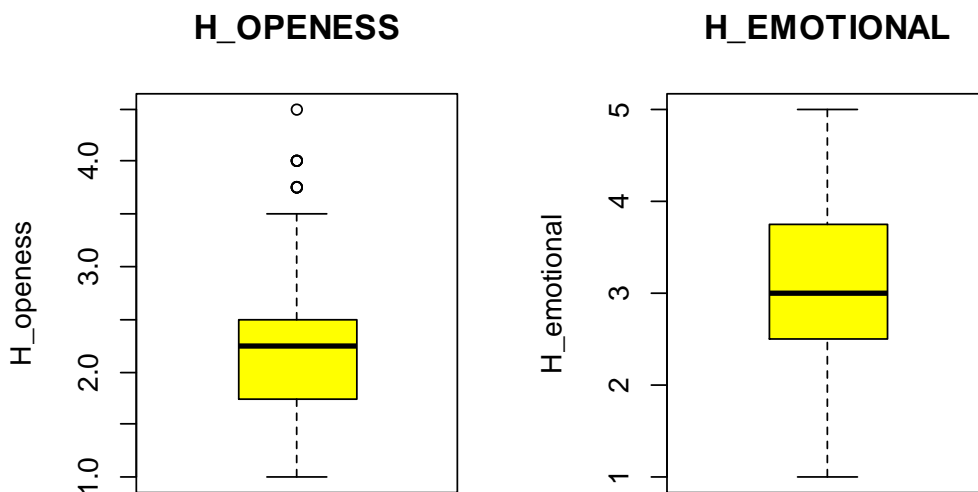
**Figura 24 - Boxplot di BETWEENNESS separati per CORSO DI STUDIO**

La distribuzione di Betweenness nei 3 corsi di studio non è simmetrica rispetto alla mediana, man mano che si passa da “cl1” a “cl3” i valori di primo e terzo quartile si allontanano tra di loro e, in particolare, si allontana anche il valore del terzo quartile dalla mediana sbilanciando via via la distribuzione di Betweenness verso destra. Da citare la presenza di outlier posizionati oltre il valore adiacente superiore.

	<b>Betweenness</b>
<b>Degree</b>	0.64

**Tabella 8 - Correlazione tra DEGREE e BETWEENNESS**

Si faccia riferimento alla Tabella 8 per la correlazione presente tra i due indici di centralità della rete a livello nodo, il valore di 0.64 è da tenere in considerazione durante la scelta delle covariate da includere nel modello essendo il valore superiore 0.5. Interessante osservare come si distribuiscono anche le altre variabili esplicative, saranno descritte prima le variabili appartenenti al gruppo dei *Big Five* e successivamente le variabile ricavate dal *SEHS*. Di seguito una prima interpretazione visiva per la distribuzione dei punti sul piano cartesiano. Vengono riportati di seguito i boxplot per la scala dei *Big Five* con la distribuzione marginale, si faccia riferimento a Figura 25, Figura 26 e Figura 27. Successivamente sarà riportata anche una tabella riassuntiva con i valori nei 4 quantili e il valore medio, si veda Tabella 9.



**Figura 25 - Boxplot di H\_OPENESS e H\_EMOTIONAL**

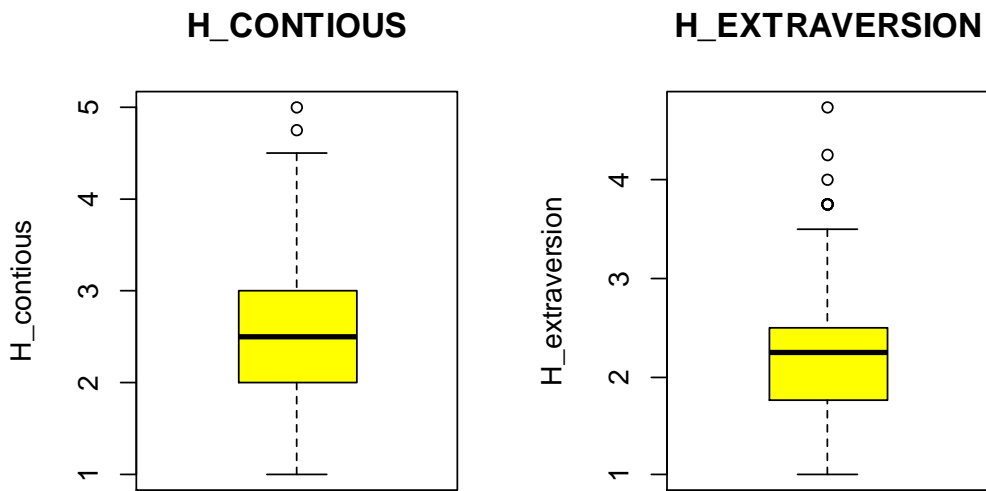


Figura 26 - Boxplot di H\_CONTIOUS e H\_EXTRAVERSION

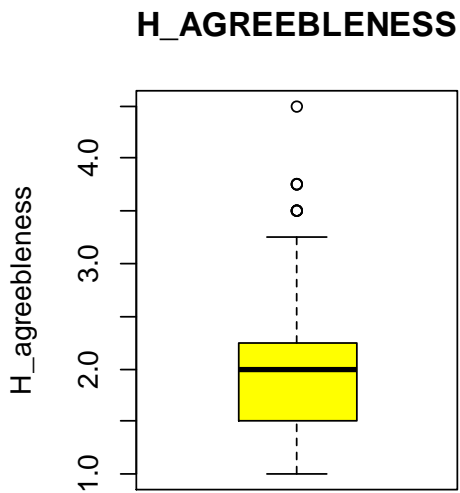


Figura 27 - Boxplot di H\_AGREEBLENESS

Le distribuzioni delle cinque variabili appartenenti alla scala H cosiddetta “*dei Big5*” risultano piuttosto simmetriche a meno di outlier che non hanno valori particolarmente distanti riseptivamente da primo e terzo quartile. Si faccia riferimento alla Tabella 9 per la specifica dei valori numerici.

	<b>Min</b>	<b>25%</b>	<b>50%</b>	<b>Media</b>	<b>75%</b>	<b>Max</b>
<b>H_openess</b>	1.00	1.75	2.25	2.20	2.50	4.50
<b>H_emotional</b>	1.00	2.50	3.00	3.05	3.75	5.00
<b>H_contious</b>	1.00	2.00	2.50	2.54	3.00	5.00
<b>H_extraversion</b>	1.00	1.75	2.25	2.23	2.50	4.75
<b>H_agreebleness</b>	1.00	1.50	2.00	1.94	2.25	4.50

**Tabella 9 - Quantili della Scala H (BIG FIVE)**

Si analizza ora la correlazione tra le 5 entità, si faccia riferimento alla Tabella 10.

	<b>H_openess</b>	<b>H_emotional</b>	<b>H_contious</b>	<b>H_extraversion</b>	<b>H_agreebleness</b>
<b>H_openess</b>	1.00	0.04	0.06	0.24	0.27
<b>H_emotional</b>	0.04	1.00	-0.03	0.22	-0.02
<b>H_contious</b>	0.06	-0.03	1.00	0.24	0.18
<b>H_extraversion</b>	0.24	0.22	0.24	1.00	0.39
<b>H_agreebleness</b>	0.27	-0.02	0.18	0.39	1.00

**Tabella 10 - Correlazione della Scala H (BIG FIVE)**



Non si registrano evidenti segnali di correlazione tra le variabili appartenenti al gruppo dei BIG5, il valore più alto pari a 0.39 sta ad indicare una lieve correlazione positiva tra le due variabili della coppia  $H_{extraversion}$  e  $H_{agreeableness}$ , ciò sta ad indicare che una persona con un alto indice di estroversione tende ad avere anche un alto indice di piacevolezza, con doti di cooperatività, empatia e un atteggiamento cordiale: *plausibile*. Si vedano di seguito i boxplot separati per *Sesso*, facendo riferimento a Figura 28, Figura 29, Figura 30.

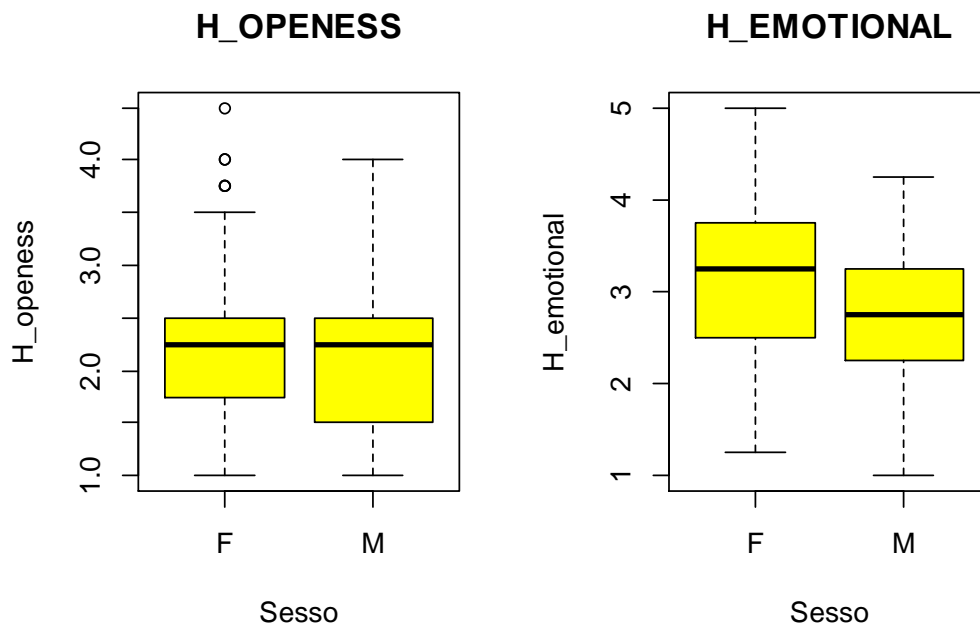


Figura 28 - Boxplot di H\_O e H\_E separati per SESSO

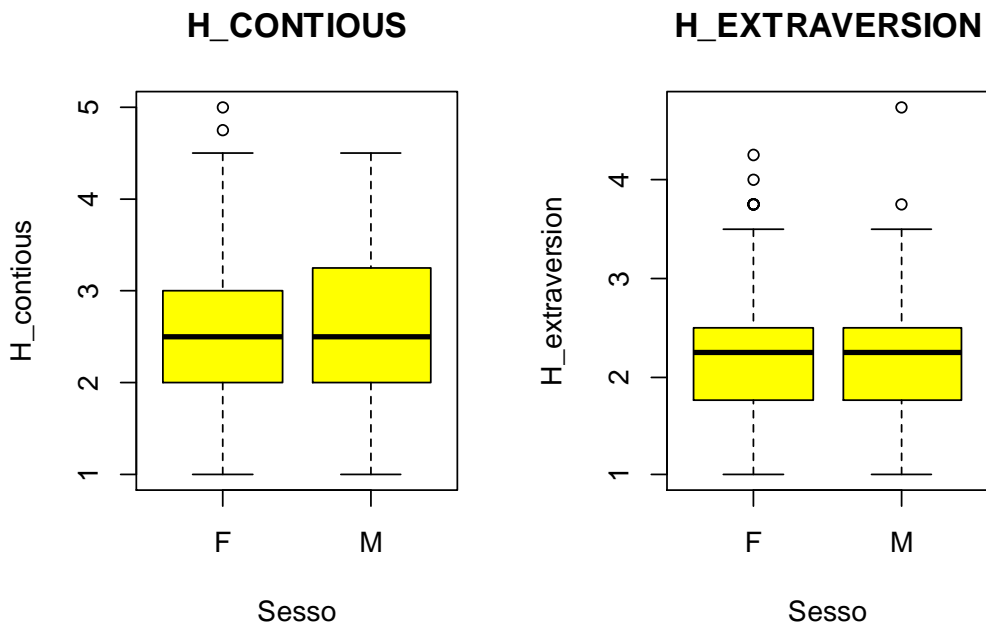


Figura 29 - Boxplot di H\_C e H\_E separati per SESSO

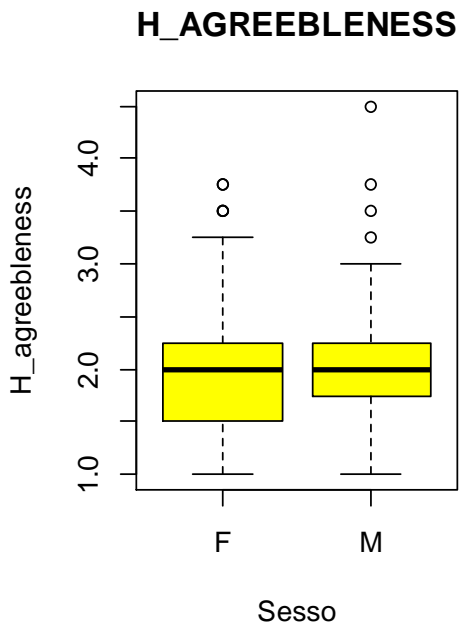


Figura 30 - Boxplot di H\_AGREEBLENESS separato per SESSO

Si vedranno ora i boxplot marginali delle variabili esplicative provenienti dallo studio SEHS, riguardanti la sfera del benessere personale. Si faccia riferimento a Figura 31, Figura 32 e Figura 33.

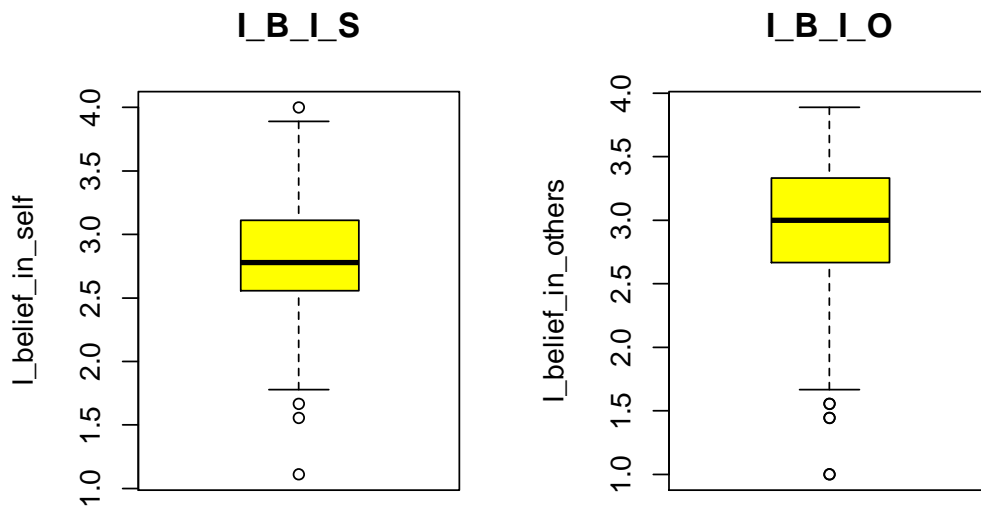


Figura 31 - Boxplot di I\_B\_I\_S e I\_B\_I\_O

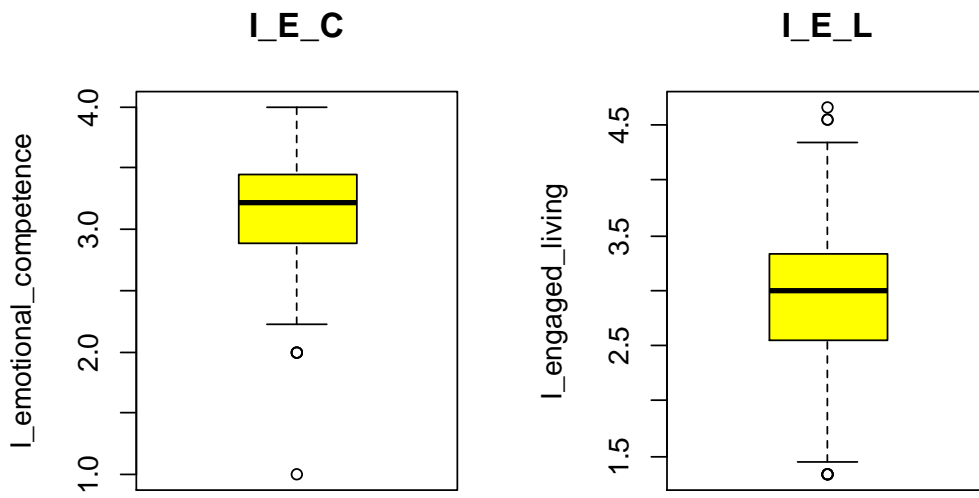


Figura 32 - Boxplot di I\_E\_C e I\_E\_L

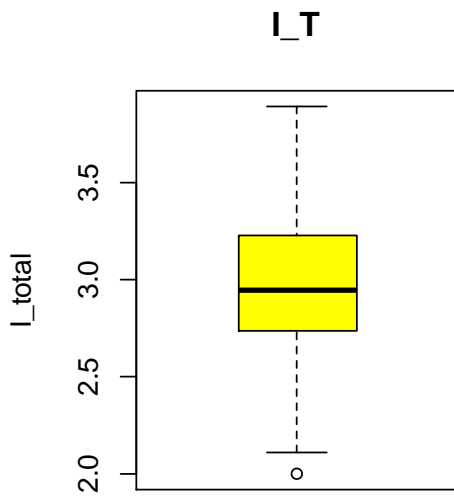


Figura 33 - Boxplot di I\_TOTAL

Le distribuzioni delle cinque variabili risultano simmetriche con limitata presenza di valori outlier, rispettivamente sopra e sotto i due valori adiacenti. Per la specifica dei precisi valori numerici di quartili e media, si veda la Tabella 11 che segue.

	<b>Min</b>	<b>25%</b>	<b>50%</b>	<b>Media</b>	<b>75%</b>	<b>Max</b>
<b>I_B_I_S</b>	1.11	2.56	2.78	2.82	3.11	4.00
<b>I_B_I_O</b>	1.00	2.67	3.00	2.93	3.33	3.89
<b>I_E_C</b>	1.00	2.89	3.22	3.18	3.44	4.00
<b>I_E_L</b>	1.33	2.56	3.00	2.94	3.33	4.67
<b>I_T</b>	2.00	2.74	2.94	2.97	3.22	3.89

**Tabella 11 - Quantili della Scala I (SEHS)**

L'ultimo step dell'analisi descrittiva sulle variabili esplicative del gruppo *Benessere* rimane la verifica della possibile correlazione esistente con le altre dello stesso gruppo. Sapendo che *I\_total* è un indice totale, ci si aspetta che sia presente correlazione con le quattro precedenti.

	<b>I_B_I_S</b>	<b>I_B_I_O</b>	<b>I_E_C</b>	<b>I_E_L</b>	<b>I_T</b>
<b>I_B_I_S</b>	1.00	0.32	0.24	0.48	0.71
<b>I_B_I_O</b>	0.32	1.00	0.13	0.48	0.73
<b>I_E_C</b>	0.24	0.13	1.00	0.16	0.49
<b>I_E_L</b>	0.48	0.48	0.16	1.00	0.81
<b>I_T</b>	0.71	0.73	0.49	0.81	1.00

**Tabella 12 - Correlazione della Scala I (SEHS)**

Facendo riferimento alla Tabella 12, si individua presenza di correlazione tra le variabili esplicative, in particolare le coppie:

- *I\_belief\_in\_self & I\_engaged\_living*
- *I\_belief\_in\_others & I\_engaged\_living*

Facendo il ragionamento dal punto di vista sociologico e psicologico risulta plausibile pensare che ci sia correlazione tra le variabili riguardanti la fiducia che si ha in se stessi, negli altri e l'impegno che si utilizza per portare a termine una operazione. La correlazione è positiva perciò più sale l'indice secondo il quale il soggetto crede, in se stesso o negli altri, più aumenta l'impegno che esso metterà nelle operazioni che andrà a svolgere; che esse siano individuali o di gruppo.

La correlazione positiva, tra la variabile *I\_total* e tutte le precedenti, è confermata e specificata numericamente sempre in Tabella 12, dove basta osservare l'ultima colonna per capire che la correlazione in questione non scende mai sotto il valore di 0.49 il che sta ad indicare che non si può tralasciare. Data la presenza di correlazione tra le variabili della *scala I*, e più precisamente una correlazione marcata che coinvolge la variabile *I\_total* è plausibile aspettarsi che ci siano dei problemi durante la stima dei coefficienti da parte del modello. Si vedano di seguito i boxplot separati per *Sesso*, facendo riferimento a Figura 34, Figura 35 e Figura 36.

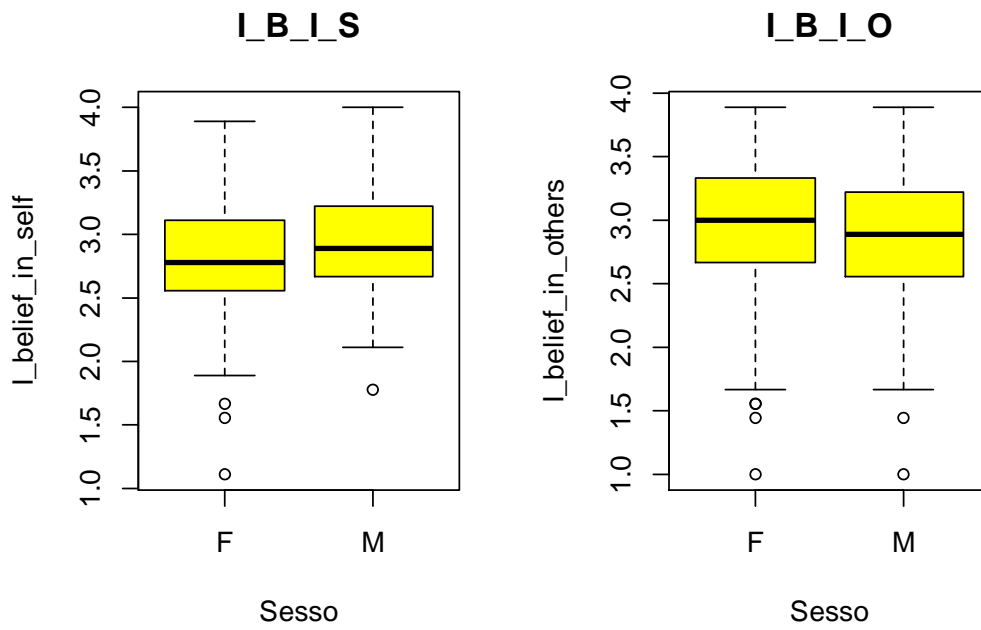


Figura 34 - Boxplot di I\_B\_I\_S e I\_B\_I\_O separati per SESSO

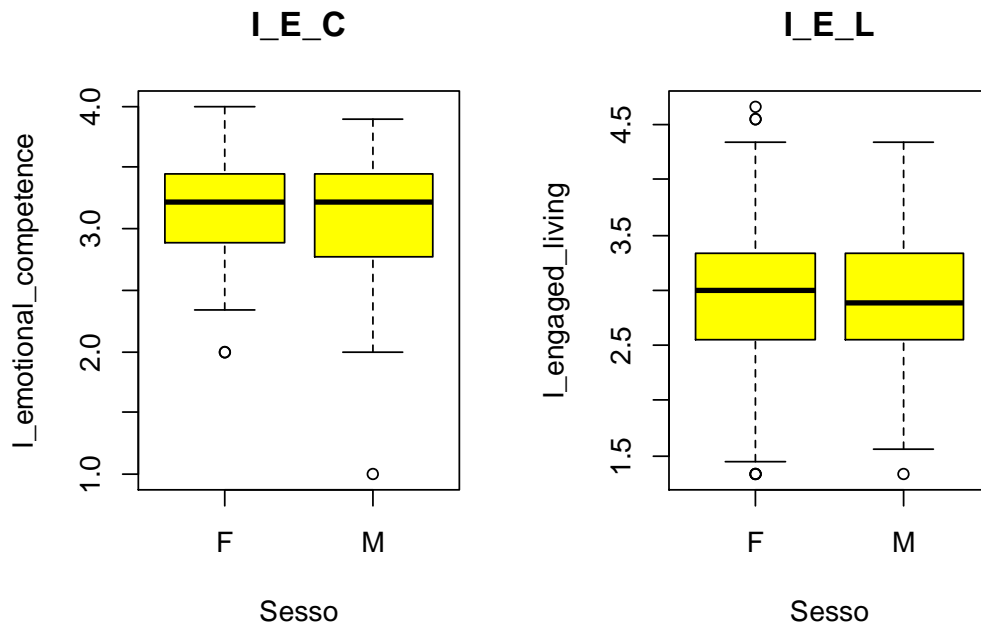


Figura 35 - Boxplot di I\_E\_C e I\_E\_L separati per SESSO

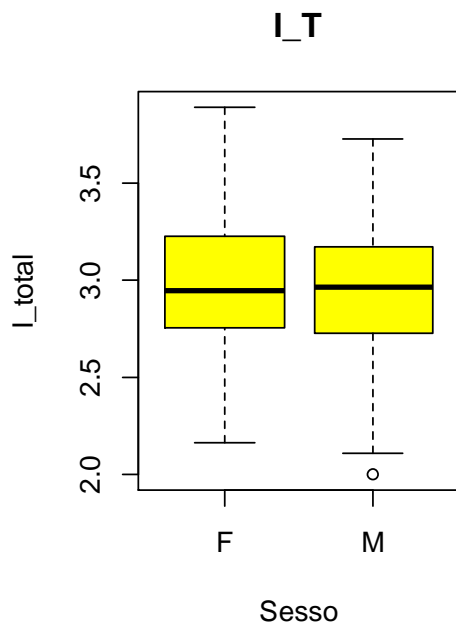


Figura 36 - Boxplot di I\_TOTAL separato per SESSO

Ultima tra le variabili psicologiche è l'indice di *Autostima*, segue le distribuzione rispettivamente marginale, separata per sesso e separata per corso di studio, nei boxplot di Figura 37 e Figura 38.



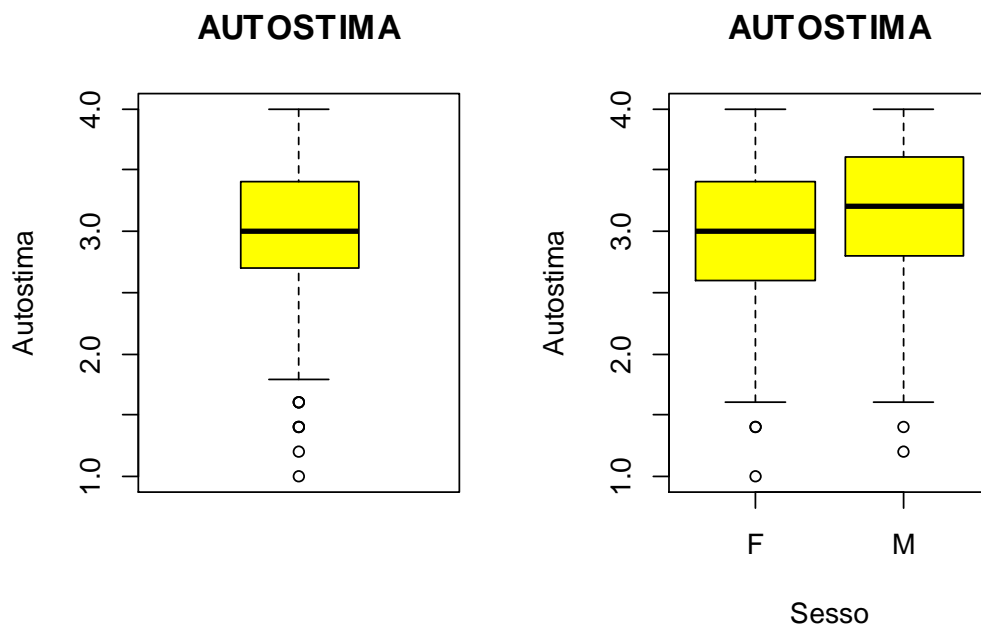


Figura 37 - Boxplot di AUTOSTIMA marginale e separato per SESSO

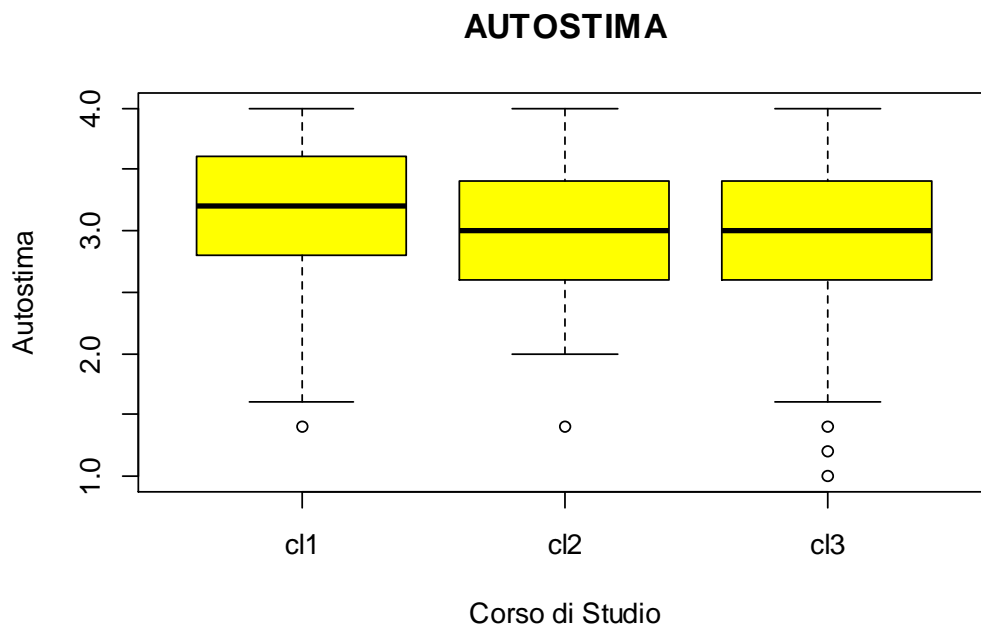


Figura 38 - Boxplot di AUTOSTIMA separati per CORSO DI STUDIO



## 8. IL MODELLO ERGM

### 8.1. Il modello ERGM – parte teorica

I modelli ERGM<sup>8</sup> rappresentano una classe generale di modelli basati sulla teoria della famiglia esponenziale, vengono utilizzati per specificare la distribuzione delle probabilità nell'ambito delle reti casuali. La forma di un modello ERGM può essere scritta come segue.

$$P(Y = y) = \frac{\exp(\theta'g(y))}{k(\theta)}$$

- $Y$  è la rete;
- $g(y)$  è un vettore contenente le statistiche di rete;
- $\theta$  è il vettore dei coefficienti;
- $k(\theta)$  è la costante di normalizzazione.

Quanto appena detto può essere riespresso in termini di log-odds con riferimento al legame tra una coppia di attori appartenenti alla medesima rete sociale.

$$\text{logit}(\text{Pr}(Y_{ij} = 1|Y^c)) = \theta' \Delta(g(y))_{ij}$$

---

<sup>8</sup> Acronimo di Exponential Random Graph Models.

- $Y_{ij}$  è una coppia  $(i,j)$  di attori all'interno della rete  $Y$ ;
- $Y^c$  è il resto della rete  $Y$ ;
- $\Delta(g(y))_{ij}$  è la variazione di  $g(y)$  quando  $Y_{ij}$  assume valore 1.

## 8.2. Il modello più semplice

Si parte nell'analisi considerando il modello più semplice, che contiene solamente il termine *edge*. Prima di riportare l'output del modello, sono riportati gli attributi dell'oggetto network utilizzato per le analisi.

```
Network attributes:
vertices = 359
directed = FALSE
hyper = FALSE
loops = FALSE
multiple = FALSE
bipartite = FALSE
total edges= 2537
  missing edges= 0
  non-missing edges= 2537

Vertex attribute names:
  corso_di_studio sesso vertex.names wave

Edge attribute names not shown
```

E la summary del *Modello 1* di seguito.

```
=====
Summary of model fit
=====

Formula:   DATASET.2_NETWORK ~ edges

Iterations: 6 out of 20

Monte Carlo MLE Results:
  Estimate Std. Error MCMC % p-value
edges -3.19169    0.02026    0 <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null Deviance: 89085 on 64261 degrees of freedom
```

Residual Deviance: 21372 on 64260 degrees of freedom
AIC: 21374    BIC: 21383    (Smaller is better.)

La *probabilità* che due nodi siano legati è quindi pari a:

$$P(Y_{ij} = 1) = \frac{\exp(-3.19169)}{1 + \exp(-3.19169)} = 0.03947$$

È esattamente il valore che ci saremmo potuti aspettare dato che la rete contiene 359 nodi, e la probabilità che due nodi siano collegati è:

$$P = \frac{EDGE\ presenti}{EDGE\ massimi} = \frac{2537}{64261} = 0.03947$$

### 8.3. Il modello con covariata quantitativa

Ora aggiungiamo anche una covariata quantitativa di nodo per aumentare la complessità del modello e interpretare i parametri, si faccia riferimento all'output che segue:

```

=====
Summary of model fit
=====

Formula:   DATASET.2_NETWORK ~ edges + nodecov("I_total")

Iterations: 6 out of 20

Monte Carlo MLE Results:
              Estimate Std. Error MCMC % p-value
edges          -4.86302   0.25023     0 <1e-04 ***
nodecov.I_total 0.28034   0.04162     0 <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      Null Deviance: 89085 on 64261 degrees of freedom
Residual Deviance: 21326 on 64259 degrees of freedom

AIC: 21330    BIC: 21348    (Smaller is better.)

```

La variabile esplicativa  $I_{total}$  rappresenta la variabile totale della scala dei Big Five, è quantitativa. Come si può osservare dal valore del  $p$ -value, c'è un *effetto positivo* sulla probabilità di legame tra due nodi della rete. In particolare l'interpretazione dei parametri è indicata di seguito. Il log-odds condizionale tra due attori è:

$$LogOdds_{TOT} = (-4.86302 * 1) + (0.28034 * I_{total_i}) + (0.2834 * I_{total_j})$$

che si può riassumere in:

$$LogOdds_{TOT} = (-4.86302 * 1) + 0.28034 * (I_{total_i} + I_{total_j})$$

Dopo aver trovato il valore del *LogOdds totale*, per calcolare la probabilità di legame per la coppia si ricorre al solito rapporto.

$$P(Y_{ij} = 1) = \frac{\exp(LogOdds_{TOT})}{1 + \exp(LogOdds_{TOT})}$$

#### **8.4. Il modello con covariata qualitativa (*Homophily*)**

In questa terza casistica viene considerato il modello con covariate qualitative, si faccia riferimento all'output riportato di seguito.

```

=====
Summary of model fit
=====

Formula:   DATASET.2_NETWORK ~ edges + nodematch("sesso", diff = T) + nod
ematch("corso_di_studio",
        diff = T)

Iterations: 7 out of 20

Monte Carlo MLE Results:
              Estimate Std. Error MCMC % p-value
edges                -4.35803    0.05255     0 <1e-04 ***
nodematch.sesso.F      0.06754    0.04510     0  0.134
nodematch.sesso.M      0.05686    0.10465     0  0.587
nodematch.corso_di_studio.c11 2.03377    0.05938     0 <1e-04 ***
nodematch.corso_di_studio.c12 1.80853    0.07520     0 <1e-04 ***
nodematch.corso_di_studio.c13 2.00379    0.05376     0 <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      Null Deviance: 89085 on 64261 degrees of freedom
Residual Deviance: 19274 on 64255 degrees of freedom

AIC: 19286   BIC: 19341   (Smaller is better.)

```

Come nel caso precedente si veda la procedura per il calcolo della probabilità di legame tra una coppia  $(i,j)$  di attori. Nello specifico viene presa in considerazione la variabile significativa delle due presenti nel modello, “*corso\_di\_studio*” con 3 modalità:

- “*c11*”, corso di studio i cui dati sono stati raccolti durante la *Wave 1*;
- “*c12*”, corso di Psicologia di Comunità (Prof. Vieno), *Wave 2*;
- “*c13*”, corso di Turismo (Prof.ssa Furlan), *Wave 2*.

Si può notare che è presente un *effetto positivo* sulla probabilità che esista il legame tra la coppia  $(i,j)$ . Per stimare la probabilità bisogna prima calcolare il *LogOdds totale* e poi calcolare la probabilità effettiva conoscendo il corso di studio al quale appartengono i due attori in questione.

$$\text{LogOdds}_{TOT} = (-4.35803 * 1) + (2.03377 * A_{TOT}) + (1.80853 * B_{TOT}) + (2.00379 * C_{TOT})$$

- $A$  è il numero di attori, nella coppia, che frequentano corso di studio “cl1”;
- $B$  è il numero di attori, nella coppia, che frequentano corso di studio “cl2”;
- $C$  è il numero di attori, nella coppia, che frequentano corso di studio “cl3”;

Da sottolineare che il massimo valore che  $A_{TOT}$ ,  $B_{TOT}$  e  $C_{TOT}$  possono assumere è 2, perché la probabilità si riferisce ad coppia di attori. Dopo aver trovato il valore del  $\text{LogOdds}_{TOT}$ , come visto in precedenza si calcola la probabilità che esista una relazione tra i due attori coinvolti, tramite il rapporto:

$$P(Y_{ij} = 1) = \frac{\exp(\text{LogOdds}_{TOT})}{1 + \exp(\text{LogOdds}_{TOT})}$$

## 8.5. Il modello completo

Dopo aver visto una panoramica sulle due principali tipologie di variabili esplicative oggetto di questa relazione, si procede con l’elaborazione del modello completo di tutte le covariate descritte nel Capitolo 7. Seguiranno *Modello 1*, *Modello 2* e *Modello 3*, verranno tolte man mano le variabili i cui coefficienti non risultano significativi almeno al 5%. Verranno inoltre utilizzati AIC<sup>9</sup> e BIC<sup>10</sup> per la riduzione del modello di partenza.

---

<sup>9</sup> Acronimo di Akaike’s Information Criterion, fornisce una misura della qualità della stima di un modello statistico tenendo conto sia della bontà di adattamento che della complessità del modello.

<sup>10</sup> Acronimo di Bayesian Information Criterion, è un criterio per la selezione di un modello tra una classe di modelli parametrici con un diverso numero di parametri.



## Modello 1

```
=====  
Summary of model fit  
=====  
  
Formula:  DATASET.2_NETWORK ~ edges + nodematch("sesso", diff = T) + nod  
ematch("corso_di_studio",  
      diff = T) + nodecov("H_openess") + nodecov("H_emotional") +  
      nodecov("H_contious") + nodecov("H_extraversion") + nodecov("H_agreeb  
leness") +  
      nodecov("I_belief_in_self") + nodecov("I_belief_in_others") +  
      nodecov("I_emotional_competence") + nodecov("I_engaged_living") +  
      nodecov("autostima") + nodecov("degree.2") + nodecov("betweenness.2")  
  
Iterations:  7 out of 20  
  
Monte Carlo MLE Results:  
  
              Estimate Std. Error MCMC %  p-value  
edges                -8.4382720  0.6030793    0 < 1e-04 ***  
nodematch.sesso.F      0.1848154  0.0509448    0 0.000286 ***  
nodematch.sesso.M      0.0876848  0.1166113    0 0.452089  
nodematch.corso_di_studio.c11  2.6868867  0.0663218    0 < 1e-04 ***  
nodematch.corso_di_studio.c12  2.3527667  0.0822845    0 < 1e-04 ***  
nodematch.corso_di_studio.c13  1.6296706  0.0573218    0 < 1e-04 ***  
nodecov.H_openess      0.0256782  0.0263310    0 0.329462  
nodecov.H_emotional    -0.0128638  0.0226936    0 0.570822  
nodecov.H_contious     -0.0004557  0.0210265    0 0.982709  
nodecov.H_extraversion -0.0868306  0.0321466    0 0.006913 **  
nodecov.H_agreebleness -0.0565047  0.0359486    0 0.115998  
nodecov.I_belief_in_self -0.0619659  0.0547370    0 0.257611  
nodecov.I_belief_in_others  0.0164730  0.0350881    0 0.638732  
nodecov.I_emotional_competence -0.0451145  0.0482444    0 0.349728  
nodecov.I_engaged_living -0.0557752  0.0364190    0 0.125654  
nodecov.autostima      -0.0121317  0.0344128    0 0.724438  
nodecov.degree.2       0.6658427  0.0163693    0 < 1e-04 ***  
nodecov.betweenness.2  -0.0044633  0.0016978    0 0.008568 **  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Null Deviance: 89085 on 64261 degrees of freedom  
Residual Deviance: 15567 on 64243 degrees of freedom  
  
AIC: 15603    BIC: 15766    (Smaller is better.)
```

In questo primo modello non è stata inserita la variabile esplicativa  $I_{total}$ , essendo essa somma dei punteggi relativi alle variabili esplicative della *Scala I*, è combinazione lineare di variabili già presenti nel modello. Già a partire da questo primo modello sono state modificate le due variabili che potevano assumere valori zero, *Degree* e *Betweenness*, in alternativa sono state create due nuove variabili chiamate *Degree.2* e *Betweenness.2* con la logica indicata di seguito.

$$Degree.2 = \sqrt{Degree + 1}$$

$$Betweenness.2 = \sqrt{Betweenness + 1}$$

La prima variabile psicologica ad avere coefficiente significativo è *H\_extraversion*, seppur con effetto negativo, come se l'aumento di estroversione facesse diminuire la probabilità di un'amicizia tra i due nodi della rete. I valori di AIC e BIC sono rispettivamente di 15603 e 15766, si noti che nei modelli più semplici visti in precedenza i due valori erano sensibilmente più elevati, ciò significa che l'introduzione delle variabili esplicative ha contribuito a spiegare la variabilità presente nella rete.

## Modello 2

```

=====
Summary of model fit
=====

Formula:  DATASET.2_NETWORK ~ edges + nodematch("sesso", diff = T) + nod
ematch("corso_di_studio",
      diff = T) + nodecov("H_openess") + nodecov("H_emotional") +
      nodecov("H_contious") + nodecov("H_extraversion") + nodecov("H_agreeb
leness") +
      nodecov("I_belief_in_self") + nodecov("I_belief_in_others") +
      nodecov("I_emotional_competence") + nodecov("I_engaged_living") +
      nodecov("autostima")

Iterations:  7 out of 20

Monte Carlo MLE Results:

```

	Estimate	Std. Error	MCMC %	p-value
edges	-4.60992	0.53024	0	< 1e-04 ***
nodematch.sesso.F	0.02109	0.04643	0	0.649700
nodematch.sesso.M	0.09315	0.10546	0	0.377090
nodematch.corso_di_studio.c11	2.07593	0.06053	0	< 1e-04 ***
nodematch.corso_di_studio.c12	1.83033	0.07607	0	< 1e-04 ***
nodematch.corso_di_studio.c13	1.96657	0.05469	0	< 1e-04 ***
nodecov.H_openess	0.05691	0.02400	0	0.017740 *
nodecov.H_emotional	0.02450	0.02120	0	0.247728
nodecov.H_contious	0.07228	0.01888	0	0.000129 ***
nodecov.H_extraversion	-0.21689	0.02918	0	< 1e-04 ***
nodecov.H_agreebleness	-0.15768	0.03300	0	< 1e-04 ***
nodecov.I_belief_in_self	-0.09004	0.04960	0	0.069474 .
nodecov.I_belief_in_others	0.01692	0.03266	0	0.604364
nodecov.I_emotional_competence	0.20088	0.04277	0	< 1e-04 ***

```

nodecov.I_engaged_living      0.05112    0.03216    0 0.111966
nodecov.autostima            -0.01968    0.03265    0 0.546724
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null Deviance: 89085 on 64261 degrees of freedom
Residual Deviance: 19080 on 64245 degrees of freedom

AIC: 19112    BIC: 19257    (Smaller is better.)

```

Dopo aver tolto gli indici di centralità *Degree.2* e *Betweenness.2* si può notare un miglioramento dal punto di vista della significatività, risultato infatti significative al 5% anche le variabili psicologiche *H\_openess* con effetto positivo, *H\_contious* con efetto positivo, *H\_agreebleness* con effetto negativo per la *scala H* e *I\_emotional\_competence* con effetto positivo per la *scala I*. Per contro è aumentato il valore di AIC e BIC, indicando che la bontà della stima e l'adattamento del modello diminuiscono.

### Modello 3

```

=====
Summary of model fit
=====

Formula:  DATASET.2_NETWORK ~ edges + nodematch("corso_di_studio", diff
= T) +
          nodecov("H_openess") + nodecov("H_contious") + nodecov("H_extraversio
n") +
          nodecov("H_agreebleness") + nodecov("I_emotional_competence")

Iterations:  7 out of 20

Monte Carlo MLE Results:

          Estimate Std. Error MCMC % p-value
edges          -4.62806    0.36389    0 < 1e-04 ***
nodematch.corso_di_studio.c11    2.07434    0.05984    0 < 1e-04 ***
nodematch.corso_di_studio.c12    1.84009    0.07587    0 < 1e-04 ***
nodematch.corso_di_studio.c13    1.96082    0.05437    0 < 1e-04 ***
nodecov.H_openess          0.06531    0.02345    0 0.00536 **
nodecov.H_contious          0.07699    0.01845    0 < 1e-04 ***
nodecov.H_extraversion        -0.19547    0.02540    0 < 1e-04 ***
nodecov.H_agreebleness        -0.17964    0.03182    0 < 1e-04 ***
nodecov.I_emotional_competence    0.18348    0.04173    0 < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null Deviance: 89085 on 64261 degrees of freedom

```

Residual Deviance: 19091 on 64252 degrees of freedom  
AIC: 19109    BIC: 19190    (Smaller is better.)

In questo terzo modello proposto è stato tolto il *Sesso*, *H\_emotional*, *I\_belief\_in\_self*, *I\_belief\_in\_others* e l'indice di *Autostima* in quanto sono risultate non significative nel *Modello 2*; I valori di AIC e BIC sono leggermente diminuiti rispetto al modello precedente. Analizzando quindi i coefficienti significativi al 5% è possibile dedurre che:

- La probabilità che una coppia abbia un legame all'interno della rete analizzata è di per sé bassa, coefficiente di edges con valore -4.62806;
- La prima modalità del *Corso di Studio* sembra essere quella che ha l'effetto maggiore, essendo tutte e 3 con effetto positivo, sulla probabilità che una generica coppia abbia un legame nella rete.

Sulle variabili psicologiche della *scala H* possiamo dire che:

- *H\_openness*, l'apertura mentale, ha un effetto positivo;
- *H\_contious*, indice di consapevolezza, ha un effetto positivo;
- *H\_extraversion*, indice di estroversione, ha un effetto negativo;
- *H\_agreeableness*, indice di gradevolezza personale, ha un effetto negativo.

Sulle variabili psicologiche della *scala I* possiamo dire che:

- *I\_emotional\_competence*, l'indice di quanto una persona riesce a gestire le proprie emozioni, ha un effetto positivo.

## CONCLUSIONI

Un social network come *Facebook*, dandoci la possibilità di scaricare il profilo digitale ci mette a disposizione le informazioni da raffinare prima di essere sottoposte ad analisi. Ciò che si può ricavare è sensazionale, una rete di relazioni tra utenti, bisogna però fissare delle regole per la creazione della *Matrice di Adiacenza* come indicato nel *Capitolo 5*. Gli indici di rete hanno dato la possibilità di confermare la struttura del *Mondo Piccolo* (Milgram & Travers, 1969) adottando le relative implicazioni e l'analisi sulle variabili esplicative ha evidenziato la presenza di dati, di carattere psicologico, pronti per essere utilizzati. L'analisi statistica della rete con modello ERGM ha riscontrato l'influenza superiore da parte delle cinque grandi dimensioni della personalità, i cosiddetti *Big Five*, risultando quasi tutte significative seppur con effetti diversi. Per quanto riguarda gli indici del benessere personale, una soltanto risulta significativa, meno di quanto ci si sarebbe potuti aspettare. Il profilo *Facebook* contiene migliaia di informazioni che potrebbero dare vita a suggestive analisi, già utilizzando la parte di messaggistica diventa possibile valutare tutta la sezione delle comunicazioni tra gli attori della rete. Inoltre, spostando l'attenzione sulle immagini che ogni utente ha caricato se ne potrebbero ricavare preziose indicazioni, a partire dall'immagine del profilo che con buone probabilità si trova in quella posizione per determinati motivi, anche personali, la sfida è scoprire quali. Per concludere *l'era digitale* sta cambiando anche il modo di instaurare la relazione, tra due o più soggetti, e sta fornendo migliaia di informazioni che hanno solo bisogno di essere analizzate per prevedere man mano il cambiamento della civiltà.



## INDICE DELLE FIGURE

Figura 1 - Esempio di RETE SOCIALE.....	2
Figura 2 - Esempio di TRANSITIVITY nella triade i, j, h.....	4
Figura 3 - I sei gradi di separazione.....	11
Figura 4 - GRAFO orientato con 4 nodi .....	13
Figura 5 - Pagina iniziale FACEBOOK .....	15
Figura 6 - Cartella.zip .....	16
Figura 7 - Il contenuto del file Facebook.....	16
Figura 8 - La cartella HTML .....	17
Figura 9 - La cartella PHOTOS .....	17
Figura 10 - La cartella VIDEOS .....	18
Figura 11 - La pagina INDEX .....	19
Figura 12 - Scala di Benessere.....	23
Figura 13 – Il grafico della RETE divisa per WAVE.....	29
Figura 14 - Nodi coinvolti nel DIAMETRO della RETE.....	31
Figura 15 - Distribuzione di DEGREE .....	32
Figura 16 - Valori di GRADO per WAVE .....	33
Figura 17 - Distribuzione di BETWEENNESS .....	34
Figura 18 - Valori di BETWEENNESS per WAVE .....	35
Figura 19 - Grafico a torta di SESSO con percentuali.....	38
Figura 20 – RETE colorata per CORSO DI STUDIO .....	39
Figura 21 - Boxplot di DEGREE e BETWEENNESS .....	40
Figura 22 - Boxplot di GRADO e BETWEENNESS separati per SESSO .....	41
Figura 23 - Boxplot di DEGREE separati per CORSO DI STUDIO .....	42
Figura 24 - Boxplot di BETWEENNESS separati per CORSO DI STUDIO .....	43
Figura 25 - Boxplot di H_OPENESS e H_EMOTIONAL .....	44
Figura 26 - Boxplot di H_CONTIOUS e H_EXTRAVERSION .....	45
Figura 27 - Boxplot di H_AGREEBLENESS .....	45

Figura 28 - Boxplot di H_O e H_E separati per SESSO .....	47
Figura 29 - Boxplot di H_C e H_E separati per SESSO .....	48
Figura 30 - Boxplot di H_AGREEBLENESS separato per SESSO.....	48
Figura 31 - Boxplot di I_B_I_S e I_B_I_O .....	49
Figura 32 - Boxplot di I_E_C e I_E_L .....	50
Figura 33 - Boxplot di I_TOTAL .....	50
Figura 34 - Boxplot di I_B_I_S e I_B_I_O separati per SESSO.....	53
Figura 35 - Boxplot di I_E_C e I_E_L separati per SESSO.....	53
Figura 36 - Boxplot di I_TOTAL separato per SESSO.....	54
Figura 37 - Boxplot di AUTOSTIMA marginale e separato per SESSO.....	55
Figura 38 - Boxplot di AUTOSTIMA separati per CORSO DI STUDIO .....	55

## INDICE DELLE TABELLE

Tabella 1 - Divisione in gruppi per il calcolo di ASSORTATIVITY .....	7
Tabella 2 - Approccio CLASSICO .....	12
Tabella 3 - Approccio SNA (Matrice di Adiacenza) .....	12
Tabella 4 - Analisi descrittive della RETE .....	30
Tabella 5 - Variabile esplicativa SESSO .....	38
Tabella 6 – Frequenza di CORSO DI STUDIO .....	40
Tabella 7 - Quantili di DEGREE e BETWEENNESS .....	41
Tabella 8 - Correlazione tra DEGREE e BETWEENNESS.....	44
Tabella 9 - Quantili della Scala H (BIG FIVE) .....	46
Tabella 10 - Correlazione della Scala H (BIG FIVE).....	46
Tabella 11 - Quantili della Scala I (SEHS).....	51
Tabella 12 - Correlazione della Scala I (SEHS) .....	51



## BIBLIOGRAFIA

Dallago, F. & Roccato, M., 2010. Right-wing authoritarianism, Big Five and perceived threat to safety. *European Journal of Personality*.

Finos, L., Marino, C., Meneghetti, D. & Righetto, S., 2017. *Package myFBr - Read and visualizes the infos of your FaceBook profile*, Padova: s.n.

Furlong, M. J. et al., 2013. Preliminary Development and Validation of the Social and Emotional Health Survey for Secondary School Students. *An International and Interdisciplinary Journal for Quality-of-Life Measurement*.

Handcock, M. et al., 2017. *Package ERGM - Fit, Simulate and Diagnose Exponential-Family Models for Networks*. [Online] Available at: <http://statnet.org>

Milgram, S. & Travers, J., 1969. An Experimental Study of the Small World Problem. *Sociometry*.

Newman, M., 2002. Mixing Patterns in Networks. *Physical Review Letters*.