

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

**CORSO DI LAUREA IN STATISTICA E TECNOLOGIE
INFORMATICHE**



Relazione finale

INDICI DI SOCIAL NETWORK ANALYSIS PER MISURARE IL CAPITALE SOCIALE:
COSTRUZIONE DI UNA LIBRERIA CON IL SOFTWARE R E VERIFICA DI IPOTESI SU UN
CASO DI STUDIO.

Relatore: dott. Livio Finos

Correlatore: dott. Andrea Sciandra

Laureando: Federico Gioachin

Anno Accademico 2009-2010

Indice Generale

<i>Introduzione</i>	3
<i>1.1 Social Network Analysis</i>	6
<i>1.1.1 Un esempio: Come si cerca un lavoro</i>	9
<i>1.2 Rete ego-centrata</i>	12
<i>1.3 Il Capitale Sociale e i buchi strutturali</i>	13
<i>2.1 Libreria egonet</i>	19
<i>2.2 Indici</i>	22
<i>3 Un caso di studio</i>	35
<i>Riferimenti bibliografici</i>	46
<i>Ringraziamenti</i>	48

Introduzione

Questo lavoro nasce dalla esigenza di avere un pacchetto specifico di funzioni tali da affrontare un' analisi strutturale di reti sociali ego-centrate ("egonet") sfruttando l'ambiente R (noto frame work statistico). Questo pacchetto, o libreria, sarà costituito, in una prima fase, da una funzione adatta per l'importazione dei dati e successivamente, da una serie di indici utili per condurre la nostra analisi.

In R sono già stati implementati metodi che permettono lo studio di una rete sociale. Quest'ultimi sono tutti racchiusi nella libreria SNA che, nella documentazione allegata, viene descritta come:

"Un insieme di strumenti per la Social Network Analysis, che include indici a livello sia di nodo che di grafo, metodi di covarianza e di distanza strutturale, di equivalenza strutturale, di modellazione p^* , di generazione di grafi casuali, e di visualizzazione sia 2D che 3D"

L'obbiettivo che si pone questo progetto è quello di estrarre gli indici più adatti al nostro caso dalla suddetta libreria, modificandone alcuni ove lo si esige ed aggiungendone altri mancanti creati su misura.

Il lavoro è strutturato come segue: nel primo capitolo vengono esposti i concetti teorici del progetto. Il primo paragrafo spiega in maniera sintetica cosa si intende per social network analysis: su cosa si basa questa metodologia, il suo scopo ed il perché sia stata adottata da molti studiosi per analizzare le reti sociali. Nel secondo paragrafo, invece, si parlerà della rete ego-centrata e più nello specifico delle particolarità che la contraddistinguono tra le tante reti sociali. Nel terzo paragrafo si espone il concetto di capitale sociale, le controversie teoriche che ha suscitato tra i tanti studiosi della materia ed il suo legame con la social network analysis.

Il secondo capitolo tratta la specificazione della libreria "egonet": il perché della sua introduzione all'interno dell' ambiente R e la spiegazione delle sue varie componenti: le due funzioni create, ossia la prima per l'importazione di matrici e la seconda per il calcolo degli indici. Sempre in questo capitolo si trova una spiegazione di queste misure e la loro relazione con il capitale sociale.

Infine, nel terzo capitolo, si illustra un'applicazione pratica delle funzioni create attraverso un'analisi condotta sui gruppi di studio esistenti all'interno della facoltà di statistica dell'università degli studi di Padova.

1.1 Social Network Analysis

La Social Network Analysis (SNA), ovvero analisi dei reticoli sociali, è una recente metodologia di analisi delle relazioni sociali sviluppatasi a partire dai contributi di Jacob Levi Moreno, il fondatore della sociometria, scienza che analizza le relazioni interpersonali.

Nella teoria delle reti sociali, la società è vista e studiata come rete di relazioni, più o meno estese e strutturate. Il presupposto fondante è che ogni individuo (o attore) si relaziona con altri soggetti e questo tipo di iterazione condiziona o modifica il comportamento di entrambi.

Lo scopo principale dell'analisi di network è appunto quello di individuare e analizzare tali legami tra gli individui così da scoprire schemi riconducibili a strutture relazionali, determinandone le condizioni di origine ed infine di rilevarne le conseguenze per l'azione.

La SNA consiste in un insieme di tecniche di analisi strutturale, che però si basano su quattro postulati relativi alla realtà sociale:

- a) Il comportamento dell' attore è interpretabile principalmente in termini di vincoli strutturali all' azione, piuttosto che in termini di libertà di scelta tra corsi di azione alternativi;
- b) La spiegazione dei fenomeni sociali deve essere ricercata nelle relazioni tra gli elementi, piuttosto che nelle caratteristiche degli elementi;
- c) Le tecniche di analisi si concentrano sulla natura relazionale della struttura sociale e sostituiscono (o integrano) le tecniche statistiche classiche che si basano su elementi considerati indipendenti tra loro;
- d) La forma delle relazioni sociali può a sua volta essere spiegata in parte come l'esito delle scelte degli attori, individuali o collettivi, che rappresentano i nodi del reticolo. (Chiesi,1999)

Benché nelle scienze sociali nessuna procedura di ricerca, e tantomeno uno strumento analitico, possano essere considerate teoricamente neutrali, l'analisi dei reticoli si presta ad applicazioni teoricamente orientate in direzioni molto diverse, allo stesso modo in cui le tecniche di analisi della matrice dei dati tradizionale possono essere utilizzate in ambiti teorici molto diversi. Il

fatto che l'analisi dei reticoli sia stata utilizzata nelle ricerche empiriche di svariate discipline conferma questa sua versatilità.

La social network analysis sta prendendo molto piede nel campo delle scienze sociali ed il motivo è presto detto: l'idea che i fenomeni sociali possano essere rappresentati da reti di relazioni, che la società possa essere assimilata ad una sorta di rete e che le azioni degli attori sociali possano essere raffigurate come esito di vincoli prende molto consenso in questo ambito perché, come molti studiosi affermano, questi concetti si avvicinano molto alla natura della società moderna e quindi l'analisi dei reticoli è molto più adeguata della classica analisi su campioni di casi. Infatti, la network analysis si occupa direttamente degli elementi costitutivi della società – le relazioni tra gli attori – al contrario delle tecniche tradizionali, che si sforzano di interpretare la realtà solo indirettamente attraverso gli attributi di individui singolarmente presi. Welmann (1988) ha sottolineato che le analisi delle reti personali si sono ben integrate con i metodi di ricerca tradizionali individuabili nelle surveys.

1.1.1 Un esempio: Come si cerca un lavoro

Per assimilare meglio queste informazioni vi porto alla lettura un piccolo esempio che spiega le differenze basilari tra un' analisi tramite impostazioni tradizionali ed una approcciando il metodo della rete da parte di studiosi sul funzionamento del mercato di lavoro:

Utilizzando il primo criterio si separa, innanzi tutto, l'approccio macro da quello micro, mettendo a disposizione tecniche differenti. Per il primo livello, di solito, si va ad analizzare l'andamento di indicatori aggregati su base territoriale per avere una stima indiretta di comportamenti concreti sottoposti a vincoli di mercato. A livello micro, invece, un risultato soddisfacente si ottiene modellizzando il comportamento razionale e massimizzante dell'attore che cerca lavoro sulla base del capitale investito per la proprio formazione ed eventualmente sulla base delle strategie di retribuzione delle attività tra i membri del nucleo familiare. Nel caso questo studio fosse condotto da sociologi si inseriscono anche variabili culturali sempre legate all' identità sociale dell' attore. In definitiva vengono privilegiate le informazioni relative alle caratteristiche individuali senza cogliere le modalità che permettono agli attori di individuare e ricoprire posti vacanti.

L'approccio di rete, invece, pone al centro dello studio il modo in cui le persone ottengono le informazioni sulle opportunità di lavoro. In questo modo vengono analizzati contemporaneamente sia gli aspetti micro che quelli macro del mercato. L'analisi dei reticoli permette anche di individuare la posizione strategica di chi meglio è in grado di fornire informazioni utili alla ricerca di un lavoro, il tipo di legame che intercorre tra l'attore e l'informatore, la lunghezza della catena di conoscenze entro cui le segnalazioni passano.

Dagli studi di Mark Granovetter si mette in luce che per i lavoratori professionalizzati le informazioni rilevanti provengono più frequentemente da contatti di semplice conoscenza, mentre l'aiuto di amici e parenti stretti è molto raro. È evidente il fatto che gli ultimi che ho descritto siano molto motivati ad aiutare chi cerca lavoro, ma non si trovano nella posizione strutturale migliore per poterlo fare, molto probabilmente perché condividono lo stesso ambiente con il soggetto preso in esame e difficilmente possono disporre di informazioni ulteriori rispetto a quelle già in possesso dall'attore. In conclusione, il fatto di condividere legami forti non aiuta la ricerca di un lavoro. Effetto contrario si ha quando il lavoro è dequalificato: soprattutto nei settori caratterizzati da stagionalità nell'occupazione (alimentare) non solo il

reclutamento, ma anche l'organizzazione produttiva passano ancora attraverso le reti familiari.

1.2 Rete ego-centrata

Una rete sociale consiste in un gruppo di persone connesse tra di loro da diversi legami sociali, quali possono essere la semplice conoscenza, vincoli familiari, rapporti di lavoro. Esempi di social network possono essere comunità di sportivi che si ritrovano per tifare la propria squadra, gruppi di fedeli uniti sotto la stessa fede, confraternite segrete tipo la massoneria etc....

La rete ego-centrata costituisce un particolare sottogruppo delle tradizionali reti sociali. Queste sono caratterizzate da un attore focale, chiamato ego, da un insieme di soggetti chiamati alter e dai legami che li collegano. Nel formare questo tipo di rete è l'attore focale a fornire i nominativi degli alter con cui ha contatti specificando caratteristiche e legami che, a suo parere, intercorrono tra i soggetti nominati. In questo particolare caso, di fatto, è ego il centro della rete.

1.3 Il Capitale sociale e buchi strutturali

Molti approcci teorici hanno cercato di definire il capitale sociale nel corso del tempo, tuttora esiste una sorta di confusione su questo termine tra i tanti studiosi, rendendo improbabile la crescita cumulativa di conoscenze sull'argomento. Dello stesso parere sono anche Nal Lin e Bonnie Erickson (Lin e Erickson 2008,2) che sostengono : "approcci differenti hanno condotto a misure di capitale sociale numerose e liberamente costruite, rendendo impossibile valutare la validità del concetto e della teoria o l'accumulazione di riscontri empirici". La prima apparizione del concetto di capitale sociale si trova già a partire dal 1916 che viene definito da Lydia J. Hanifan (al tempo ispettore scolastico del West Virginia) come "quegli elementi tangibili che contano più di ogni altra cosa nella vita quotidiana delle persone: la buona volontà, l'amicizia, la partecipazione e i rapporti sociali tra coloro che costituiscono un gruppo sociale. Se una persona entra in contatto con i suoi vicini e questi a propria volta con altri vicini, si determina un'accumulazione del capitale sociale" (Hanifan, 1916). Questo concetto scomparve, tranne qualche accenno che non ha avuto particolare riscontro fino al contributo di Pierre Bourdieu che definisce il capitale sociale a partire da una più ampia

teoria sulle forme di capitale. Per capitale sociale si intende: “ l’insieme delle risorse attuali e potenziali che sono legate al possesso di una rete durevole di relazioni, più o meno istituzionalizzate di interconoscenza e di inter-riconoscimento; o in altri termini, all’appartenenza a un gruppo (...) Il volume di capitale sociale posseduto da un particolare agente dipende dunque dall’ampiezza della rete di legami che egli può efficacemente mobilitare e dal volume di capitale (economico, culturale e simbolico) detenuto da ciascuno di coloro cui egli è legato” (Bourdieu 1980,2).

Da questa definizione si intuisce come l’Autore voglia esprimere il concetto di capitale sociale come risorsa individuale nata dalla rete di relazioni con altri individui tralasciando il fatto che possa essere una proprietà della relazione interindividuale. Di altra opinione è invece James Coleman (1988) : “il capitale sociale è definito dalla sua funzione. Non si tratta di un’entità singola, ma di una verità di differenti entità che hanno due caratteristiche in comune: consistono tutte di un determinato aspetto della struttura sociale e rendono possibili determinate azioni degli individui che si trovano dentro questa struttura. (...) Una data forma di capitale sociale che è preziosa nel facilitare certe azioni può essere non utile o perfino dannosa per altre. Diversamente

da altre forme di capitale, il capitale sociale risiede nella struttura delle relazioni tra gli attori: esso non si trova negli individui, né negli input fisici alla produzione.” (Coleman 1990, trad. it. 2005,388). L’ Autore, quindi, specifica che il capitale sociale risiede nella struttura delle relazioni anche se, dal punto di vista metodologico fa riferimento esplicito ad azioni individuali indipendenti.

Nan Lin, invece, si interessa soprattutto alla posizione del soggetto all’interno di una rete e definisce il capitale sociale come le “risorse *embedded* (traducibile come “incorporata”) in una struttura sociale che sono reperite e/o mobilitate dall’individuo in azioni dotate di fini”. Elementi principali dell’analisi e della ricerca sono i concetti di risorsa (*embedded resource*) e di relazione (intesa come *network location*). Per *embedded resource* si intende la ricchezza, il potere e lo status propri degli individui con i quali il soggetto di cui ci interessa studiare il capitale sociale è in interazione; per *network location* si intende invece la posizione occupata dal soggetto-ego (posizione di ponte) all’interno della sua rete di relazioni (con gli alter) in modo da poter mobilitare le migliori risorse e ottenere maggiori vantaggi.

L’ultimo parere che richiamo all’attenzione del lettore è quello fornito da Ronald Burt che approfondisce ed in qualche maniera innova la teoria di Lin.

In un suo saggio l'Autore definisce il capitale sociale come una metafora del vantaggio: "La società può essere vista come un mercato nel quale i soggetti scambiano una varietà di beni e di idee, nel proseguimento dei loro fini. Alcuni individui, o alcuni gruppi di individui, raggiungono meglio le proprie finalità, nel senso che ottengono vantaggi più grandi dei loro sforzi. Alcuni godono di benefici più elevati e diventano importanti più velocemente di altri (...). In termini di capitale umano, la spiegazione di tale ineguaglianza risiede nel fatto che gli individui che riescono meglio sono più abili, sono cioè più intelligenti, più qualificati (...). Il capitale sociale è il completamento contestuale del capitale umano. La metafora del capitale sociale sostiene che i soggetti che agiscono meglio sono anche in qualche modo "connessi meglio" (Burt, 2002)

Ci sono due modi di affrontare, in maniera diversa, la "migliore connessione": da una parte c'è il dibattito sulla *network closure* (chiusura della rete) per il quale i reticoli altamente densi dotano coloro che li formano di un alto livello di capitale sociale, poiché garantiscono un immediato e rapido accesso alle informazioni e facilitano l'efficacia delle sanzioni favorendo la propensione a concedere fiducia interpersonale; dall'altra, il dibattito degli *structural holes*

(buchi strutturali che designano l'assenza di legami tra contatti non ridondanti) porta a considerare il capitale sociale in funzione delle opportunità che si hanno, all'interno di una struttura sociale, di esercitare la funzione di *broker*. Ognuno di questi buchi nella struttura sociale, infatti rappresenta l'opportunità per il soggetto, la cui rete abbraccia questo buco strutturale, di mediare il flusso di informazioni fra soggetti e di controllare i progetti che connettono gli individui ai lati opposti del buco strutturale. Questo concetto si estende anche alla creatività, la scintilla creativa consiste nel vedere ponti laddove gli altri vedono buchi strutturali. La mediazione fra buchi strutturali rappresenta capitale sociale. Per gli individui e i gruppi, le reti che abbracciano i buchi strutturali sono associate a creatività e successo, promozioni anticipate e ricompense più alte.

Quest'ultimi due Autori chiariscono che il capitale sociale è una risorsa che non appartiene né agli individui né alla collettività, bensì consiste nelle relazioni che mediano il rapporto tra individuo e società. Ed è proprio da questi concetti che si partirà per affrontare l'analisi perché, in qualche modo, sono i fondamenti per l'analisi strutturale.

2.1 Libreria “egonet”

Come già esposto nel capitolo introduttivo, l'idea di creare un nuovo pacchetto di funzioni per l'analisi strutturale di una rete sociale nasce dalla difficoltà d'utilizzo degli strumenti della libreria SNA quando ci trova ad esaminare una rete ego-centrata.

Innanzitutto si deve spiegare che i reticoli ego-centrati, che andremo ad analizzare, si presentano come matrici di adiacenza di individui, quindi, per lavorare con i dati a disposizione nell'ambiente R si è proceduto ad implementare una funzione (chiamata `read.egonet()`) che importa questi reticoli da file salvati in formato CSV, li pulisce e li porta in forma matriciale.

Nella prossima pagina è implementato lo script della funzione:

Funzione `read.egonet`

```
function(file, sep=";", dec=".", ego.name="EGO") {
```

```

dat <- read.csv(file, header=T, sep= sep)

dat <- dat[!is.na(dat[,ego.name]),, drop = FALSE]

rownms <- dat[,1]

colnms <- rownms

dat <- as.matrix(dat[,-1, drop = FALSE])

rownames(dat) <- rownms

colnames(dat) <- colnms

idego <- which(rownames(dat)==ego.name)

dat <- dat[c(idego,setdiff(1:dim(dat)[1],idego)),, drop = FALSE]

idego <- which(colnames(dat)==ego.name)

dat <- dat[,c(idego,setdiff(1:dim(dat)[1],idego)), drop = FALSE]

keep <- intersect(rownames(dat), colnames(dat))

dat <- dat[keep,keep, drop = FALSE]

dat

}

```

Il risultato della fase di importazione (matrici di adiacenza) sarà dunque trattato in modo da ottenere diversi tipi di indici che saranno poi utilizzati per condurre un'analisi di tipo statistico sui reticoli a disposizione.

Ecco due esempi di matrici di adiacenza: gli elementi della matrice rappresentano i legami tra gli attori (1 dove esiste un legame e 0 altrimenti) .

Nella prima (A) tutti gli attori hanno legami con gli altri, nella seconda (B), invece, alcuni non sono in relazione con altri attori. Si useranno queste due matrici per una dimostrazione pratica del funzionamento degli indici.

0	1	1	1		0	1	0	0
1	0	1	1		1	0	1	1
1	1	0	1		1	1	0	0
1	1	1	0		0	1	1	0

(A)

(B)

Nelle matrici A e B, come si può notare, gli elementi della diagonale sono degli zeri. Il motivo è che essendo una matrice di adiacenza gli attori non possono avere relazioni con loro stessi.

Gli indici derivano in parte dalla libreria SNA (density, degree, efficiency, hierarchy, centralization) altri sono stati implementati per consentire un'analisi più ampia e specifica dei reticoli ego-centrati (effective size e constraint) . Nel prossimo paragrafo vi è la spiegazione dettagliata di tutti gli indici presenti nella libreria creata ed in alcuni il loro legame con il capitale sociale.

2.2 Indici

- **Densità (density)**

La densità di una rete è una proprietà strutturale a livello di gruppo che considera il rapporto tra i legami presenti nella rete su quelli possibili. Per un grafo non orientato, formalmente avremo:

$$\Delta = \frac{L}{g(g-1)/2} = \frac{2L}{g(g-1)}$$

La densità di una rete varia tra 0 e 1. Vale 0 se nessun legame della rete è presente, ovvero se tutti i nodi sono isolati; vale 1 se ogni attore è adiacente a tutti gli altri. In quest' ultimo caso il grafo si dice completo. La densità può avere relazioni positive o negative con il capitale sociale, poiché i reticoli particolarmente densi riescono a facilitare i loro membri nel raggiungimento di alcune finalità (ad esempio, nell' ottenere rapidamente un'informazione), mentre possono essere di ostacolo per il raggiungimento di altre (ad esempio, nell'evitare di essere contagiati da una malattia). Questo indice, nel pacchetto creato, è stato richiamato dalla libreria SNA tramite la funzione `gden()` che si occupa di calcolare la densità di tutta la rete. Per quanto

riguarda l'approccio ego-centrico, è necessaria una precisazione importante sul modo di calcolare la densità. In una rete ego-centrata è consuetudine misurare la densità trascurando l'attore su cui è focalizzata l'attenzione e i suoi contatti diretti per concentrarsi solo sui legami che esistono tra questi contatti. Quindi, si è costruito un altro indicatore, chiamato ego.gden() che calcola quello che è stato appena descritto.

Applicando l'indice alle matrici A di esempio abbiamo che sia la densità totale che quella ego-centrata sono uguali a 1 essendo la matrice completa.

Nella matrice B abbiamo invece:

density	ego.gden
0.66667	0.83333

- **Gradi (degree)**

L'interpretazione più semplice del concetto di centralità è data dal computo dei gradi: questa misura si concentra sugli archi che collegano un dato vertice al suo vicinato, e semplicemente ne conta il numero. L'attore con il grado più alto rappresenta metaforicamente il luogo nel gruppo dove "le cose

accadono". In contrasto, gli attori con un basso grado rappresentano le posizioni periferiche nella rete: estremizzando, se avessimo un attore isolato (grado 0), e decidessimo di eliminarlo, nulla cambierebbe nella disposizione dei legami tra gli altri attori. Il grado di un vertice varia da 0 a $g - 1$ (il vertice è collegato con tutti gli altri). Per calcolare il grado di un vertice è quindi sufficiente contare il numero degli archi incidenti ad esso, oppure sommare i valori di riga (o di colonna) a cui il vertice appartiene nella matrice di adiacenza.

Se il grafo è orientato, il concetto di grado si amplia, poiché possiamo prendere in considerazione la direzione della relazione. Tutti gli archi che originano dal vertice n_i (ovvero le scelte compiute dall'attore che il vertice rappresenta) costituiscono il suo outdegree, o grado in uscita; tutti gli archi che arrivano al vertice n_i (ovvero le preferenze ricevute dall'attore che il vertice rappresenta) costituiscono invece il suo indegree, o grado in entrata.

L'introduzione di questi due concetti ci permette di distinguere quattro tipologie all'interno delle quali i teorici dei grafi classificano i vertici:

isolato:

un vertice è isolato, nessun arco lo coinvolge ($d_i(n_i) = d_o(n_i) = 0$).

Trasmittitore:

dal vertice originano degli archi ($d_i(n_i) = 0$ AND $d_o(n_i) > 0$).

Ricevitore:

al vertice arrivano degli archi ($d_i(n_i) > 0$ AND $d_o(n_i) = 0$).

Portatore:

o ordinario: al vertice arrivano archi ed il vertice origina archi

($d_i(n_i) > 0$ AND $d_o(n_i) > 0$).

Esiste una condizione che distingue un portatore da un ordinario: sebbene entrambi abbiano indegree ed outdegree positivi, nel nodo portatore questi due valori sono entrambi uguali ad 1.

La relazione con il capitale sociale è positiva: più sono le persone con cui si hanno relazioni, più alta è la probabilità che una di esse posseda la risorsa di cui si ha bisogno. Nella libreria egonet questo indice è richiamato dalla libreria SNA (`degree()`). Nelle reti ego-centrate, essendo il nostro punto d'interesse "EGO" (ossia l'attore focale) è molto più utile misurare il grado in uscita e quindi, come parametro della funzione specificheremo la modalità outdegree.

Nell'esempio abbiamo:

outdegree A	outdegree B
3	1

Nella libreria si è implementata anche la modalità indegree anche se, come vedremo nell'applicazione, i due indici danno sempre gli stessi risultati (come è lecito immaginare).

Le prossime due misure sono state proposte da Burt (1992) e sono relative ai buchi strutturali.

- **Effective Size**

Esprime il numero di alter cui ego è connesso meno il grado medio degli alter presenti all'interno della rete di ego, non conteggiando i legami con ego.

Esprime, in sostanza, il numero di alter cui ego è connesso, meno un fattore di ridondanza. L' effective size rende conto della capacità di ego di gettare ponti su buchi strutturali ed è in relazione positiva con il capitale sociale detenuto da ego. Maggiori sono le differenti zone del network con cui ego è in relazione, maggiori saranno i vantaggi in termini informativi e di controllo di ego. L' effective size di un soggetto i , all' interno di un network si calcola nel modo seguente:

$$\sum_j [1 - \sum_q p_{iq} m_{jq}] \quad q \neq i, j \quad (1)$$

dove :

$$p_{iq} = \frac{(z_{iq} + z_{qi})}{\sum_j (z_{ij} + z_{ji})} \quad i \neq j \quad (2)$$

e :

$$m_{jq} = \frac{(z_{jq} + z_{qj})}{\max_k (z_{jk} + z_{kj})} \quad j \neq k \quad (3)$$

Dove il soggetto i rappresenta ego ed i soggetti j rappresentano gli alter.

Se l' effective size è uguale a 1 ogni contatto di i è fortemente legato agli altri contatti, se l' effective size coincide con la dimensione della rete i non c'è alcuna relazione tra gli individui in contatto con il soggetto i . Il valore di p_{iq} è la "proporzione" di legame q sul totale delle relazioni, quello di m_{jq} , invece, esprime la forza del legame tra j e q . Questo indicatore è stato implementato perché non presente nella libreria SNA. La funzione che lo calcola ha preso il nome di `effsize()`. Di seguito l'implementazione della funzione `effsize` e l'applicazione sulle due matrici di esempio.

```
function(dati, ego.name) {
  n <- dim(dati)[1]
  if(n < 2) return(NaN)
  Sj <- 0
  for(y in 2:n)      Sj <- Sj + (dati[ego.name,y] + dati[y,ego.name])
  ris <- 0
  for (j in 2: n){
    cont1 <- setdiff(1:n,j)
    Vetmax <- rep(NA,(length(cont1)))
    for (k in setdiff(1:n,j)){ Vetmax[k] <- (dati[j,k] +dati[k,j])}
    massimo <- max(Vetmax,na.rm=T)
    sumPM <- 0
    for(f in setdiff(2:n,j)){ #f=q in formula originale
      Piq <- (dati[ego.name, f] + dati[f,ego.name])/ Sj
      Mjq <- (dati[j,f] + dati[f,j])/ massimo
      sumPM <- sumPM +Piq * Mjq
    }
  }
}
```

```

}
ris <- ris + (1 - (sumPM))
}
ris
}

```

effsize A	effsize B
1.00000	1.16667

- **Constraint**

Il constraint totale rende conto del grado in cui ego si trova chiuso e in dipendenza dagli alter all' interno del suo reticolo ed ha pertanto una relazione tendenzialmente negativa con il capitale sociale proprio di ego.

Il constraint totale (C_i) che pesa su un soggetto i è uguale a:

$$C_i = \sum_j c_{ij}$$

Dove:

$$c_{ij} = (p_{ij} + \sum_q p_{iq} p_{qj})^2, \quad q \neq i, j$$

Se C_i è uguale a 0, i ha molti contatti isolati tra loro, se C_i è uguale a 1, i ha un solo contatto. (per il significato di ρ si veda la formula 2 dell'indice effective size). L'indice di constraint ha una relazione negativa con il capitale sociale: Più l'attore è vincolato, minori saranno le opportunità per l'azione. Anche quest'ultimo è stato implementato da zero perché non esisteva nel pacchetto SNA. L'indice viene richiamato utilizzando la funzione constraint()

Constraint applicato nelle matrici A e B:

constraint A	constraint B
0.925926	1.0964506

- **Efficienza (Efficiency)**

“L'efficienza è, essenzialmente, il grado in cui un grafo usa il minor numero possibile di archi per connettere i propri vertici già connessi nel grafo”.

(Krackhardt, 1994)

L'efficienza è quindi una funzione della densità media delle componenti presenti nel grafo e rappresenta una misura della ridondanza della connettività.

L'indice di efficienza varia tra 0 (presenza di collegamenti ridondanti) e 1 (all'interno di ogni componente del grafo troviamo soltanto il numero minimo di legami indispensabili per tenere connessi i punti che ne fanno parte). Un grafo connesso e completo ha quindi efficiency pari a 0. Come conseguenza della definizione di efficienza, per valori vicini allo 0 ci aspettiamo di trovare legami simmetrici; viceversa, per valori vicini ad 1, ci aspettiamo di trovare legami asimmetrici. Anche questo indice è stato richiamato dalla libreria SNA (efficiency()).

Efficienza calcolata nelle matrici A e B:

efficiency A	efficiency B
0.000000	0.4444444

Nella matrice A, essendoci la presenza di tutti i collegamenti possibili tra gli attori, l'efficienza è pari a 0.

- **Gerarchizzazione (Hierarchy)**

Il concetto di gerarchia fa pensare a situazioni in cui un attore è in una posizione particolare rispetto ad altri appartenenti al medesimo gruppo. La gerarchizzazione misura quanto gli attori siano collegati da percorsi asimmetrici.

L'asimmetria della relazione può nascere sia da una gerarchia di comando, che da una gerarchia da attribuzione di prestigio. Nel primo caso, l'attore in posizione strategica esercita sugli altri il proprio volere; nel secondo caso lo stesso attore è destinatario delle preferenze di quelli che lo hanno scelto.

Il calcolo viene effettuato sulla matrice di raggiungibilità e rappresenta il rapporto tra percorsi direzionati non reciprocati sul totale di quelli presenti nella matrice. L'indice varia tra 1 (nessun percorso è reciprocato, le diadi sono tutte asimmetriche ed il grafo è fortemente gerarchizzato) e 0 (tutti i percorsi sono reciprocati, le diadi - "una coppia non ordinata di attori e gli

archi che esistono tra i due attori della coppia” - Wasserman - sono tutte simmetriche ed il grafo non è strutturato gerarchicamente) .

Anche questo indice viene richiamato dalla libreria SNA (hierarchy()).

Indice Hierarchy applicato delle due matrici:

hierarchy A	hierarchy B
0.000000	0.3333333

Nell'esempio riportato sopra la matrice A è simmetrica quindi anche le diadi lo sono: ecco spiegato il valore dell'indice.

- **Centralizzazione (Centralization)**

La centralizzazione di un grafo, insieme alla densità, rappresenta una proprietà a livello di grafo, e si costruisce a partire da diverse misure di centralità. Un indice di centralizzazione varia tra 0 ed 1: quanto più si avvicina

ad uno, tanto più avremo il grafo avrà un solo attore centrale, e molti attori periferici; viceversa, quanto più sarà vicino a 0, tanto più gli attori saranno ugualmente centrali. Per questa caratteristica l'indice di centralizzazione può essere considerato come una rozza misura di variabilità o dispersione, un modo di misurare quanto diseguali siano i valori dei singoli attori.

Anche questo indice è stato richiamato dalla libreria SNA (`centralization()`).

Indice `centralization` applicato sulle due matrici:

centralization A	centralization B
0.000000	0.6666667

Nella matrice A tutti gli attori sono legati tra loro e quindi l'indice vale 0.

3 Un caso di studio

In questa ultima parte del progetto si è svolta un'applicazione pratica per meglio comprendere i concetti teorici riportati nei capitoli precedenti.

Si tratta di un'indagine che mira a capire se, in qualche modo, i metodi di studio abbiano qualche dipendenza su fattori biografici degli studenti.

Raccolta dei dati:

L'indagine è stata svolta all'interno della facoltà di statistica dell'università degli studi di Padova. Su un campione di 50 studenti sono state rilevate:

- Corso di laurea (STI, SI, SEFA, SEF, SPS, SGI)
- Età (in anni compiuti)
- Genere (maschio, femmina)
- Lavoratore (SI, NO)
- Pendolare giornaliero o Domiciliato

Ad ogni rispondente si è poi chiesto di elencare tutti gli amici con i quali aveva studiato nel corso dell'ultimo anno scolastico ed il loro parere sul fatto che anche le persone che aveva menzionato avessero studiato assieme o meno.

Riportiamo un esempio di un questionario compilato. Si può notare che nell'ultima parte dell'intervista il rispondente "crea", con le sue risposte, una

matrice simmetrica rispetto alla diagonale. Questo, come già spiegato nei concetti teorici, è un fattore fondamentale per un reticolo ego-centrato.

Esempio di questionario compilato.

Il presente questionario ha l'obiettivo di rilevare le relazioni tra gli studenti della Facoltà di Scienze Statistiche per quanto riguarda i gruppi di studio

Rispondere riempiendo solo le celle bianche

Dati del rispondente

id	34	<i>numero progressivo</i>
corso di laurea	SEFA	<i>indicare la sigla</i>
genere	M	<i>F/M</i>
età	24	<i>in anni compiuti</i>
pendolare giornaliero/domiciliato	Domiciliato	<i>Pendolare/Domiciliato</i>
lavoratore	NO	<i>SI/NO</i>

Nell'ultimo anno con chi ti è capitato di studiare o fare esercitazioni in vista di un esame?

- Indica nelle celle bianche della prima colonna i nomi o nickname delle persone con cui hai studiato

Alla fine delle interviste si avranno:

- Un data-set contenente tutti i dati personali dei rispondenti
- n reticoli ego-centrati quanti sono gli intervistati (in questo caso 50)

Analisi dei dati:

Attraverso la funzione `read.egonet()` contenuta nella libreria creata, importiamo tutti i reticoli nell'ambiente R trasformandoli così in matrici di adiacenza utili per il calcolo degli indici sociali . Applichiamo dunque la funzione `index.egonet()` alle matrici per l'implementazione degli indici ed

uniamo il risultato di quest'ultima con i dati personali degli intervistati formando così un unico dataset su cui si andrà a lavorare.

Riportiamo un esempio di codice per costruire il dataset. Qualcosa di simile si può trovare anche nell'help del package `egonet`:

```
# ho un piccolo dataset di persone e le egonetwork nella cartella ./dati
dati <- read.csv("bio.csv", sep=";")
# importo tutti i file e li infilo in una oggetto "list"
mats <- lapply(files, read.egonet, sep=";")
# calcolo indici su tutte le matrici
idx <- sapply(mats, index.egonet)
# trasforma la matrice in dataset e aggiunge la colonna di nomi dei file
idx <- as.data.frame(t(idx))
idx <- cbind(idx, id=rownames(idx))
dati <- merge(dati, idx, by="id")
```

Questo caso di studio risulta particolarmente interessante anche nell'aspetto inferenziale. E' infatti verosimile presupporre che le reti ego centrate rilevate in questo studio non siano reciprocamente indipendenti. Questo problema si verifica ogni qualvolta la popolazione di riferimento risulta piccola rispetto al numero di soggetti campionati. Per gli studi riguardanti le reti ego centrate

questo problema risulta ancora più rilevante, è infatti immediato capire che la rete di studio (e quindi gli indici ad essa connessa) di due persone che studiano assieme sia molto simile in quando parzialmente sovrapposta. Tale aspetto critico può essere può inficiare il valore inferenziale dell'intera analisi. Una soluzione è ottenibile operando stratificazioni del campione in gruppi omogenei; ovviamente questi "gruppi omogenei" non sono noti, ma è ragionevole supporre che variabili quali il corso di laurea e il genere operino una prima importante riduzione di eterogeneità tra individui.

Affrontare con rigore questa questione non è tra gli scopi di questa tesi, in questa sede ci limiteremo solo a sottolineare questo punto e ad adottare la strategia della stratificazione ove possibile.

I test inferenziali adottati sono ANOVA e la regressione lineare esatte (non parametriche) che consentono di analizzare adeguatamente le distribuzioni non normali di questo campione relativamente piccolo. La strategia della stratificazione è inoltre facilmente raggiunto tramite una strategia di permutazione entro gli strati (Pesarin 2001).

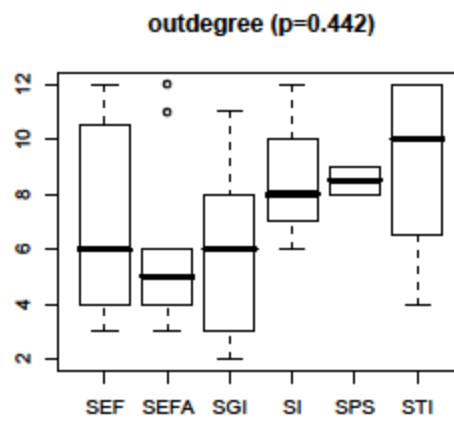
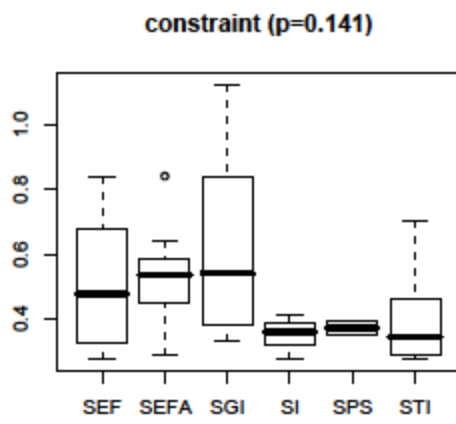
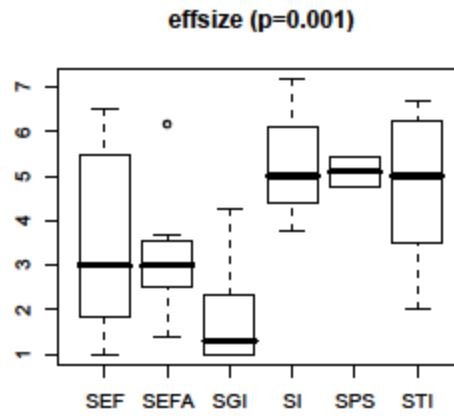
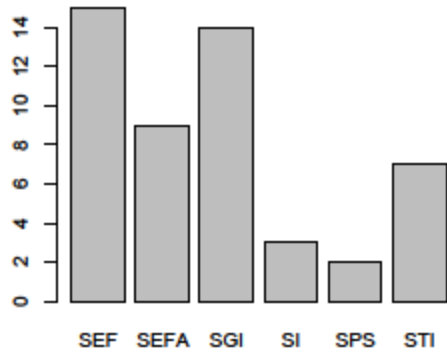
P-value dei test ANOVA non parametrica effettuati tra i dati biografici e gli indici.

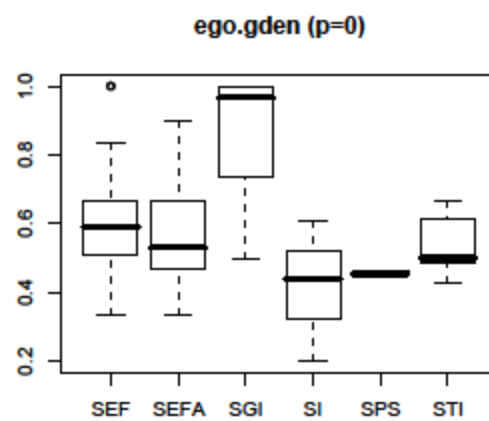
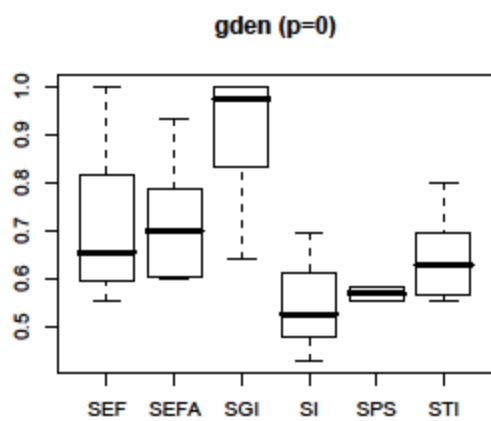
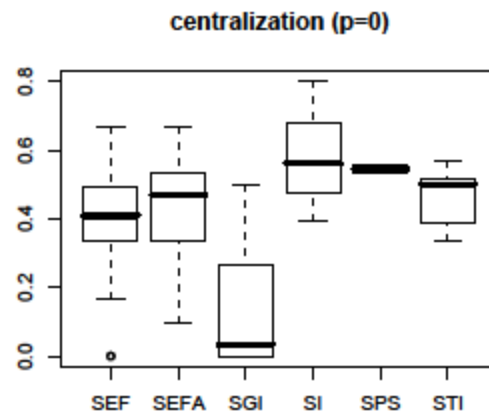
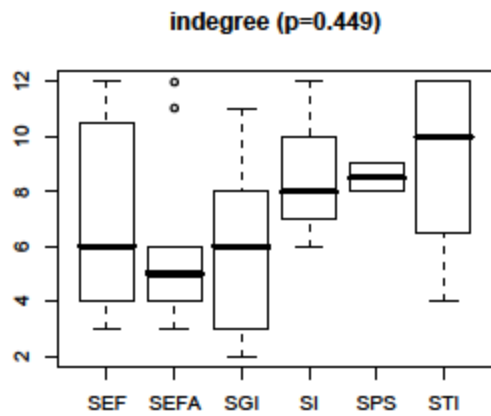
	lavoratore	genere	pendolare	laurea	età
effsize	0,7485	0,4951	0,2440	0,0002	0,1028

constraint	0,1702	0,8012	0,9898	0,1556	0,3483
outdegree	0,0818	0,8148	0,1974	0,4459	0,9914
indegree	0,0882	0,8200	0,2004	0,4499	0,9926
centralization	0,2394	0,2102	0,4049	0,0000	0,0022
gden	0,5133	0,2841	0,8714	0,0000	0,0072
ego.gden	0,2368	0,2110	0,4107	0,0000	0,0022

Dalla tabella che illustra le varie significatività dei p-value si nota come le variabili laurea ed età influiscano molto sugli indici di densità e di centralizzazione. Focalizziamo allora, con l'aiuto di alcuni grafici, l'analisi di questi fenomeni.

Box plot Corsi di Laurea vs Indici sociali

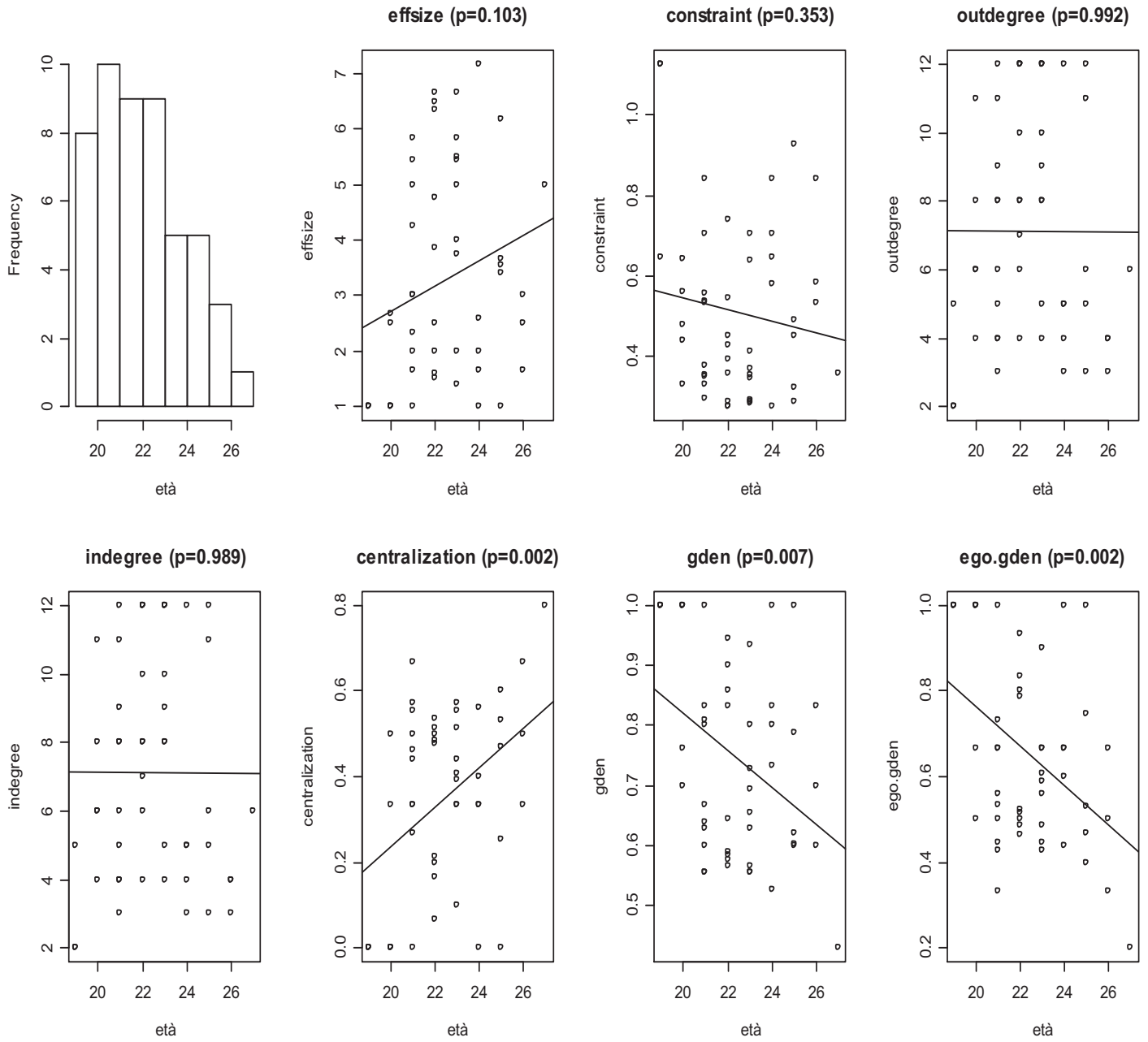




Da queste illustrazioni notiamo come l'appartenenza a diversi corsi di laurea provochino effetti diversi sugli indici sociali, soprattutto per gli ultimi tre (densità, densità ego-centrata e centralizzazione).

Una situazione molto interessante si nota, però, mettendo a confronto l'età con gli indici sociali.

Plot Età vs Indici sociali



Anche in questo caso si denota una dipendenza, anche se non fortissima, tra l'età e la variabilità degli indici. Quello che notiamo, di interessante, in questi grafici è come variano gli indici a seconda dell'età degli studenti. In particolare, aumentando l'età, la densità tende a diminuire mentre la centralizzazione tende a crescere. Questo denota un fatto preciso: con l'avanzare degli anni il reticolo di uno studente diventa sempre meno denso (ossia diminuiscono le relazioni all'interno della matrice) e ha sempre di più un solo attore centrale. Alla luce di questi risultati ci si aspettava anche che gli indici che misurano il grado del reticolo si comportassero in modo inversamente proporzionale alla variabile età (anche se il test non dà risultati molto significativi). Un motivo può essere trovato visionando il grafico che mette a confronto l'età con l'effsize. Essendo quest'ultimo un indicatore della capacità di ego di gettare ponti su buchi strutturali si può presumere che con l'aumentare dell'età l'attore principale acquisisce la capacità di essere mediatore di tante reti differenti.

Alla luce dei problemi metodologici accennati all'inizio del paragrafo, la significatività della relazione dell'età con gli indici considerati è stata valutata anche operando una stratificazione per corso di laurea e genere. I p-value ottenuti danno indicazioni analoghe all'analisi senza stratificazione con la sola differenza che le evidenze sono meno forti. Questo risulta facilmente giustificabile anche alla luce della ridotta numerosità campionaria.

Conclusioni

I risultati ottenuti da quest'analisi ci portano a concludere che con l'aumentare dell'età un singolo studente ha sempre meno bisogno di studiare in gruppo, un motivo potrebbe essere che nei primi due anni scolastici i corsi di studio raggruppano molti studenti, anche di corsi di laurea differenti. Con il passare del tempo i singoli attori si "specializzano", i corsi diminuiscono la loro numerosità ed ogni studente si affida a persone che possono avere accesso a informazioni differenti dalle proprie.

Riferimenti bibliografici

Borgatti, S.P., Jones, C. e Everett, M.G.

2005 *Le misure reticolari del capitale sociale*, in «Sociologia e Politiche Sociali», 1, pp. 91-100.

Bourdieu, P.

1986 *The forms of capital*, in Richardson (ed.) *Handbook of Theory and Research for the Sociology of Education*, New York, Greenwood, 241-258. Originally: in "Ökonomisches Kapital, kulturelles Kapital, soziales Kapital." in *Soziale Ungleichheiten* (Soziale Welt, Sonderheft 2), edited by Reinhard Kreckel. Goettingen: Otto Schartz & Co.. 1983. pp. 183-98.

Burt, R.S.

1992 *Structural Holes. The Social Structure of Competition*, Cambridge, MA, Harvard University Press.

2001 *Structural Holes versus Network Closure as Social Capital*, in Lin N., Cook K. e Id. (eds.), *Social Capital. Theory and Research*, Aldine de Gruyter, New York, pp.31-56.

2002 *The social capital of structural holes*, in M. F. Guillén: *The New Economic Sociology: Developments in an Emerging Field*. New York: Russell Sage Foundation, pp. 148-90

Chiesi, A.

1999 *L'analisi dei reticoli*, Milano, FrancoAngeli.

Coleman, J.

1990 *Foundations of Social Theory*, Cambridge and London, The Belknap Press of Harvard University Press, trad. it. *Fondamenti di teoria sociale*, Bologna, Il Mulino, 2005.

Donati, P. e Tronca, L.

2008 *Il capitale sociale degli italiani. Le radici familiari, comunitarie e associative del civismo*, Milano, FrancoAngeli

Pesarin, F.

2001 *Multivariate permutation tests: With application in biostatistics*. Padova, John Wiley and Sons

Scott, J.

1991 *Social Network Analysis. A Handbook*, London, Sage, trad. it. *L'analisi delle reti sociali*. Roma, La Nuova Italia Scientifica, 1997.

Tronca, L.

2007 *L'analisi del capitale sociale*, Padova, Cedam.

Wasserman, S. e Faust, K.

1994 *Social Network Analysis. Method and Applications*, Cambridge (MA), Cambridge University Press.

Ringraziamenti

Un grazie particolare v`a al dott. Livio Finos e al dott. Andrea Sciandra per il grande aiuto datomi per lo svolgere di questa tesi (non `e assolutamente di cortesia ma viene dal profondo del cuore).

Un grazie affettuoso v`a ai miei amici per avermi supportato e "sopportato": grazie Nicola, Luca, Donato, Andrea, Leon Muras, Grande Socio Cini.

Un grazie a Te, Michela.

Un grazie speciale v`a a tutti i miei compagni di universit`a, siete veramente i migliori....(Fabio e Francy su tutti).

Un grazie a pap`a (ti voglio bene anche se non si vede) e a tutta la famiglia Marsotto (siete unici!!!).

Nonna Edda e zio Massimo, che dire..... vi amo!!!

L'ultimo grazie ma il primo per importanza v`a alla donna che conta di pi`u nella mia vita: un esempio da seguire, una persona adorabile....Mamma, questa `e per Te. Sei la migliore al mondo!!!!