

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea Magistrale in Scienze di Internet

**RETI DI PARTECIPAZIONE FRA
SOCIETA' DI CAPITALE IN ITALIA:**

**PRESENZA DI TOPOLOGIE
CARATTERIZZATE DA
DISTRIBUZIONI POWER LAW**

Relatore:
Chiar.mo Prof.
MARCO RUFFINO

Presentata da:
ANDREA CAVICCHINI

Sessione III

Anno Accademico 2012-2013

Indice

1	Introduzione	9
2	Letteratura	13
2.1	Reti tra imprese	13
2.1.1	Classificazione dei gruppi aziendali	14
2.2	(Social) Network Analysis	18
2.2.1	Teoria dei grafi	20
2.2.2	Cluster analysis	21
2.2.3	Centralità	24
2.3	Modelli di rete	25
2.4	Struttura sociale della competizione	28
2.4.1	Forza della relazione	29
2.4.2	L'importanza dell'informazione	30
2.4.3	Structural holes	31
2.5	Distribuzione power-law	31
2.5.1	Stima dei parametri	33
2.5.2	Test di bontà della stima	35
3	Ipotesi di lavoro	36
4	Disegno e applicazione del metodo	39
4.1	Metodo	39
4.1.1	Input	39
4.1.2	Workflow	40
4.2	Il modello	48
4.3	Strumenti utilizzati	49
4.3.1	Node XL	49
4.3.2	Gephi	54

5	Dati	58
5.1	Fonti e Limiti	58
5.2	Dataset 1	59
5.2.1	Caratterizzazione	60
5.2.2	Rappresentazione della rete e analisi delle metriche . .	63
5.3	Dataset 2	66
5.3.1	Caratterizzazione	66
5.3.2	Rappresentazione della rete e analisi delle metriche . .	68
5.4	Dicotomizzazione basata sulla partecipazione azionaria	69
5.4.1	Rappresentazione della rete e analisi delle metriche . .	72
6	Analisi su componenti connesse	75
6.1	Considerazioni sui gruppi	75
6.2	Dataset 1	77
6.2.1	Componente gigante	77
6.2.2	Altre componenti	80
6.3	Dataset 2	88
6.3.1	Componente gigante	88
6.3.2	Altre componenti	91
6.4	Dicotomizzazione su Dataset 2	96
6.4.1	Componente gigante	96
6.4.2	Altre componenti	100
7	Centralità e potere	106
7.1	Potere e influenza nelle organizzazioni	110
7.2	Degree centrality	110
7.2.1	Degree centrality come coinvolgimento nella rete	111
7.3	Betweenness centrality	115
7.3.1	Betweenness centrality come capacità di intermediazione	116
7.3.2	Betweenness centrality come influenza nella rete	119
7.4	Eccentricità	122
7.4.1	Eccentricità come diametro delle partecipazioni	122
7.5	Closeness centrality	126
7.5.1	Closeness centrality come misura di potere	129
7.6	Risultati finali	134
8	Conclusioni e lavori futuri	138

<i>INDICE</i>	3
Bibliografia	141

Elenco delle figure

2.2.1 Crescita esponenziale delle pubblicazione contenenti la parola "social network" nell'abstract o nel titolo[7].	19
4.1.1 Matrice di adiacenza	40
4.1.2 Edge-list	41
4.1.3 Lista di adiacenza	41
4.1.4 Workflow	43
4.1.5 Esempio di edge-list dataset 2	44
4.1.6 Dataset unico	45
4.1.7 Rappresentazione iniziale di un grafo. I colori rappresentano i diversi cluster a seconda dei componenti connessi ad essi	47
4.2.1 Il modello	50
4.3.1 Esempio di edge-list con NodeXL	51
4.3.2 Esempio di utilizzo di Gephi	55
5.2.1 Dataset 1: vertex-list con vertici in ordine di degree	60
5.2.2 Dataset 1: distribuzione del degree	61
5.2.3 Dataset 1: distribuzione di frequenze in-degree e out-degree	62
5.2.4 Dataset 1: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza	63
5.2.5 Dataset 1: esempio di grafo	65
5.3.1 Dataset 2: distribuzione	67
5.3.2 Dataset 2: distribuzione di frequenze in-degree e out-degree	67
5.3.3 Dataset 2: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza	68
5.4.1 Dataset dicotomizzato: distribuzione degree su dataset dicotomizzato	71

5.4.2 Dataset dicotomizzato: distribuzione di frequenze in-degree e out-degree dicotomizzati	71
5.4.3 Dataset dicotomizzato: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza.	72
5.4.4 Rappresentazione parziale del grafo dicotomizzato basato su degree e betweenness	73
6.2.1 Dataset 1: distribuzione del degree nel gruppo principale	77
6.2.2 Dataset 1: distribuzione di in e out-degree nel gruppo principale	78
6.2.3 Dataset 1: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza. a) grafo diretto; b) grafo non diretto (in-degree); c) grafo non diretto (out-degree)	79
6.2.4 Esempio di power-law con cut-off	80
6.2.5 Dataset 1: distribuzione del degree nelle altre componenti di rete	82
6.2.6 Dataset 1: distribuzione di in e out-degree	83
6.2.7 Gruppo 2 (Rhodia Holding)	86
6.2.8 Gruppo 5 (Sigma Campania)	87
6.2.9 Gruppo 9 (Demag Cranes & Components GMBH)	89
6.3.1 Dataset 2: distribuzione del degree nel gruppo principale	90
6.3.2 Dataset 2: distribuzione di in e out-degree nel gruppo principale	90
6.3.3 Dataset 2: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza. a) grafo diretto; b) grafo non diretto (in-degree); c) grafo non diretto (out-degree)	91
6.3.4 Dataset 2: distribuzione del degree nelle altre componenti di rete	92
6.3.5 Dataset 2: distribuzione di in e out-degree nelle altre componenti di rete	94
6.3.6 Gruppo 6 (Poggiorsini)	95
6.3.7 Gruppo 7 (D.A.EM)	97
6.4.1 Dataset dicotomizzato: distribuzione del degree nel gruppo principale	98
6.4.2 Dataset dicotomizzato: distribuzione di in e out-degree nel gruppo principale	99

6.4.3 Dataset dicotomizzato: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza. a) grafo non diretto; b) grafo diretto (in-degree); c) grafo diretto (out-degree)	100
6.4.4 Dataset dicotomizzato: distribuzione del degree, altre componenti	101
6.4.5 Dataset dicotomizzato: distribuzione di in e out-degree nelle altre componenti di rete	103
6.4.6 Gruppo 10 (B.S.E. Business Service Express SRL)	104
7.0.1 Esempio di rete a stella	108
7.2.1 Degree centrality basata su out-degree (rete di influenza) . . .	112
7.2.2 Gruppi di gruppi [6]	113
7.2.3 Ego network "Fiat S.p.A."	114
7.3.1 Esempio di rete lineare. a) senza misura di betweenness; b) con misura di betweenness	115
7.3.2 Betweenness centrality (rete di influenza)	117
7.3.3 Ego network "Rizzoli Corriere della Sera mediagroup S.p.A." .	120
7.3.4 Betweenness centrality (rete di influenza)	121
7.3.5 Ego network "Impregilo S.p.A."	123
7.4.1 Eccentricità	125
7.4.2 Ego network "Marco Tronchetti Provera & C. Sapa"	127
7.5.1 Esempio di rete lineare. a) senza misura di closeness; b) con misura di closeness	127
7.5.2 Closeness centrality (rete di potere)	130
7.5.3 Ego network "Sintonia S.A."	132
7.5.4 Ego network "Aurelia S.p.A."	133
7.5.5 Ego network "Bim Fiduciaria e di Revisione S.p.A."	135
7.6.1 Riassunto dei principali gruppi di potere	137
8.0.1 Paesi con betweenness e closeness media più alta hanno sofferto meno durante la crisi economica[11]	140

Elenco delle tabelle

2.3.1 Comparazione qualitativa tra le caratterizzazioni di diversi modelli di rete e risultati empirici nel mondo reale	28
5.2.1 Numero di relazioni e vertici del dataset 1	59
5.2.2 Dataset 1: Risultati dell'analisi sulla distribuzione power-law. ”/” indica la divisione tra in-degree/out-degree	63
5.2.3 Dataset 1: risultati metriche	64
5.3.1 Numero di relazioni e vertici del dataset 1	66
5.3.2 Dataset 2: risultati dell'analisi sulla distribuzione power-law	68
5.3.3 Dataset 2: risultati metriche	69
5.4.1 Numero di relazioni e vertici del dataset 2 dicotomizzato	70
5.4.2 Dataset dicotomizzato: risultati dell'analisi sulla distribuzione power-law.	71
5.4.3 Dataset 2: risultati metriche con dicotomizzazione	74
6.1.1 Suddivisione del dataset 1 in gruppi	76
6.1.2 Suddivisione del dataset 2 in gruppi	76
6.2.1 Dataset 1: analisi delle metriche riguardanti il gruppo principale	78
6.2.2 Dataset 1: Risultati dell'analisi sulla distribuzione power-law, componente gigante. ”/” indica la divisione tra in-degree/out-degree	79
6.2.3 Dataset 1: analisi della distribuzione di in-degree e out-degree nelle altre componenti di rete	81
6.2.4 Dataset 1: analisi della distribuzione dell'in e out-degree	84
6.3.1 Dataset 2: analisi delle metriche riguardanti il gruppo principale	88

6.3.2 Dataset 2: Risultati dell'analisi sulla distribuzione power-law, componente gigante.	88
6.3.3 Dataset 2: analisi della distribuzione del degree, altre componenti	93
6.3.4 Dataset 2: analisi della distribuzione di in-degree e out-degree, altre componenti	96
6.4.1 Dataset dicotomizzato: analisi delle metriche riguardanti il gruppo principale	99
6.4.2 Dataset dicotomizzato: Risultati dell'analisi sulla distribuzione power-law, componente principale	100
6.4.3 Dataset dicotomizzato: analisi della distribuzione del degree, altre componenti	102
6.4.4 Dataset dicotomizzato: analisi della distribuzione di in-degree e out-degree, altre componenti	105

Capitolo 1

Introduzione

Una fra le questioni centrali della *digital era* è il passaggio dall'informazione alla conoscenza, questa seconda intesa come il significato contenuto potenzialmente nell'informazione. I Big Data possono essere definiti come una raccolta di dataset (insieme di dati) per loro natura di grandi dimensioni. Se da un lato ciò può far immaginare rilevanti potenzialità intrinseche, dall'altro è chiaro che la quantità e l'organizzazione dell'informazione non per "pattern di conoscenza" determina costi cognitivi di accesso ed interpretazione umanamente non affrontabili. In tal modo la quantità, da risorsa, si trasforma presto in un forte vincolo d'uso. Anche approcci statistici tradizionali di natura inferenziale possono rivelarsi non totalmente funzionali alla "conversione" in conoscenza di questo tipo di informazione, soprattutto laddove il significato potenziale è legato alla natura ed alla topologia delle relazioni che legano gli oggetti rappresentati dai big data. In questi casi, approcci strutturalisti o connessionisti, basati su tecniche computazionali e di rappresentazione che partono dall'assunto di definire "gli oggetti a partire dalle loro relazioni", possono rivelarsi molto più efficaci.

Questo è il caso trattato dalla tesi. Il dataset di riferimento è formato da un enorme insieme di dati in formato tabellare, rappresentanti le partecipazioni azionarie di 171.889 imprese, per oltre 500.000 record, a cui si aggiunge, per un importante subset, un ricco insieme di dati economico-finanziari. Siamo nel campo dei big data, rappresentativi di un mondo reale (la porzione manifatturiera dell'economia italiana, vista nel 2007). L'idea alla base di questo elaborato è quella di leggere i dati utilizzando le tecniche e i metodi dall'analisi delle reti, modellizzandoli in termini di nodi (le singole imprese) e relazioni (le quote di capitale possedute, in termini di partecipazioni).

Sulla scorta dei lavori di Barabási e Albert [4], la dimensione del dataset porta ad ipotizzare che l'insieme delle imprese così definito segua una ben precisa legge naturale, nota come distribuzione power-law, o legge di potenza, scale-free. Ciò come conseguenza di due meccanismi genetici secondo i quali le reti si espandono in continuazione con l'aggiunta di nuovi nodi, e questi nuovi nodi si aggregano, preferibilmente a nodi che sono già ben collegati (*preferential attachment*). Secondo gli autori, lo sviluppo di grandi reti è governato da robusti fenomeni di auto-organizzazione che vanno oltre gli elementi dei singoli sistemi. Oltre che nel campo big data siamo in questo caso anche in quello delle macro-organizzazioni complesse. In senso economico, è del resto abbastanza intuitivo comprendere come la asimmetria di dotazione di risorse (e dunque di potere) favorisca in un prospettiva ecologica di *self-organization* proprio gli attori maggiori, maggiormente capaci di svolgere un ruolo "attrattivo" delle ulteriori risorse presenti nel sistema. Questo fenomeno è noto anche con il nome di "*rich get richer*" (i ricchi diventano sempre più ricchi).

Bak, Tang e Wiesenfeld nel 1987 [2] furono i primi ad introdurre il concetto di criticità auto-organizzata (*self-organized criticality*) per spiegare il comportamento del loro modello a pile di sabbia (detto anche modello Abeliano o appunto modello di Bak-Tang-Wiesenfeld). Questa teoria venne utilizzata in seguito per spiegare diversi fenomeni naturali, compreso quello dei mercati azionari. Una delle caratteristiche più importanti dei sistemi critici auto-organizzati è che mostrano un comportamento che segue una distribuzione power-law.

Quando la probabilità di misurare un particolare valore di una certa quantità varia inversamente come potenza di tale valore, il quantitativo è detto come seguente una power-law, conosciuta anche come legge di Zipf o distribuzione di Pareto. Le power-law appaiono ampiamente in vari campi scientifici quali la fisica, la biologia, l'economia e la finanza, l'informatica, le scienze sociali. Ad esempio [31]: la frequenza di occorrenze di parole uniche nel libro Moby Dick di Herman Melville [17]. Il grado (*degree*) di nodi nella rappresentazione parzialmente conosciuta di Internet al livello dei sistemi autonomi¹ riguardante Maggio 2006 [26]. Il numero di chiamate a lunga distanza ricevute dai clienti di AT&T negli Stati Uniti durante una singola giornata [25].

¹Un sistema autonomo è un gruppo di indirizzi IP in Internet tramite i quali il routing è gestito internamente o autonomamente, piuttosto di usare i meccanismi di routing del protocollo di gateway a larga scala di Internet.

Il numero di best seller venduti negli stati uniti durante il periodo 1895-1965 [21]. La popolazione delle città statunitensi nel rapporto Census del 2000 [31]. e molti altri..

Uno dei problemi principali deriva dalla rappresentazione dei dati. Una rete formata da centinaia di migliaia di nodi e relazioni non potrà mai risultare rappresentabile in senso grafico, né la sola osservazione “a vista”, del resto, può essere sufficiente per esprimerne le proprietà. E’ dunque necessario un approccio modellistico che ne consenta la computabilità, attraverso la scelta e l’applicazione di alcune misure sintetiche ed esplorative, compatte e di immediata rappresentabilità.

Ritornando alla definizione di big data precedente, questi dataset devono possedere principalmente 4 caratteristiche che riguardano il *volume*, inteso come accesso a grandi volumi di dati, la *velocità*, come capacità di analizzare dati in un breve arco di tempo, la *varietà* e la *veridicità*.

La rappresentazione chiara e veloce di una mole importante di dati è uno dei punti chiave del nostro lavoro di tesi. Una volta creato un modello chiaro e scalabile è possibile applicarlo per:

- analizzare nel dettaglio alcune strutture presenti all’interno della rete, arrivando ad una possibile generalizzazione di struttura societaria all’interno di un mercato finanziario;
- analizzare i nodi influenti, cioè quelle entità che, meglio di altre, hanno maggior potere all’interno della rete;
- rappresentare i dati che rispecchiano determinate caratteristiche, in termini di rete;
- dividere la rete in sottoreti di connessioni tra nodi;
- ecc...

Obiettivo di questa tesi sarà principalmente quello di verificare se il campione esteso di imprese segue la power-law (e se sì, in che limiti). A tale fine si configureranno i dati in un formato di rete monomodale, della quale si studieranno alcune macro-proprietà di struttura a livello complessivo e con riferimento alle componenti (i singoli subnet distinti) di maggior dimensione. Successivamente si compiranno alcuni approfondimenti sulla struttura fine di alcuni subnet, essenzialmente rivolti ad evidenziare la potenza di un

approccio network-based, anche al fine di rivelare rilevanti proprietà nascoste del sistema economico soggiacente, sempre, ovviamente, nei limiti della modellizzazione adottata.

In sintesi, ciò che questo lavoro intende ottenere è lo sviluppo di un approccio alternativo al trattamento dei *big data* a componente relazionale intrinseca (in questo caso le partecipazioni di capitale), verso la loro conversione in “*big knowledge*”: da un insieme di dati cognitivamente inaccessibili, attraverso la strutturazione dell’informazione in modalità di rete, giungere ad una conoscenza sufficientemente chiara e giustificata.

Capitolo 2

Letteratura

2.1 Reti tra imprese

La *rete* può essere concepita come una forma di organizzazione delle attività economiche in grado di governare la ragnatela di interdipendenze che connette gli individui. Le interdipendenze e le relazioni tra individui, i gruppi e le unità organizzative così come quelle verso l'esterno sono al centro della vita dell'organizzazione [36].

Le *relazioni* tra aziende possono avere svariate forme: dallo scambio dei beni, servizi o informazioni, alle relazioni di associazione fondate sulla messa in comune di obiettivi, risorse e la divisione del rischio. Le relazioni tra organizzazioni rappresentate in questo elaborato possono evidenziare intrecci di scambi, l'esistenza di un orientamento reciproco, la presenza di fiducia tra le parti e, soprattutto, la condivisione di informazioni.

Il concetto di *reti tra imprese* è molto ampio, si può riferire alla trama di relazioni non competitive che connette entità istituzionalmente diverse, senza intaccarne l'autonomia formale in assenza di una direzione e un controllo unitario. In questo caso si parla di forme di organizzazione basate sulla cooperazione e il coordinamento tra imprese o altre organizzazioni che si trovano sotto condizioni di interdipendenza, che può essere verticale, cioè lungo la catena di valore (fornitore o cliente), o orizzontale, cioè tra imprese che svolgono le medesime attività e operano nello stesso ambiente.

Il concetto di rete può anche rappresentare l'insieme di relazioni date dalla partecipazione al capitale sociale di un'azienda, dove non si creano meccanismi di cooperazione e di coordinamento. Questa relazione può derivare

quindi dalla teoria della crescita dell'impresa di Penrose [34], secondo cui le imprese possono ridurre sostanzialmente i propri rischi diversificando la produzione, creando economie di scala che riducono i costi medi per far fronte alla crescita dimensionale necessaria a causa della struttura del mercato, aumentando il potere di mercato mediante l'attuazione di strategie volte al controllo di altre società.

In questo caso l'utilizzo degli strumenti offerti dall'analisi delle reti sociali può aiutare nell'individuazione dei confini di un gruppo, e quindi anche dell'impresa. Le relazioni di cui si dispone all'interno di questo elaborato saranno solo relazioni di partecipazione azionaria, in cui ogni azione costituisce una partecipazione al capitale sociale, e attribuisce al suo titolare un complesso unitario di diritti e poteri di natura amministrativa e patrimoniale. Alcuni diritti dell'azionista sono indipendenti dal numero di azioni possedute (ad esempio il diritto di intervento in assemblea), mentre è chiaro che i diritti più significativi spettano in proporzione del numero di azioni possedute (ad esempio diritto di voto, diritto agli utili e alla quota di liquidazione, diritto di opzione).

Una relazione diretta tra due nodi A e B non solo indicherà che, ad esempio, il nodo A possiede una determinata quantità del capitale sociale di B , ma indica anche che B potrà avere una dipendenza soggettiva da A , a seconda della percentuale di partecipazione al nodo.

Di Carlo [15] propone una distinzione tra "*dipendenza soggettiva*" e "*autonomia decisionale*" indicando che una società è dipendente quando è dominata da un soggetto economico che esercita il suo dominio attraverso una società controllante; è invece indipendente quella società non controllata da altri soggetti, ma direttamente da persone fisiche, che nel caso di un gruppo coincide con la holding di primo livello. Il soggetto economico di gruppo, detentore del potere di comando (o anche "*ultimate owner*"), può delegare alle singole l'autonomia decisionale, ma tale autonomia non si può definire assoluta, bensì relativa, data dalla possibilità di revoca del soggetto controllante.

2.1.1 Classificazione dei gruppi aziendali

Di Carlo [15] propone una classificazione dei gruppi aziendali in base a diversi criteri:

1. Natura delle combinazioni economiche aggregate:

- (a) *gruppi verticali*: le unità del gruppo hanno una serie continua di scambi tra loro;
- (b) *gruppi orizzontali*: le varie unità operano nello stesso settore economico, producendo prodotti o servizi simili;
- (c) *gruppi diversificati*: le unità sono tra loro autonome a livello di business praticato.

2. Natura dei legami esistenti tra le diverse unità del gruppo:

- (a) *gruppi finanziari*: i legami tra le varie entità sono puramente finanziari;
- (b) *gruppi economici*: presenza di intensi legami tecnico-economici e finanziari;
- (c) *gruppi misti*.

3. Direzione strategica della capogruppo:

- (a) *gruppi patrimoniali*: la capogruppo si limita a gestire le partecipazioni sulla base di valutazioni economiche, senza intervenire nella gestione delle controllate;
- (b) *gruppi finanziari*: la capogruppo esercita una funzione di controllo attraverso la nomina e la revoca degli amministratori, influenzando nella scelta dei dirigenti;
- (c) *gruppi industriali*: la holding interviene in maniera più o meno marcata nella direzione strategica delle controllate;
- (d) *gruppi imprenditoriali*: la capogruppo gestisce le unità controllate in modo completamente integrato e accentrato.

4. Natura del soggetto economico che controlla il gruppo:

- (a) *gruppi pubblici*: è lo Stato o un ente pubblico a dominare il gruppo;
- (b) *gruppi privati*: chi controlla le unità dell'aggregazione è un soggetto economico privato.

5. Proprietà e modello di governance:

- (a) *gruppi ad azionariato diffuso (widely-held)*: il capitale della holding è suddiviso in una miriade di azionisti, nessuno dei quali è in grado di esercitare il controllo;
- (b) *gruppi ad azionariato concentrato (large shareholders)*: gruppi in cui la maggioranza del capitale è detenuta da un individuo, da una famiglia (*family-owned*) o dallo Stato (*State-owned*).

6. Estensione geografica dell'attività:

- (a) *gruppi locali*: l'attività delle consociate è svolta in una zona specifica (ad esempio, regione, provincia o comune) del territorio nazionale;
- (b) *gruppi nazionali*: le consociate hanno tutte la medesima nazionalità e operano nell'intero territorio nazionale;
- (c) *gruppi internazionali*: realizzano processi di internazionalizzazione, l'attività svolta all'estero rimane secondaria;
- (d) *gruppi multinazionali*: agiscono in più stati con consociate di diversa nazionalità, e quindi non hanno un mercato di riferimento.

7. Forma giuridica delle unità produttiva:

- (a) *gruppi societari*: le varie unità del gruppo assumono forma societaria, S.p.A. o S.r.l., e sono controllate dalla holding;
- (b) *gruppi personali*: le unità hanno forma individuale e non societaria;
- (c) *gruppi misti*: nei gruppi di minore dimensione le unità produttive non sono controllate da una società capogruppo, ma direttamente da persone fisiche.

8. Grado di integrazione tecnico-economica e comportamento direzionale perseguito:

- (a) *gruppi strategici*: caratterizzati da elevata integrazione economica e da una logica di gestione unitaria;
- (b) *gruppi finanziari*: modesta integrazione tecnico-economica e direzione unitaria;

- (c) *gruppi patrimoniali*: caratterizzati da una bassa integrazione tecnico-economica e dall'assenza di logica di gestione unitaria;
- (d) *gruppi formali*: elevata integrazione tecnico-economica con l'assenza di logica di gestione unitaria.

9. Modalità di configurazione del controllo:

- (a) *gruppi a struttura semplice*: caratterizzati da partecipazioni dirette di controllo della capogruppo in tutte le società controllate;
- (b) *gruppi a struttura complessa*: nei quali il controllo viene esercitato a cascata o a mezzo di successivi livelli di controllo;
- (c) *gruppi a catena*: nei quali i poteri di controllo si compongono variamente a mezzo di partecipazioni incrociate.

10. Natura del vincolo di controllo:

- (a) *gruppi partecipativi*: si realizzano quando una società riesce a controllarne un'altra attraverso una quota di maggioranza, assoluta o relativa, del suo capitale (controllo *equity*);
- (b) *gruppi contrattualistici*: due imprese stipulano un contratto che può essere di subordinazione o di dominio con il quale una delle due società rinuncia alla propria indipendenza per partecipare al gruppo (controllo *non equity formale*);
- (c) *gruppi con vincoli commerciali o finanziari*: quando due imprese stabiliscono tra loro rapporti e vincoli commerciali e/o finanziari tali per cui si limita l'autonomia del soggetto passivo di tale rapporto, rendendolo così dipendente (controllo *non equity informale*).

11. Natura della capogruppo:

- (a) *gruppi controllati da holding pure*: caratterizzati dalla presenza di holding nelle quali è assente l'attività produttiva (*funzione headquarter*);
- (b) *gruppi controllati da holding miste*: sommano alle funzioni proprie, di pianificazione, finanza e controllo anche quelle produttive.

12. Attività svolta dal gruppo:

- (a) *gruppi bancari*;
- (b) *gruppi industriali*;
- (c) *gruppi assicurativi*;
- (d) *gruppi commerciali*.

2.2 (Social) Network Analysis

L'analisi delle reti sociali è una moderna metodologia di analisi delle relazioni sociali sviluppatasi a partire dai contributi di Jacob Levi Moreno [30], fondatore della *sociometria*, scienza che analizza le relazioni interpersonali.

Nonostante l'origine, negli ultimi anni (come mostrato in figura 2.2.1), la social network analysis è diventata una generale metodologia di analisi dei dati basata su un approccio relazionale che è entrata a far parte di tutti i domini. Da Internet al World Wide Web fino ad arrivare alle reti in economia, di trasmissione delle malattie, e di studio di punti sensibili ad attacchi terroristici, l'analisi delle reti sociali si è diffusa ampiamente nella cultura moderna. Secondo Newman, Barabàsi e Watts [32] le reti si possono trovare ovunque e possono servire per descrivere molti eventi naturali.

In questo elaborato si utilizzerà, in modo metaforico dato che in questo elaborato si studieranno interazioni tra attori economici, la definizione di analisi delle reti sociali proposta da Linton Freeman [18] secondo cui:

Nelle scienze sociali, l'approccio strutturale che è basato sullo studio dell'interazione tra attori sociali è chiamato analisi delle reti sociali. L'approccio si basa sulla nozione intuitiva che i modelli dei legami sociali in cui gli attori sono incardinati abbiano conseguenze importanti per quegli stessi attori. L'analisi delle reti sociali è motivata da un'intuizione strutturale basata sui legami che collegano gli attori sociali, si radica in dati empirici e sistematici. In questa analisi è presente l'uso, in modo consistente, di rappresentazioni grafiche e si basa sull'uso di modelli matematici e/o computazionali.

L'analisi delle reti sociali, che sta espandendo sempre di più il suo campo negli ultimi anni, si distingue dai precedenti lavori sulle reti in 3 modi:

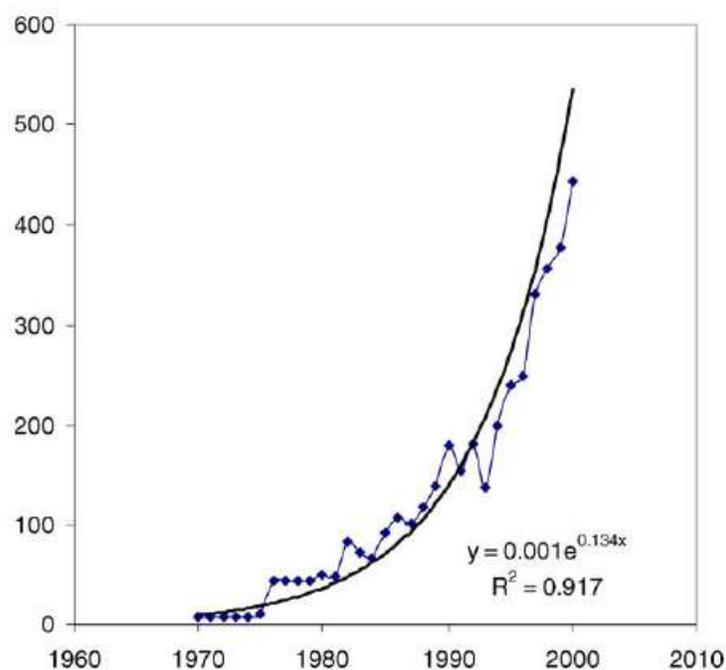


Figura 2.2.1: Crescita esponenziale delle pubblicazione contenenti la parola "social network" nell'abstract o nel titolo[7].

- focalizza la sua analisi sulle proprietà delle reti rappresentanti il mondo reale;
- ritiene che le reti non siano statiche, ma che possano evolvere nel tempo in base ad una serie di regole dinamiche;
- cerca di capire le reti non solo come oggetti topologici, ma anche come modelli su cui i sistemi dinamici sono costruiti.

Questa materia veicola l'idea in base alla quale la società può essere considerata con un *intreccio complesso di relazioni sociali variamente strutturate* [35], questo intreccio costituisce il focus centrale dell'analisi.; la metodologia nella social network analysis è posta al centro dell'analisi, ne costituisce una delle parti principali.

2.2.1 Teoria dei grafi

Un *grafo* $G = (V, E)$ è costituito da due set: un set finito di elementi V chiamati *vertici* e un set finito di elementi E chiamate *relazioni*. Ogni relazione è identificata da un paio di vertici. Se le relazioni del grafo G sono identificate da coppie di vertici ordinate, allora G è chiamato *grafo diretto*, o *orientato*, altrimenti, G si definisce *grafo non diretto*, o *non orientato*. I grafi sono utilizzati per la loro funzione di modello matematico della struttura di rete.

Due vertici sono *adiacenti* se sono i vertici terminali di qualche relazione. Se due relazione hanno un vertice terminale in comune, allora queste relazioni sono definite come adiacenti.

Il numero di relazioni incidenti su un un vertice v è chiamato *degree* di un vertice, denotato con $d(v_i)$, un vertice con $degree = 0$ è chiamato *vertice isolato*. In un grafo non diretto, la somma dei degree dei vertici di un grafo G è uguale a $2m$, dove m è il numero di relazioni di G .

Considerando un grafo $G = (V, E)$, $G' = (V', E')$ è definito come *sottografo* di G se V' e E' sono, rispettivamente, sottoinsiemi di V e E tali che una relazione (v_i, v_j) appartiene a E' solo se v_i e v_j sono in V' . G' sarà un sottografo *corretto* di G se anche E' è un sottoinsieme corretto di E o V' è un sottoinsieme corretto di V . Se tutti i vertici di un grafo G sono presenti in un sottografo G' di G , allora G' è definito come un sottografo *ricoprente* di G .

Un *cammino* in un grafo $G = (V, E)$ è una sequenza finita alternativa di vertici e relazioni $v_0, e_1, v_1, e_2, \dots, v_{k-1}, e_k, v_k$ che inizia e termina con vertici tali che v_{i-1} e v_i sono i vertici terminali della relazione e_i , con $1 \leq i \leq k$. Il numero di relazioni in un cammino è chiamato lunghezza del cammino.

La *distanza* tra due vertici u e v in G , denotata con $d(u, v)$, è la lunghezza del cammino minimo $u - v$ in G . Se non esiste questo cammino, allora si definiscono $d(u, v)$ come infiniti. Il *diametro* di G , indicato con $diam(G)$, è la distanza massima tra due vertici e G .

Un concetto importante della teoria dei grafi riguarda la *connessione*. Due vertici v_i e v_j si dicono aventi una connessione in un grafo G se esiste un cammino $v_i - v_j$ in G , un grafo si definisce *connesso* se esiste un cammino che connette ogni paio di vertici in G .

Se si considera un grafo $G = (V, E)$ *non connesso*, l'insieme di vertici V e G possono essere partizionati in sottoinsiemi V_1, V_2, \dots, V_p tali che il sottografo di vertici indotti V_i sono connessi e nessun vertice nel sottoinsieme

V è connesso a nessun vertice del sottoinsieme V_j , con $j \neq i$. I sottografi V_i sono chiamati *componenti* di G . Se il grafo non è collegato, allora si rompe naturalmente in una serie di pezzi, cioè gruppi di nodi con la proprietà per cui ogni gruppo è connesso se considerato come grafo in isolamento, e non come sovrapposizione di gruppi.

Il *coefficiente di clustering* è la misura che indica il grado con cui i nodi all'interno di un grafo tendono a raggrupparsi assieme. In letteratura, nella maggior parte delle reti sociali rappresentanti dati empirici, i nodi tendono a creare gruppi caratterizzati da una *densità* relativamente elevata di legami a maglia. Il coefficiente di clustering globale è basato sulle terzine di nodi. Supponendo che un vertice v abbia k_v vicini, quindi possono esistere al massimo $k_v(k_v - 1)/2$ relazioni tra di loro (questo accade quando ogni vicino di v è connesso ad ogni altro vicino di v). C_v indica la frazione di queste relazioni ammissibili che in realtà esistono. C è definita come la *media* di C_v su tutte le v [38].

All'interno di sistemi complessi di larghe dimensioni, spesso si trova ciò che si chiama il *componente gigante*, termine informale per definire un componente connesso che contiene una frazione significativa di tutti i nodi. Una rete solitamente contiene solo un componente gigante al suo interno, nella maggior parte dei casi è essenzialmente inconcepibile che due grandi insiemi di relazioni non abbiano punti di contatto, infatti, nelle reti reali, è un fenomeno che non si vede quasi mai, è quindi generalmente unico e distinguibile come componente che supera di gran lunga tutti gli altri.

2.2.2 Cluster analysis

Al di là dell'apparente sinonimia con il termine "coefficiente di clustering", presentato nel paragrafo precedente, la cluster analysis non ha a che fare con l'analisi delle reti sociali ma è un insieme di tecniche di *analisi multivariata* dei dati volta alla selezione o raggruppamento di elementi omogenei in un insieme di dati. In questa tesi le tecniche di cluster analysis non saranno utilizzate per classificare nella stesso stesso cluster casi indipendenti, bensì si applica alla rete per classificare nello stesso cluster casi dipendenti, in modo che ad ogni gruppo sia corrisposta componente di rete.

In questo caso abbiamo un insieme di relazioni all'interno di un'unico gruppo, tramite tali relazioni saremo in grado di ottenere una composizione di $N = \{n_1, n_2, \dots, n_{t-1}, n_t\}$ reti, con $N > 0$.

A seconda della tecnica di clustering ci si può basare principalmente su due teorie:

- *Bottom-Up* o *Metodi Aggregativi* (dal basso verso l'alto): questa tecnica prevede che inizialmente tutti gli elementi siano considerati cluster a sè, l'algoritmo provvede ad unire i cluster più vicini. L'algoritmo continuerà ad unire elementi al cluster fino a che la distanza minima tra i cluster non supera un determinato criterio statistico prefissato.
- *Top-Down* o *Metodi Divisivi* (dall'alto verso il basso): all'inizio tutti gli elementi sono un unico cluster, l'algoritmo inizia poi a dividere il cluster in tanti cluster di dimensione inferiore. Il criterio che guida la divisione è quello di ottenere gruppi sempre più omogenei. L'algoritmo procede fino a che non viene soddisfatta una regola di arresto.

Esistono molte tecniche per la divisione di un dataset in gruppi, queste dipendono dal tipo di relazione che si desidera mantenere:

- *raggruppamento basato sugli attributi dei vertici*: si creano gruppi basandosi sui vertici del grafo utilizzando un discriminatore X scelto a priori;
- *raggruppamento basato sul componente connesso*: i gruppi sono creati in base alle componenti dei grafi strettamente connesse. Il risultato del raggruppamento è la creazione di un grafo connesso. In teoria dei grafi, un grafo $G = (V, E)$ è detto connesso se, per ogni coppia di vertici $(u, v) \in V$, esiste un cammino che collega u a v . Un sottografo connesso di un grafo non orientato è detto componente connessa di tale grafo. Questo raggruppamento si basa solo sulle relazioni tra i nodi.
- *Algoritmo di Clauset-Newman-Moore*: questo metodo si basa sulla *modularità*, come proprietà di una rete e di una specifica divisione di tale rete in comunità. Dati gli elementi di una matrice di adiacenza A_{vw} che avranno come valori

$$\begin{cases} 1 & \text{se i vertici } v \text{ e } w \text{ sono connessi} \\ 0 & \text{altrimenti} \end{cases},$$

e si suppone che i vertici siano divisi in comunità tali che il vertice v appartenga alla comunità c_v . Quindi la frazione di relazione che fa parte

di una comunità, ad esempio in grado di connettere vertici che fanno parte della stessa comunità è

$$\frac{\sum_{vw} A_{vw} \delta(c_v, c_w)}{\sum_{vw} A_{vw}} = \frac{1}{2m} \sum A_{vw} \delta(c_v, c_w)$$

Dove la funzione $\delta(i, j)$ è 1 se $i = j$ e 0 altrimenti, $m = \frac{1}{2} \sum_{vw} A_{vw}$ è il numero di relazioni nel grafo. A questi elementi si aggiunge il degree k_v , inteso come il degree del vertice v e definito come:

$$k_w = \sum A_{vw}$$

Quindi si definisce la *modularità* Q come:

$$Q = \frac{1}{2m} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w).$$

- *Algoritmo di Wakita-Tsurumi*: algoritmo simile al precedente ma migliorato in termini di velocità di calcolo. Questo tipo di algoritmo è più idoneo per reti con più di 500,000 nodi arrivando a processare i legami 7 volte più velocemente dell'algoritmo CNM
- *Algoritmo di Girvan-Newman*: individua comunità dalla rimozione progressiva delle relazioni dalla rete originale. Le componenti connesse della restante rete sono le comunità, l'algoritmo si focalizza quindi sulle relazioni tra comunità. La *betwenness* viene studiata come misura di centralità e di influenza dei nodi nelle reti. Per ogni nodo i , la *betwenness* del vertice è definita come il numero di cammini minimi tra coppie di nodi che lo percorrono. L'algoritmo si comporta in questo modo:

1. calcola la *betwenness* di tutte le relazioni esistenti nella rete;
2. la relazione con la *betwenness* più alta viene rimossa;
3. la *betwenness* di tutte le relazioni colpite dall'eliminazione viene ricalcolata;
4. si ripetono i passi 2 e 3 fino a quando non rimangono più relazioni.

Il risultato finale dell'algoritmo è la distanza dei casi rispetto ai cluster. Questo risultato viene spesso rappresentato sotto forma di dendrogramma.

2.2.3 Centralità

Nella teoria dei grafi e nell'analisi delle reti, la *centralità* di un vertice è la misura della sua relativa importanza all'interno di un grafo, mostra la quantità, e la qualità, delle connessioni di un nodo avente connessioni con altri. Le misure principali di misurazione della centralità sono 4:

- *degree centrality*: è definita come il numero di link incidenti su un nodo (numero di legami del nodo). Nel caso di reti dirette (dove i legami hanno associata una direzione) il degree si differenzia in *in-degree* e *out-degree*. L'in-degree è il conteggio del numero di legami diretti al nodo e l'out-degree il numero di legami che il nodo dirige verso altri.

La degree centrality di un vertice v , di un dato grafo $G = (V, E)$ con V vertici e E relazioni, è definita come: $C_D(v) = deg(v)$;

- *betweenness centrality*: è una misura di centralità del vertice all'interno del grafo, quantifica il numero di volte in cui un nodo agisce come intermediario lungo il cammino più breve tra gli altri due nodi. La betweenness di un vertice v in un grafo $G = (V, E)$ con V vertici è calcolata come segue:

1. per ogni coppia di vertici (s, t) , calcola il *cammino minimo* tra di loro;
2. Per ogni coppia di vertici (s, t) , si determina la frazione del cammino minimo che passa dal vertice in questione (in questo caso v);
3. somma la frazione su tutte le coppie di vertici (s, t) .

La betweenness centrality può quindi essere rappresentata come:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

dove σ_{st} è il numero totale di cammini minimi dal nodo s al nodo t e $\sigma_{st}(v)$ è il numero di questi cammini passanti per v ;

- *closeness centrality*: è definita come la distanza naturale tra tutte le coppie di nodi, definita dalla lunghezza del loro percorso più breve.

La distanza (*farness*) di un nodo è definita come la somma delle sue distanze con gli altri nodi, e la sua vicinanza (*closeness*) è definita come l'inverso della sua distanza;

- *Eigenvector centrality*: è la misura dell'influenza di un nodo in una rete. Assegna punteggi relativi a tutti i nodi di una rete basandosi sul concetto che le connessioni a nodi con alto punteggio contribuiscono maggiormente al punteggio del nodo in questione rispetto a connessioni di nodi uguali con un punteggio basso. Dato un grafo $G = (V, E)$ con V vertici e sia $A = (a_{v,t})$ la matrice di adiacenza, il punteggio di centralità di un vertice v può essere definito come

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum a_{v,t} x_t,$$

dove $M(v)$ è un'insieme di vicini di v e λ è una costante.

2.3 Modelli di rete

I lavori di ricerca di Watts [39] e Barabàsi [3] descrivono tre modelli generali di rete: *casuali*, *small-world* e *scale-free*. Questi modelli sono caratterizzati dal modo in cui queste reti sono generate, e da diverse statistiche risultanti, come la distribuzione del degree, il cammino minimo medio tra i nodi e il grado di clustering medio.

Nei modelli descritti si assume che ci siano n nodi che formano una data rete; si assume anche che i collegamenti siano non diretti.

Reti casuali: una rete casuale è creata specificando che ogni paio di nodi è connesso da un collegamento con probabilità uniforme p . Queste reti sono state studiate da un punto di vista puramente matematico da Erdős e Rènnyi [16]. Nel limite di n grande, molte delle proprietà delle reti casuali possono essere espresse analiticamente, tuttavia, alcune proprietà chiave del modello di rete casuale Erdős-Rènnyi sono piuttosto divergenti da quelle viste nelle reti contenenti dati empirici.

Innanzitutto, le reti casuali non esibiscono un coefficiente di clustering forte dei nodi visti tramite dati empirici. I comportamenti di clustering avvengono solitamente quando, ad esempio, una persona è più

facilmente amica di un amico di amico rispetto ad una persona a caso. Il coefficiente di clustering di una rete casuale è quindi esattamente p . Tuttavia in molti casi di studio empirici sul mondo reale si nota che il valore del coefficiente di clustering è di ordini superiori rispetto al valore del grafo casuale con lo stesso numero di nodi e relazioni.

Secondo, si può dimostrare che la distribuzione del degree nelle reti casuali ha la forma di una gaussiana approssimata (Poisson per grandi n), che è diversa dalla distribuzione power-law che solitamente si osserva nelle reti reali

Reti Small-World: Nel 1998, Watts e Strogatz introdussero il modello di rete small-world. Per costruire tale modello di rete, si inizia con un automa cellulare con forma di anello reticolare in cui ogni nodo si collega a k dei suoi nodi vicini. Per ogni link, con probabilità piccola p si ricollega un nodo ad un altro selezionato casualmente all'interno della rete. Per $p = 0$ il risultato è una rete completamente regolare; per $p = 1$ il risultato è una rete completamente casuale. Per p bassi e diversi da zero, il risultato è una rete con molte connessioni locali e poche connessioni a lunga distanza. Tale rete è definita come avente proprietà small-world, anche se sono presenti relativamente poche connessioni a lunga distanza, il cammino minimo (in salti tra link) tra due nodi scala logaritmicamente o più lentamente con dimensioni della rete n con un degree medio fisso. Questo indica che, anche nelle reti small-world con molti nodi, il cammino minimo tra due nodi individuali è probabilmente basso, da cui deriva il nome "small-world". Al contrario, in una rete connessa regolarmente, la lunghezza del cammino medio scala linearmente con la dimensione della rete.

Reti small-world con p basso e diverso da zero mostrano anche un alto coefficiente di clustering. Al contrario nelle reti casuali si aveva un cammino minimo medio basso e coefficiente di clustering basso. Tuttavia, il modello di Watts e Strogatz di solito dà come risultato una distribuzione del degree che non corrisponde con la maggior parte delle reti basate sul mondo reale.

Reti Scale-Free: Barabási e Albert svilupparono un modello alternativo a quello di Watts e Strogatz, in cui la crescita della rete è basata su "attaccamento preferenziale" (*preferential attachment*), che produce la

rete scale-free. Un network scale-free è semplicemente una rete in cui la distribuzione del degree segue una power-law.

All'interno di una rete con distribuzione gaussiana del degree, la maggioranza dei nodi ha approssimativamente lo stesso numero di legami, è molto raro trovare una grande deviazione dalla media. Al contrario, nelle reti scale-free, sono presenti grandi variazioni nel numero di collegamenti per un nodo. Sono presenti piccoli insiemi di nodi, detti "hub", con molti link, e molti insiemi di nodi con un piccolo numero di nodi.

Nel 1999 Barabàsi e Albert proposero il meccanismo della "crescita per attaccamento preferenziale"[4] come spiegazione dei dati empirici delle reti scale-free. A loro insaputa, riscoprirono le osservazioni fatte da Simon e Price precedentemente, e cioè la semplice idea che i ricchi diventano sempre più ricchi ("rich get richer"). Le reti crescono in un modo tale che i nodi con il degree maggiore ricevono più collegamenti nuovi rispetto a nodi con un degree più basso. Intuitivamente si possono riscontrare molte analogie con il mondo reale, ad esempio, le persone con molti amici tendono a conoscere sempre più nuove persone e quindi con l'elevata possibilità di fare nuovi amici, diversamente dalla possibilità delle persone che possiedono pochi amici.

La crescita di rete proposta da Barabàsi e Albert si comporta come segue: la rete inizia con un piccolo numero di nodi iniziali. Ad ogni passo, si aggiunge un nuovo nodo, e si connette agli m nodi esistenti, per qualche m . La probabilità di fare uno di quei collegamenti al nodo i è proporzionale al degree del nodo i .

In molti studi di reti nel mondo reali, Barabàsi e Albert hanno ipotizzato che la distribuzione empirica del degree sia illustrata tramite una power-law con α compreso tra 2 e 3 (dipendente dalla rete), concludendo che il meccanismo che guida l'evoluzione di reti nel mondo reale è la "crescita con attaccamento preferenziale". Nel mondo reale, piccole differenze nell'abilità o anche fluttuazioni puramente casuali possono rimanere bloccate e portare a grandi disuguaglianze nel tempo [3]. Si può anche dimostrare che alcune reti scale-free, non tutte, godono anche della proprietà small-world.

Mitchell in [29] descrive le caratteristiche qualitative delle varie caratterizzazioni di reti con varie distribuzioni del degree, in particolare

si sofferma sul coefficiente di clustering e sulla lunghezza media del cammino (tabella 2.3.1).

Modello di rete	Distribuzione del degree	Coefficiente di clustering	Lunghezza media del cammino
Regolare	costante	alto	alto
casuale	Poisson	basso	basso
Small world (probabilità p bassa e ≥ 0)	dipende da p	alto	basso
Scale-free	power-law	alto	basso
Risultati empirici su rete del mondo reale	power-law	alto	basso

Tabella 2.3.1: Comparazione qualitativa tra le caratterizzazioni di diversi modelli di rete e risultati empirici nel mondo reale

Secondo Mitchell, non è realistico assumere che la probabilità della creazione di un collegamento ad un nodo sia esattamente proporzionale al suo degree; nel mondo reale, tutti i nodi non sono identici, ma variano le loro proprietà intrinseche. Tuttavia, se viene modificata l'assunzione di probabilità proporzionale, la distribuzione del degree potrebbe non essere più una power-law.

Mitchell conclude che il fatto sorprendente che emerge dalla presentazione di questi modelli è che, nonostante tutti i presupposti elencati in precedenza, questi modelli sembrano catturare alcuni aspetti essenziali delle reti reali.

2.4 Struttura sociale della competizione

Secondo Burt [9] un soggetto porta 3 tipologie di capitale all'interno di un contesto competitivo:

1. *capitale finanziario*, disponibilità economiche;
2. *capitale umano*, le qualità umane;

3. *capitale sociale*, relazioni con altri soggetti.

Dal capitale sociale si ottengono opportunità, nel senso più ampio del termine, per l'utilizzo del capitale umano e finanziario. I capitali umani e finanziari sono distinti in due modi dal capitale sociale, innanzitutto, perchè riguardano le proprietà dell'individuo, possedute in parte o completamente da un singolo individuo, persona o corporazione. La seconda differenza è che questi due capitali riguardano gli investimenti nell'equazione di produzione del mercato, vengono utilizzati per creare possibilità di produzione.

Il capitale sociale è differente in entrambi gli aspetti: è una cosa posseduta congiuntamente dalle parti della relazione, nessuno può avere diritti di proprietà sul capitale sociale, e il capitale sociale riguarda il tasso di rendimento nell'equazione di produzione del mercato.

Più precisamente, il capitale sociale è importante quanto l'imperfezione della concorrenza e l'abbondanza dei capitali di investimento. In concorrenza perfetta, il capitale sociale sarebbe una costante nell'equazione di produzione, ci sarebbe un unico tasso di rendimento perchè il capitale si muoverebbe liberamente dal basso rendimento degli investimenti all'alto, fino a quando i tassi di rendimento sono omogenei tra investimenti alternativi. Quando la concorrenza è, invece, imperfetta, il capitale è meno mobile e svolge un ruolo più complesso nell'equazione di produzione; *la competizione non è mai perfetta*. Esiste un ruolo svolto dalla struttura della rete del soggetto e dalla posizione dei contatti del soggetto stesso nella struttura sociale che forniscono un vantaggio competitivo, e quindi la possibilità di ottenere tassi più alti di rendimento.

2.4.1 Forza della relazione

Il concetto di forza delle relazioni è stato introdotto da Granovetter [20] che definisce la forza di una relazione interpersonale come la combinazione, probabilmente lineare, del grado di intensità emotiva, del livello di intimità, dei servizi reciproci e del tempo ad essa dedicato dalle parti. Il lavoro di ricerca di Granovetter ha dimostrato che, per interpretare alcuni fenomeni reticolari, le connessioni deboli sono a volte più importanti dei legami forti.

Ad esempio, i legami deboli possono costruire dei *"local bridges"* che sono posizionati in aree della rete per cui possono essere in grado di connettere due cluster altrimenti non collegabili. La maggiore efficacia dei legami deboli deriva dal caso studiato da Granovetter secondo cui nel mercato del lavoro,

i parenti o gli amici stretti (*legami forti*) si muovono in ambienti prossimi a quelli propri, mentre i semplici conoscenti (*legami deboli*) hanno più probabilità di conoscere diversi ambienti dal proprio ed hanno, quindi, più possibilità di accedere ad informazioni non note.

Un'altra determinante della forza di una relazione può essere identificata nella prossimità o similarità degli attori, come condivisione degli stessi valori e appartenenza alla stessa categoria.

2.4.2 L'importanza dell'informazione

I benefici informativi di una rete definiscono chi conosce le opportunità che la rete ha da offrire, quando le conoscono, e chi vi può partecipare. I soggetti con una rete perfettamente strutturata per accedere a questi benefici ottengono alti tassi di rendimento nei loro investimenti, come conseguenza del fatto che questi soggetti hanno più informazione e quindi più opportunità.

Un elemento fondamentale per la creazione di reti che beneficino di ricche informazioni sono i contatti all'interno di essa. Questi contatti saranno stabiliti in posizioni in cui sarà più probabile il passaggio di grandi insiemi di informazioni utili, e forniranno un flusso affidabile di informazioni da e verso quei luoghi. Da questo punto si ottiene che la scelta dei contatti è un'operazione critica, da cui si possono derivare questioni di fiducia che, in un contesto di competizione imperfetta, è fondamentale.

All'interno di un sistema competitivo perfetto, si fa affidamento al sistema per ottenere un rendimento equo sugli investimenti. In un sistema di competizione imperfetta, ogni soggetto può fare affidamento solo ai propri contatti.

La dimensione dell'insieme dei contatti può rivelarsi utile nell'ottenere informazioni, si aumenta perciò l'esposizione ad informazioni preziose, ma, aumentare la dimensione della rete senza considerare la diversità può rendere inefficiente la rete in vari modi. Secondo Burt in [9] ciò che conta all'interno di una rete è il numero di contatti non ridondanti; i contatti sono ridondanti nella misura in cui portano agli stessi soggetti, e così forniscono gli stessi benefici informativi, questo porta alla successiva definizione, sempre di Burt, di *structural holes*.

2.4.3 Structural holes

Gli structural holes (o buchi strutturali) indicano la separazione tra contatti non ridondanti, i contatti non ridondanti sono connessi da buchi strutturali. Uno structural hole è una relazione di non ridondanza tra due contatti. Come risultato di questa teoria si otterranno contatti che avranno indicatori di additività delle informazioni, eliminando le sovrapposizioni.

Il numero di buchi strutturali all'interno di una rete aumenta con l'aumentare della dimensione della stessa (soprattutto se si tratta di reti scale-free), gli structural holes rappresentano la chiave per accedere ai benefici informativi. I benefici offerti dalla rete sono raggiunti in vari modi. I contatti non ridondanti permettono l'esposizione del soggetto a diverse fonti di informazione, in cui ogni cluster di contatti è una fonte indipendente. Un cluster rappresenta una sola fonte di informazione, perchè i soggetti connessi ad esso tendono a conoscere le stesse informazioni nello stesso periodo di tempo.

2.5 Distribuzione power-law

Matematicamente, una quantità x segue una power-law se si estrae dalla distribuzione di probabilità

$$p(x) \propto x^{-\alpha}$$

dove α è un parametro costante della distribuzione conosciuto come "parametro di scala", o "parametro esponenziale". Il parametro scalare tipicamente risiede in un range $2 < \alpha < 3$. In pratica, pochi fenomeni empirici seguono una power-law per tutti i valori di x .

Di solito una power-law si applica solo per valori maggiori di un certo minimo x_{min} , in questi casi si dice che la coda della distribuzione segue una power-law.

Una distribuzione power-law si divide in due formati:

- *Distribuzione continua*: che disciplina numeri reali e continui.
- *Distribuzione discreta*: dove la quantità di interesse può prendere solo un determinato set di valori, di solito integrali positivi.

Assumendo che x rappresenti la quantità di cui si cerca la distribuzione, una distribuzione power-law continua è descritta da una probabilità di densità $p(x)$ tale che:

$$p(x) dx = Pr(x \leq X < x + dx) = Cx^{-\alpha} dx$$

dove x è il valore osservato e C la costante normalizzata. Questa densità diverge per $x \rightarrow 0$, e l'equazione non tiene per tutti gli $x \geq 0$, ci deve essere quindi un limite inferiore per la power-law. Questo limite si chiamerà x_{min} . Quindi, assunto $\alpha > 1$, si può calcolare la costante normalizzata

$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$

nel caso discreto x avrà solo un set discreto di valori. Considerando solo valori interi con una distribuzione di probabilità della forma

$$p(x) = Pr(x = X) = Cx^{-\alpha}$$

la distribuzione diverge per $x \rightarrow 0$, si deve così avere un limite inferiore $x_{min} > 0$ nella legge di potenza, e calcolando la costante normalizzata si ottiene che

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})},$$

dove $\zeta(\alpha, x_{min}) = \sum_{n=0}^{\infty} (n + x_{min})^{-\alpha}$

è la funzione generalizzata o funzione zeta di Hurwitz. La *funzione zeta di Hurwitz* è definita come:

$$\zeta(s, a) = \sum_{x=0}^{\infty} \frac{1}{(n + a)^s}$$

In molti casi è utile considerare anche la funzione di *distribuzione cumulativa complementare*, o CDF di una variabile distribuita power-law, in cui si indica $P(x)$ e che sia per il caso continuo che discreto è definito come $P(x) = Pr(X \geq x)$, per esempio, nel caso continuo:

$$P(x) = \int_x^{\infty} p(x') dx' = \left(\frac{x}{x_{min}} \right)^{-\alpha+1},$$

nel caso discreto

$$P(x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{min})}.$$

Come proposto da Clauset, Shalizi e Newman [12], la ricetta per l'analisi di dati distribuiti secondo una power-law prevede:

1. Stima dei parametri x_{min} e del modello power-law
2. Calcolo della bontà della stima (goodness-of-fit) tra i dati e la power-law. Se il risultante p -value è maggiore di 0,1 la power-law è una ipotesi plausibile per i dati, altrimenti è rigettata.

2.5.1 Stima dei parametri

Il primo obiettivo è quello di stimare correttamente i valori empirici all'interno di una distribuzione power-law. Gli studi su distribuzioni empiriche che seguono una power-law di solito danno alcune stime del parametro scalare α e anche il limite inferiore della regione scalare x_{min} . Utilizzando il logaritmo su entrambi i lati dell'equazione power-law, si nota che la distribuzione power-law segue la formula $\ln p(x) = \alpha \ln x + costante$, ciò indica che seguirà una linea retta su grafico log-log.

Una via comune, secondo Clauset, Shalizi e Newman [12], per provare un comportamento power-law, è quella di misurare la quantità x di interesse, costruire un istogramma rappresentante la sua distribuzione di frequenza, e disegnarlo su doppia scala logaritmica. Se dal grafico emerge una rappresentazione che approssimativamente segue una linea retta, si può affermare che la distribuzione segue una power-law, con un parametro di scala α dato dalla pendenza assoluta della linea retta. La pendenza è estratta eseguendo una regressione lineare ai minimi quadrati sul logaritmo dell'istogramma.

Stima di α : per stimare α nel modo corretto si ha bisogno di un valore per il limite inferiore x_{min} del comportamento power-law all'interno dei dati, al momento si assume che questo valore sia conosciuto. Il metodo utilizzato per la stima di un modello parametrizzato come la power-law è il metodo della *maximum likelihood* (massima verosimiglianza), il quale riesce ad indicare stime accurate dei parametri all'interno di grandi insiemi di dati. Supponendo che i dati siano derivati da una distribuzione che segue una power-law esattamente per $x > x_{min}$, si possono derivare gli stimatori della verosimiglianza massima (MLE)

del parametro scalare sia per il caso discreto che continuo. L'MLE per il caso continuo è descritto nel modo seguente:

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^n \ln \frac{x_i}{x_{min}} \right]^{-1}$$

dove x_i , con $i = 1 \dots n$ rappresenta i valori osservati di x in cui $x_i \geq x_{min}$.

L'equazione è equivalente allo *stimatore di Hill* [24], che riguarda un approccio di statistica inferenziale riguardante lo studio della coda di una distribuzione, che è asintoticamente normale e consistente ($\hat{\alpha} \rightarrow \alpha$ all'interno di un campione di n ampio).

Stima di x_{min} : prima di stimare il parametro di scala α , si devono scartare tutti i campioni al di sotto del valore x_{min} in modo che rimangano solo i valori per cui il modello power-law è considerabile come valido. Per ottenere una stima accurata di α si dovrà avere un metodo per la stima di un valore accurato di x_{min} . L'MLE dà un valore accurato quando x_{min} è scelto in modo aderente al valore reale, e devia rapidamente sotto questo punto (perchè la distribuzione devia da una power-law), e più lentamente al di sopra (perchè diminuisce la dimensione del campione).

Il metodo di stima di \hat{x}_{min} più comune può essere applicato sia per i dati continui che discreti. L'idea è di prendere un valore di \hat{x}_{min} che rende la distribuzione di probabilità dei dati misurati e il modello di power-law "best-fit" sopra \hat{x}_{min} il più simile possibile. Per il nostro tipo di dati, la miglior misura per quantificare la distanza tra due distribuzioni di probabilità è la *Kolmogorov-Smirnov* (o *statistica KS*), che semplicemente è la distanza massima tra la funzione di densità cumulativa dei dati e del suo modello stimato:

$$D = \max_{x \geq x_{min}} |S(x) - P(x)|$$

dove $S(x)$ è la funzione di densità cumulativa dei dati per l'osservazione con valore di almeno x_{min} , e $P(x)$ è la funzione di densità cumulativa del modello power-law che meglio rappresenta i dati nella regione $x \geq x_{min}$. Quindi il valore stimato \hat{x}_{min} è il valore di x_{min} che minimizza D .

2.5.2 Test di bontà della stima

Dati una serie di dati e una distribuzione power-law ipotizzata da cui i dati provengono, si vuole conoscere se questa ipotesi è plausibile. Un approccio standard è quello di utilizzare i *test di bontà della stima* (*goodness-of-fit test*), che genera un p -value che quantifica la plausibilità dell'ipotesi. Questi test sono basati sulla misura della distanza tra la distribuzione dei dati empirici e il modello ipotizzato; questa distanza è confrontata con le misure di distanza per dati sintetici comparabili per lo stesso modello, e il p -value è definito come la frazione della distanza sintetica maggiore rispetto alla distanza empirica. Se p è grande (vicino ad 1), allora la differenza tra i dati empirici e il modello può essere attribuita alle sole fluttuazioni statistiche, se è piccola, il modello non è plausibile per i dati.

Capitolo 3

Ipotesi di lavoro

Dato un dataset iniziale in formato tabellare, rappresentante circa 500.000 diadi di relazioni, tra soggetti, che possono essere persone fisiche o imprese, si vuole arrivare alla raffigurazione di questi dati in formato di network.

La maggior parte dei problemi complessi riguardanti il mondo reale può essere rappresentata nella forma di topologia reticolare, formata da nodi e collegamenti. In questo modo, si può mappare il problema come una rappresentazione della topologia nelle reti complesse, utilizzandone quindi gli stessi strumenti.

Ogni relazione presente all'interno del dataset porterà un incremento del degree del nodo da cui "esce" il collegamento (out-degree), nel caso di grafo diretto, o entrambi i nodi, nel caso di grafo non diretto. Le nostre prime ipotesi sono basate sulla distribuzione del degree che il dataset potrebbe avere.

Ipotesi 1 - secondo Barabási e Albert[4], buona parte delle reti complesse rappresentanti dati empirici mostrano un'indipendenza dal sistema e dall'identità delle sue componenti. La probabilità $P(k)$ che un vertice nella rete interagisca con k altri vertici segue una power-law, e quindi $P(k) \sim k^{-\gamma}$. Questo indica che le reti di grandi dimensioni si auto-organizzano in uno stato detto *scale-free*, che è caratteristica non prevista dai modelli di rete casuale di Erdős e Renyi[16]. La prima ipotesi di lavoro proposta riguarda la distribuzione dei degree del nostro dataset, rappresentante una porzione delle partecipazioni azionarie della Borsa italiana, e se questa segue una distribuzione power-law.

Ipotesi 2 - all'interno di una rete complessa di larghe dimensioni è probabile che sia presente un componente connesso che contiene una frazione significativa di tutti gli altri nodi (detto "*componente gigante*"). La seconda ipotesi riguarda la variazione della distribuzione nel caso in cui il dataset venga diviso in gruppi in base alla presenza di componenti connesse (cioè fino a quando è presente un legame tra i nodi di un gruppo), e fino a quale punto la rete possa seguire tale comportamento.

Ipotesi 3 - la tabella presentata nel capitolo precedente illustra i valori di coefficiente di clustering e di lunghezza media del cammino che i vari modelli di rete possiedono[29]. La terza ipotesi riguarda l'aderenza delle caratteristiche del nostro dataset al modello di power-law basata su dati empirici presentato in letteratura.

Una volta eseguito il calcolo del degree e lo studio della distribuzione, si arriva quindi alla costruzione di un modello, ancora parziale, al quale vanno aggiunti i dati finanziari, dell'anno 2007, presenti nel dataset originale.

A questo punto si ottiene un modello rappresentante i dati della rete insieme ai dati finanziari, ed è possibile utilizzare gli strumenti forniti dalla *Social Network Analysis* per "*estrarre conoscenza*" da questa serie di dati in formato tabellare. L'utilizzo delle *analisi di centralità* (betweenness centrality, degree centrality e closeness centrality), tipiche dell'analisi delle reti sociali potrebbe permetterci di dare una chiave di lettura della nostra rete. Le misure di centralità sono spesso accomunate a rappresentazioni del potere, locale o globale, nello studio delle reti sociali.

Secondo [13] la nozione di *potere* è un problema centrale degli studi sociologici. La questione fondamentale porta alla distinzione tra potere come proprietà di individui (potente per merito di alcuni attributi) e la misura in cui il potere sia definibile come proprietà di una struttura di rete (potente perchè detiene una posizione importante nella struttura sociale).

Nelle relazioni sociali, un nodo che gode di alta betweenness, e che quindi è al centro di molti cammini tra nodi, è solitamente indicato come un nodo potente o influente nel suo contesto[14]. Nel nostro caso, e cioè in un contesto economico, le misure di centralità saranno calcolate soprattutto sul dataset rappresentante il grafo diretto, e con percentuale di partecipazione. Questo ci porta a due successive ipotesi.

Ipotesi 3 - la misura di betweenness centrality, in letteratura [14], viene spesso accomunata alla nozione di potere all'interno di un network. Questo potere deriva dalla possibilità di accedere a maggiori flussi informativi rispetto agli altri nodi della rete. La terza ipotesi di lavoro riguarda la concezione di potere e betweenness centrality in un contesto economico, in cui le relazioni avranno un valore di partecipazione, e potrebbero portare ad una diversa interpretazione di questa misura di centralità.

Ipotesi 4 - calcolando le misure di centralità presentate in letteratura, insieme al calcolo del diametro delle partecipazioni, su dataset dicotomizzato con partecipazione minima al 20%, si ipotizza che all'interno di queste analisi si possa definire una possibile rappresentazione di "rete di potere" di un nodo e della sua influenza nei confronti del network.

Capitolo 4

Disegno e applicazione del metodo

In questo capitolo si illustreranno i metodi utilizzati per trasformare il dataset iniziale in un modello cognitivamente più semplice da utilizzare; verranno indicate le scelte per dividere questo dataset in cluster significativi. Infine verrà illustrata la composizione del nuovo dataset, fino alla creazione del modello, una prima illustrazione e la descrizione delle componenti principali che verranno studiate dettagliatamente nel capitolo successivo.

4.1 Metodo

4.1.1 Input

Un *dataset* indica una generale collezione di dati, i quali in questo caso sono un'insieme di relazioni tra vertici. Una *relazione* nell'ambito della network analysis può essere rappresentata in vari modi, tra i quali:

- *matrice di adiacenza*: è una matrice in cui ogni riga, o colonna rappresenta nodi e vertici differenti. In una matrice di adiacenza non pesata, le relazioni sono rappresentate da 0 e 1, che rappresentano l'indicazione della connessione tra due nodi. Se due nodi sono connessi si dicono adiacenti, da cui il nome matrice di adiacenza. In una matrice pesata, si possono avere valori differenti indicanti la qualità della relazione. Le matrici di adiacenza sono molto utili soprattutto quando si utilizzano grafi pesati (figura 4.1.1);

	Giovanni	Paolo	Caterina	Lucia	Riccardo	Stefania	Mario
Giovanni	0	1	0	1	0	1	0
Paolo	1	0	1	1	0	1	0
Caterina	0	1	0	0	0	0	0
Lucia	1	1	0	0	1	1	0
Riccardo	0	0	0	1	0	1	0
Stefania	1	1	0	1	1	0	1
Mario	0	0	0	0	0	1	0

Figura 4.1.1: Matrice di adiacenza

- *edge-list*: una *edge-list* è una lista di vertici, rappresentata su due colonne, connessi ad una rete. Il legame è dato dalla presenza dei due nodi nella stessa riga, ogni relazione avrà il valore di 1 nel caso di grafi non diretti. L'*edge-list* è sicuramente il metodo più intuitivo e veloce per creare una rete di relazioni da un insieme di dati (figura 4.1.2).
- *lista di adiacenze*: una lista di adiacenze, o lista di nodi, presenta il nodo focale sulla prima colonna, che rappresenta quindi il nodo stesso, e le altre eventuali relazioni (adiacenze) nelle colonne alla destra di esso (figura 4.1.3).

Nel nostro caso, si è scelto di utilizzare una “*edge-list*” per la sua facilità di compilazione e per la sua estrema compatibilità con il dataset originale.

4.1.2 Workflow

In figura 4.1.4 si mostra il *workflow* indicante tutti i processi per arrivare alla creazione della rete principale, delle sotto-reti e i processi di analisi e trasformazione. Da queste operazioni si arriverà alla creazione di due dataset:

- *dataset 1*: *edge-list* in formato (nome 1, nome 2);
- *dataset 2*: *edge-list* in formato (nome 1, nome 2, valore del legame).

V1	V2

Giovanni	Mario
Paolo	Caterina
Lucia	Paolo
Giovanni	Caterina
Mario	Giovanni
Giulio	Lucia
Giulio	Caterina
Lucia	Mario
Stefania	Lucia

Figura 4.1.2: Edge-list

Focale	Interazione 2	Interazione 3	Interazione 4	Interazione 5

Mario	Giovanni	Caterina	Lucia	
Giovanni	Mario	Paolo	Caterina	Lucia
Paolo	Giovanni			
Caterina	Mario	Giovanni	Giulio	Lucia
Giulio	Caterina	Lucia		
Lucia	Mario	Giovanni	Caterina	Giulio
Stefania	Lucia			

Figura 4.1.3: Lista di adiacenza

Non tutti i contenuti del dataset iniziale saranno utili per la rappresentazione dei dati, di seguito si descriveranno le operazioni compiute per la trasformazione del dataset:

Unione dei dataset iniziali: il dataset iniziale si presenta sotto la forma di due file dal contenuto in formato tabellare rappresentanti: i soggetti che partecipano all'azienda x , nella forma x partecipata da y con %, e le partecipazioni dell'azienda x in altre aziende, nella forma x partecipa in y con %. Viene quindi effettuata una operazione di unione tra i dataset e per fare ciò si è deciso di utilizzare una forma unica per tutto il dataset: la forma x partecipa in y con %.

Scelta dei parametri da utilizzare: si rappresentano quindi le partecipazioni azionarie di una parte della borsa italiana rappresentate sotto forma di nodi e relazioni che costituiranno N reti. Il dataset dato in input verrà trasformato sotto forma di "edge-list" composto da triple ($Vertex\ 1$, $Vertex\ 2$, %), rappresentato in figura 4.1.5.

- $Vertex\ 1$ = nome della società partecipante in $Vertex\ 2$.
- $Vertex\ 2$ = nome della partecipata.
- % = percentuale di partecipazione.

Il dataset unico è formato da una tabella di 515.343 righe per 47 colonne, come presentato parzialmente in figura 4.1.6. Vengono poi effettuate operazioni che servono a ridurre il numero di dati che non serviranno per le trasformazioni successive:

- eliminazione delle colonne non funzionali al progetto;
- eliminazione delle righe in cui mancano i valori riguardanti la percentuale (per il dataset 2);
- eliminazione delle righe duplicate;
- eliminazione dei simboli che indicano quote di partecipazione maggiori o minori di un certo numero, per motivi di semplicità, sono utilizzate senza l'operatore e mantenendo la quota indicata;
- Sostituzione dei vertici con il nome "Self Owned", indicanti la partecipazione della società in sé stessa, con il nome della società.

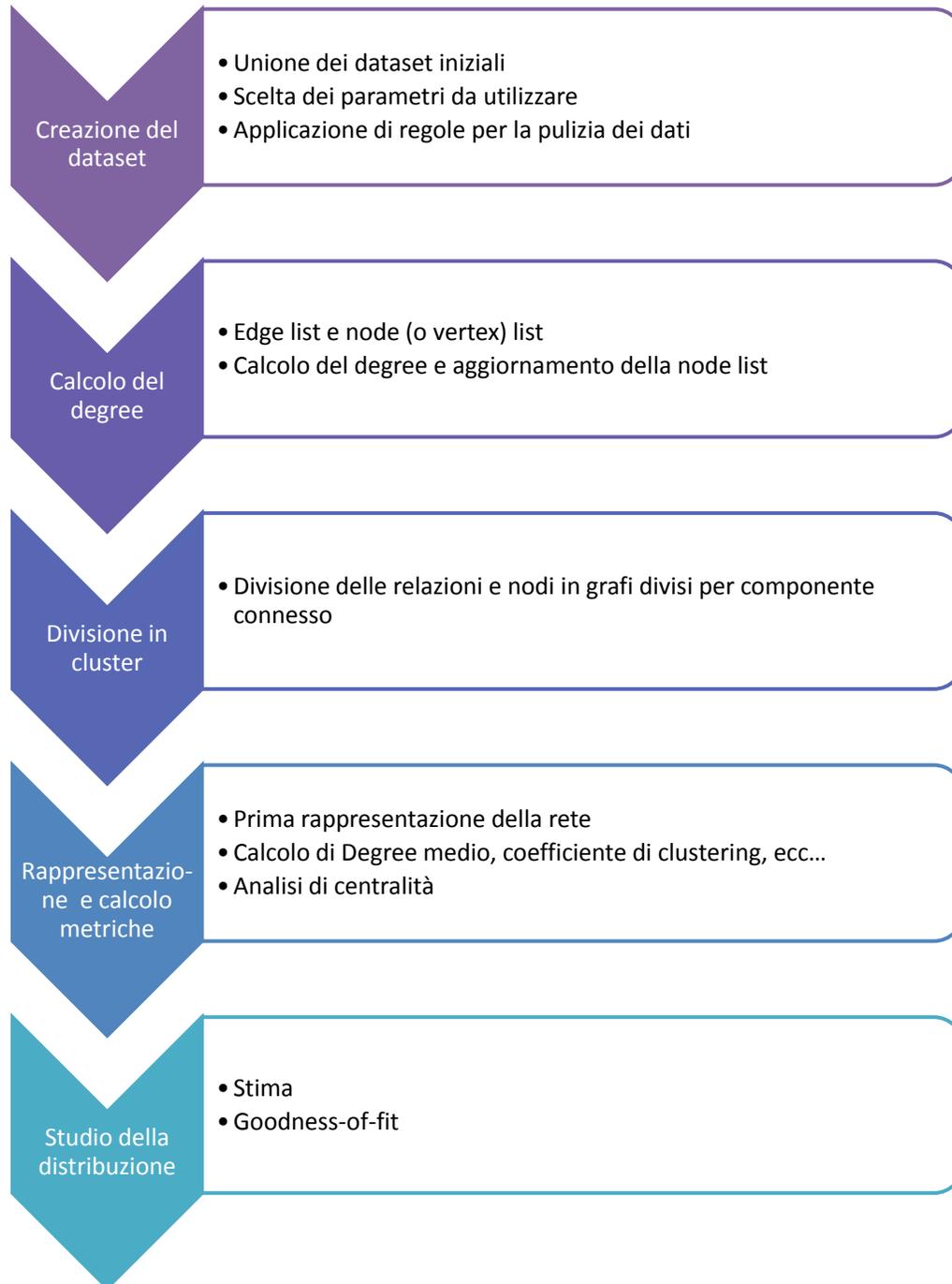


Figura 4.1.4: Workflow

Vertex 1	Vertex 2	%
SOCIETA' PER L'AUTOSTRADA DI ALEMAGNA - SOCIETA' PER AZIONI	AUTOSTRADA TORINO MILANO SPA	5,89
CONSORZIO AUTOSTRADE ITALIANE ENERGIA	AUTOSTRADA TORINO MILANO SPA	5,27
CONSORZIO COCIV	AUTOSTRADA TORINO MILANO SPA	5,00
L A S LINEA AUTO SERVICE S R L	AUTOSTRADA TORINO MILANO SPA	5,00
ROAD LINK LIMITED	AUTOSTRADA TORINO MILANO SPA	5,00
S.S.A.T. SPA	AUTOSTRADA TORINO MILANO SPA	4,00
P.S.T. S.P.A.	AUTOSTRADA TORINO MILANO SPA	3,46
AEROPORTO DELLA PROVINCIA DI PAVIA RIVANAZZANO SRL	AUTOSTRADA TORINO MILANO SPA	2,97
FILSE SPA	AUTOSTRADA TORINO MILANO SPA	1,14
ASSICURAZIONI GENERALI SPA	AUTOSTRADA TORINO MILANO SPA	0,01
DATABASE TECNOLOGIE SPA	EDS ELECTRONIC DATA SYSTEMS ITALIA SRL	100,00

Figura 4.1.5: Esempio di edge-list dataset 2

Il risultato di questa trasformazione saranno due edge-list, una senza percentuale, e una con percentuale.

Calcolo del degree: utilizzando strumenti per l'analisi delle reti sociali, in questa tesi sono stati utilizzati soprattutto NodeXL o Gephi, si può creare, a partire da una lista di relazioni, una lista di vertici. Il dataset è quindi formato. L'operazione successiva sarà quella di trasformare le relazioni in vertici con un determinato *degree*:

$$vertex_1 \leftarrow vertex_2 = vertex_2 \leftarrow vertex_1 = degree_{vertex_1} + 1, degree_{vertex_2} + 1$$

Per ogni relazione $Vertex_1$ e $Vertex_2$ aumenterà il degree di entrambi i vertici, nel caso di grafo non diretto.

Se invece si tratta di grafo diretto il degree sarà diviso in *in-degree* e *out-degree*:

$$vertex_2 \leftarrow vertex_1 = in - degree_{vertex_2} + 1, out - degree_{vertex_1} + 1.$$

Divisione in componenti di rete: il dataset, che rappresenta un grafo, è composto da un numero N di sottografi. Per dividere il grafo in sottografi connessi si utilizzano le tecniche di clustering che permettono di riconoscere ricorsivamente l'appartenenza di un vertice ad un sottografo connesso. Si utilizzerà l'algoritmo presente nel software *Gephi*, che utilizza "Depth-First Search" su grafi lineari [37], il quale introduce le seguenti operazioni, divise a seconda della tipologia di grafo:

- un algoritmo per il riconoscimento di *componenti fortemente connesse* per un grafo diretto;
- un algoritmo per individuare le *componenti bi-connesse* per un grafo non diretto.

Si individuano gruppi basati sulle componenti connesse all'interno del grafo.

Rappresentazione della rete e calcolo delle metriche: per la rappresentazione grafica della rete si utilizzerà Gephi che offre un ottimo sistema di visualizzazione dei grafi. Data una lista di relazioni (edge-list), e una lista di nodi o vertici (vertex-list) è possibile creare una rappresentazione del grafo, come presentato in figura 4.1.7. Ovviamente la prima rappresentazione del grafo non sarà altro che un un'insieme di nodi della stessa dimensione e dello stesso colore. Utilizzando i risultati delle metriche ed effettuando analisi successive sul grafo è possibile arrivare ad una rappresentazione grafica molto più chiara. Le metriche, già descritte nella sezione riguardante la letteratura sulla teoria dei grafi, analizzate nella rete principale e nelle successive divisioni sono le seguenti:

- Degree medio;
- Diametro della rete;
- Numero di Erdős;
- Coefficiente di clustering;
- Centralità di Eigenvector;
- Lunghezza media del cammino;

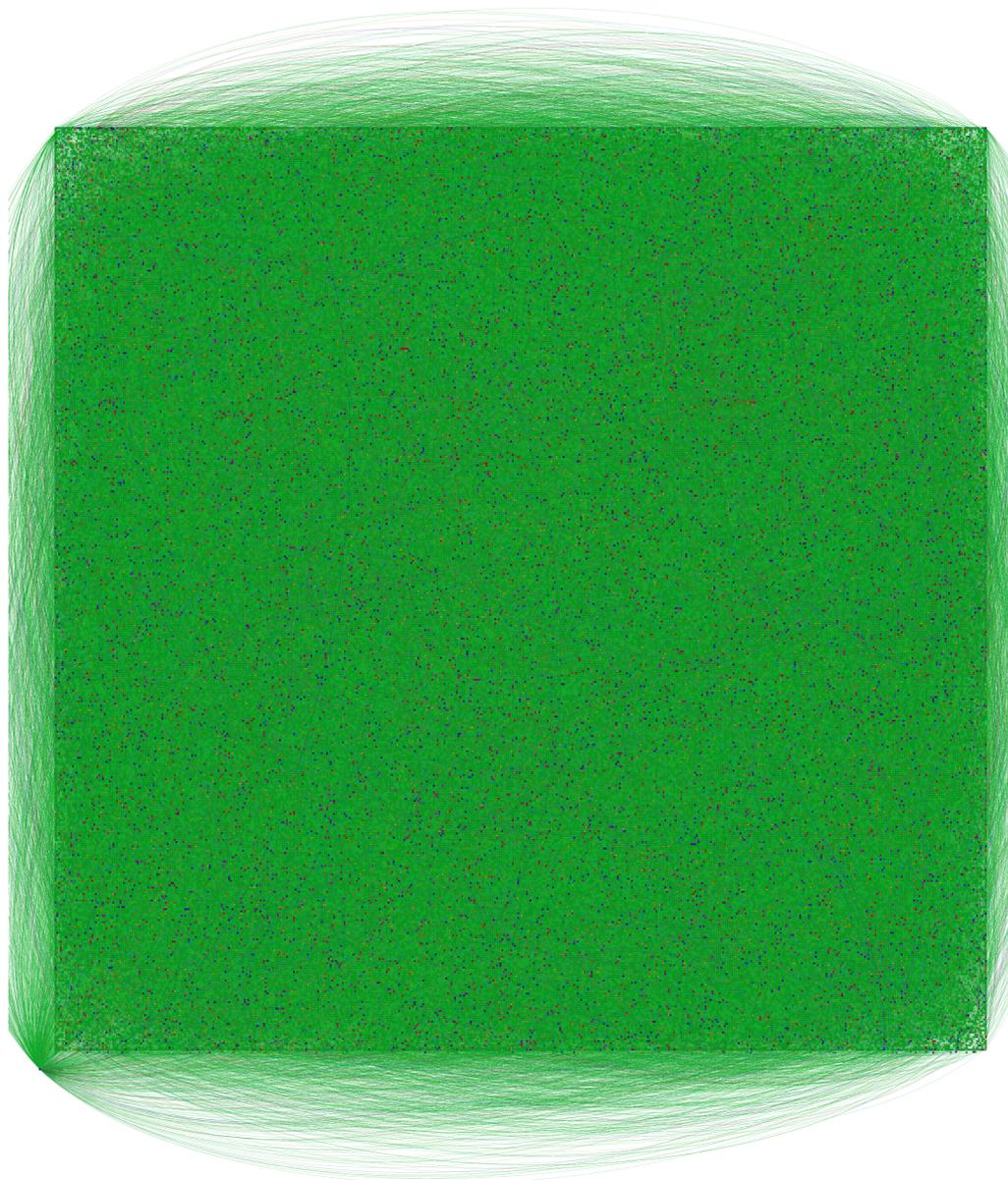


Figura 4.1.7: Rappresentazione iniziale di un grafo. I colori rappresentano i diversi cluster a seconda dei componenti connessi ad essi

Analisi della centralità: tramite le analisi di centralità del grafo, e la successiva dicotomizzazione sul dataset 2, è possibile ordinare il grafo secondo vari criteri:

- degree centrality;
- betweenness centrality;
- eccentricity;
- closeness centrality.

Studio della distribuzione: come ultima analisi verrà studiata, e rappresentata, la distribuzione su ogni livello di trasformazione del grafo per verificare l'appartenenza ad una distribuzione power-law, seguendo i seguenti 2 punti:

1. Stima dei parametri x_{min} e α del modello power-law;
2. Calcolo della *bontà della stima* (*goodness-of-fit*) tra i dati e la power-law. Se il risultante *p-value* sarà maggiore di 0.1 la power-law sarà una ipotesi plausibile per i dati, altrimenti sarà rigettata;

4.2 Il modello

Il risultato di tutte le operazioni presentate nella sezione precedente, è un modello per l'analisi del dataset sotto forma di network. Questo modello è compatibile con Gephi ma facilmente esportabile in altri formati (NodeXL, Ucinet, Pajek,..). La scelta di Gephi deriva soprattutto dalla capacità di svolgere analisi in tempo reale e dalla stabilità e compatibilità con grafi di grandi dimensioni.

Il modello ottenuto contiene tutte le componenti reticolari dei dataset (figura 4.2.1), e quindi tutti i procedimenti mostrati in precedenza, a cui si aggiungono, nella node-list, i dati finanziari presenti nel dataset originale. Riassumendo, i punti chiave contenuti nel nostro modello per ogni nodo, se presenti nel dataset iniziale, avranno:

- nome del vertice;
- degree (o in-degree e out-degree);

- misure di centralità (betweenness, closeness e degree)
- edge-list con percentuale di partecipazione (per il dataset 2);
- componente connessa in cui è presente il vertice;
- dati finanziari:
 - codice ATECO 2007 e 2002 con descrizione del codice;
 - tendenza all'import/export (con percentuali e paese di riferimento);
 - Ricavi delle vendite (Fatturato);
 - utile netto;
 - patrimonio netto;
 - posizione finanziaria netta;
 - capitale sociale;
 - Return on Investment (ROI);
 - Return on Equity (ROE);
 - numero di dipendenti;
 - nome, ruolo, genere, e titolo del manager;
 - ecc..
- dati geografici e contatti;

Questo elaborato è solo un risultato di tutto quello che è possibile ottenere da una mole così imponente di dati, il merito dell'analisi delle reti è quello di, dati una serie di dati difficilmente interpretabili, semplificare cognitivamente l'analisi, la visualizzazione e la ricerca di informazioni.

4.3 Strumenti utilizzati

4.3.1 Node XL

NodeXL (*Network Overview for Discovery and Exploration in Excel*) è un plugin open-source per Microsoft Excel 2007/2010 e permette l'analisi delle

Nodes	Id	Label	Fatturato	In-D...	Out-...	Deg...	Comp...	Stron...	Local ...	Ecce...	Close...	Betwe...
● SIEMENS HEALTHCARE DIAGNOSTICS SRL	SIEMENS HEALTHCARE DIAGNOSTICS SRL	SIEMENS HEALTHCAR	92139,00	3	1	4	0	156249	0	12	0,123	-0
● CAL SRL	CAL SRL	CAL SRL	9213,39	1	0	1	0	53218	0	0	0	-0
● CHEMTURA ITALY SRL	CHEMTURA ITALY SRL	CHEMTURA ITALY SR	92121,25	2	2	4	2206	88188	0	1	1	-0
● EFM SRL	EFM SRL	EFM SRL	9210,66	1	0	1	0	1466	0	0	0	-0
● ROSSI SRL	ROSSI SRL	ROSSI SRL	921,01	4	0	4	0	53199	0	0	0	-0
● SPIGA 1 SRL	SPIGA 1 SRL	SPIGA 1 SRL	9209,85	1	0	1	0	5737	0	0	0	-0
● ORGANIZZAZIONE DI PRODUTTORI ORTOFRUTTICOLI	ORGANIZZAZIONE DI PRODUTTORI ORTOFRUTTICOLI	ORGANIZZAZIONE DI	9209,49	0	1	1	0	161096	0	2	0,571	-0
● BONATTI SERRATURE SPA	BONATTI SERRATURE SPA	BONATTI SERRATURE	9209,44	0	2	2	4904	126004	0	1	1	-0
● SAQUELLA 1856 SRL	SAQUELLA 1856 SRL	SAQUELLA 1856 SRL	9208,82	0	1	1	7104	153792	0	1	1	-0
● SECONDA S.P.A.	SECONDA S.P.A.	SECONDA S.P.A.	9208,79	4	0	4	0	51162	0	0	0	-0
● ROMOR AURELIO & GIOV SPA	ROMOR AURELIO & GIOV SPA	ROMOR AURELIO & G	9208,62	0	1	1	0	111949	0	1	1	-0
● PANDOLFO ALLUMINIO SPA	PANDOLFO ALLUMINIO SPA	PANDOLFO ALLUMINI	92071,12	3	6	9	0	42699	0	1	1	-0
● KORIG ITALY SPA	KORIG ITALY SPA	KORIG ITALY SPA	9207,89	7	0	7	0	65543	0	0	0	-0
● CENTRO ENERGIA FERRARA SPA	CENTRO ENERGIA FERRARA SPA	CENTRO ENERGIA FE	92067,49	5	0	5	0	6232	0,05	0	0	-0
● CANTONI I.T.C. SPA	CANTONI I.T.C. SPA	CANTONI I.T.C. SPA	9206,46	2	39	41	0	53537	0,002	3	0,932	-0
● STANHOME SPA	STANHOME SPA	STANHOME SPA	92051,40	1	1	2	2083	85668	0	1	1	-0
● PITTALUGA SERVIZIO CONTAINERS SPA	PITTALUGA SERVIZIO CONTAINERS SPA	PITTALUGA SERVIZIO	9205,03	1	0	1	0	20890	0	0	0	-0
● REPETTI SRL	REPETTI SRL	REPETTI SRL	9203,42	3	2	5	0	77238	0	1	1	-0
● SOVER SRL	SOVER SRL	SOVER SRL	9203,38	0	7	7	1154	68327	0	1	1	-0
● OMIS SERVICE SRL	OMIS SERVICE SRL	OMIS SERVICE SRL	9203,19	1	0	1	2186	114725	0	0	0	-0

Figura 4.2.1: Il modello

reti sociali. NodeXL sfrutta le funzionalità di calcolo quali l'ordinamento, il filtraggio e la creazione di formule, estendendo il foglio di calcolo con l'analisi e la visualizzazione di rete e incorporando una serie di metriche di base come:

- degree;
- misure di centralità;
- clustering;
- ecc...

insieme alla visualizzazione del grafico.

I dati possono essere inseriti o importati nel modello NodeXL e visualizzati sotto forma di grafico (a seconda della dimensione del grafico).

Come si può notare dalla figura 4.3.1 l'ambiente di lavoro è quello tipico di Microsoft Excel. NodeXL è un template che si aggiunge al foglio di calcolo e presenta diverse funzioni che permettono l'elaborazione dei dati e diverse altre operazioni:

1. *Grafico*: il grafico è, in pratica, l'output desiderato a seconda dei dati inseriti dall'utente. Nella sezione dedicata al grafico, sia nella barra delle applicazioni e sia nella schermata apposita del template, è possibile lanciare l'elaborazione dei dati inseriti mediante il pulsante "Show Graph". La visualizzazione del grafico avviene secondo la scelta dell'algoritmo di visualizzazione scelto:

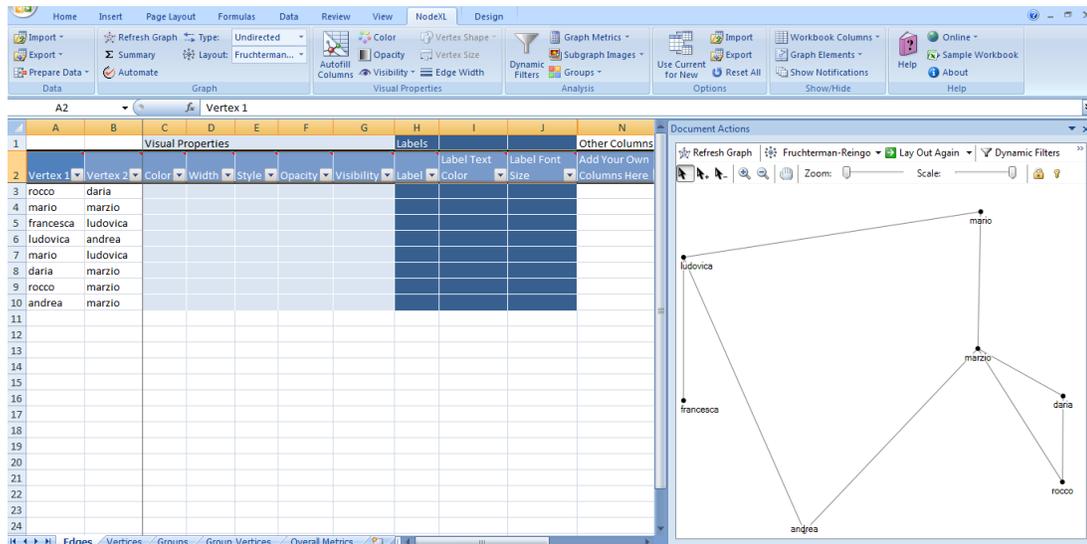


Figura 4.3.1: Esempio di edge-list con NodeXL

Algoritmi di tipo force-directed: presenta algoritmi che realizzano un layout basato sulle forze, questo tipo di algoritmo è quello più utilizzato per realizzare i grafi relazionali. L'idea è quella di considerare la forza tra due nodi qualunque del grafo andando a rendere l'energia totale del sistema la più distribuita possibile.

- Fruchterman-Reingold;
- Harel-Koren Fast Multiscale.

Algoritmi di tipo geometrico:

- circolare;
- spirale;
- a onda;
- a griglia;
- coordinate polari.

Algoritmi a livelli: i nodi sono posti in file orizzontali e i collegamenti si espandono dal livello superiore a quello inferiore, continuando ad andare verso il basso (è una raffigurazione di tipo *gerarchico*):

- Sugiyama.

Algoritmi casuali: i nodi sono assegnati a punti generati casualmente dal programma nello spazio disponibile. Molto utile quando il dataset ha un numero di dati molto alto. E' il metodo di visualizzazione più veloce.

2. *Relazioni ("Edges"):* Ogni riga del foglio presenta la relazione esistente tra due vertici del grafo inseriti nelle prime due colonne del foglio. In testa al foglio di lavoro sono presenti diverse informazioni: oltre alle relazioni tra i nodi, è riportata anche la modalità di rappresentazione di queste relazioni, sia a livello visivo, gestendo le "*Visual Properties*" o le etichette ("*Labels*"), sia a livello matematico pesando le relazioni. Questa pagina contiene quindi tutte le informazioni necessarie a definire e gestire, per mezzo di attributi, ogni relazione presente. NodeXL, mette a disposizione dell'utente diverse opzioni con le quali distinguere e personalizzare i vari legami.
3. *Vertici ("Vertices"):* in questa pagina la prima colonna presenta tutti i vertici presenti nel grafo. Le altre colonne, così come per le relazioni, rappresentano gli attributi della visualizzazione. Anche in questo caso si possono individuare le colonne per la gestione delle proprietà visive, dove è possibile abbinare a ogni vertice una forma o un'immagine che rappresenti il nodo. Per quanto riguarda le etichette è possibile anche decidere in quale posizione dovrà essere inserita l'etichetta rispetto al nodo, inoltre è possibile inserire un "*tooltip*" che permette di fornire le informazioni relative al nodo che si osserva all'interno del grafo. Oltre alle due categorie di attributi, i vertici sono caratterizzati anche da informazioni di "*Graph Metrics*" che riporta i risultati generati dal calcolo degli indici di centralità. I parametri relativi ai nodi sono:
 - degree centrality (distinta in degree per i grafi non diretti e in-degree e out-degree nei grafi diretti);
 - closeness centrality;
 - Eigenvector centrality;
 - coefficiente di clustering;

- PageRank (algoritmo, usato da Google, che assegna un peso numerico a un elemento di un insieme in base al numero di collegamenti ipertestuali che lo stesso ha con gli altri elementi);
 - reciprocated vertex pair ratio.
4. *Gruppi*: Questo strumento è molto efficace per l'analisi dei network. L'Individuazione di sotto gruppi di relazioni all'interno della rete può aiutare a stabilire come le informazioni si distribuiscono all'interno del contesto considerato. Attraverso la clusterizzazione dei vertici presenti nel grafo si possono distinguere gruppi relazionali che si formano all'interno della rete, dei quali precedentemente non si conosceva l'esistenza. Si possono utilizzare 4 metodi di clustering, già presentati nella sezione di letteratura:
- raggruppamento basato su un particolare attributo dei vertici;
 - raggruppamento relativo alle componenti connesse;
 - raggruppamento tramite cluster:
 - Clauset-Newman-Moore (CNM);
 - Wakita-Tsurumi;
 - Girvan-Newman;
5. *Calcolo delle metriche*: La funzione "*Graph Metrics*" consente di calcolare automaticamente tutti i valori che riguardano il grafo. Alcuni indici che appaiono nella schermata di conferma delle operazioni che riguardano la metrica sono riservati a grafi diretti, richiedono quindi in-degree e out-degree. Le metriche individuali calcolate con questa funzione sono utili per monitorare la posizione occupata dai vari nodi del network, individuando i vertici centrali. I risultati, visualizzati nella pagina "*Overall Metrics*" presentano una tabella riportante i risultati:
- numero totale di collegamenti;
 - *self-loops* o numero di vertici che si collegano con sè stessi;
 - percentuale delle *relazioni reciproche*: rapporto tra le relazioni reciproche e quelle totali, sia a livello di diadi, sia a livello di legami;
 - componenti connesse;

- vertici singoli;
- numero massimo di vertici di un componente connesso;
- massima distanza geodetica: distanza maggiore presente nel network tra tutte le coppie di vertici del grafo, rappresenta quindi la distanza tra i vertici più lontani;
- distanza geodetica media;
- densità del grafo: misura esatta dell'interconnessioni dei nodi.

4.3.2 Gephi

Gephi è un software open-source per l'analisi e la visualizzazione delle reti sociali scritto in Java utilizzando la piattaforma NetBeans, inizialmente sviluppato dagli studenti della Université de Technologie de Compiègne in Francia [5].

I punti di forza di Gephi sono soprattutto la visualizzazione in tempo reale del grafo, le performance (nettamente migliori per visualizzare reti di grandi dimensioni), la modularità e lo sviluppo di plug-in da parte di una base molto attiva.

L'interfaccia di Gephi è focalizzata sulla creazione e la visualizzazione di grafi in tempo reale. L'innovazione viene dalla facilità dell'utente di interagire con la rete creata, l'utente può letteralmente "giocare" con la rappresentazione visuale del network. Gephi è un software molto stabile e può caricare reti fino ad 1 milione di nodi e relazioni, in aggiunta, il programma fornisce implementazioni efficienti delle classiche metriche utilizzate nella Social Network Analysis, incluse le misure di centralità, coefficiente di clustering, modularità per l'analisi delle comunità.

Un'immagine di utilizzo di Gephi è presentata in figura 4.3.2, immagine presa dal sito ufficiale del progetto www.gephi.org.

Algoritmi di layout force-directed Gephi fornisce molti layout definiti come force-directed. Questi layout si basano sul posizionamento dei nodi a seconda della posizione degli altri. Approssimativamente, i nodi connessi tendono ad essere più vicini, mentre i nodi disconnessi tendono ad essere allontanati dalle entità connesse all'interno del grafo. I layout possono utilizzare il peso delle relazioni nel calcolo delle forze. I layout principali utilizzati da Gephi sono i seguenti:



Figura 4.3.2: Esempio di utilizzo di Gephi

- *Fruchterman-Reingold*: simula il grafo come un sistema di particelle di massa. I nodi sono le particelle di massa e le relazioni sono le sorgenti tra le particelle. L'algoritmo cerca di minimizzare l'energia del sistema fisico;
- *Force Atlas*: è uno dei layout principali offerti da questo programma. E' utilizzato per visualizzare in modo concreto reti del mondo con proprietà scale-free, o small-world;
- *Force Atlas 2*: versione migliorata di Force Atlas in grado di gestire reti di grandi dimensioni mantenendo una buona leggibilità;
- *OpenOrd*: algoritmo di layout in grado di gestire fino ad 1 milione di nodi, rendendolo ideale per i grafi di grandi dimensioni. L'algoritmo si basa su Fruchterman-Reingold e lavora con un numero di iterazioni fisse;
- *Yifan Hu*: algoritmo molto veloce con una buona qualità di visualizzazione e di partizione di grafi di grandi dimensioni. Combina varie

tecniche utilizzate dai modelli di layout force-directed per ridurre il costo computazionale;

- *Circular*: disegna i nodi ordinati a forma di cerchio in base agli attributi dei nodi. Molto utile per mostrare la distribuzione dei nodi, insieme alle relazioni;
- *Radial Axis*: raggruppa i nodi in gruppi e li distribuisce sull'asse irradiandoli verso l'esterno da un cerchio centrale. I gruppi sono generati utilizzando una metrica o un attributo;
- *Geographical*: il GeoLayout utilizza coordinate geografiche per posizionare i nodi su uno spazio geografico.

Metriche Gephi fornisce metodi statistici classici per lo studio delle. Le metriche per le reti sono statistiche relative all'intero grafo; le metriche sui nodi sono statistiche relative ad ogni nodo così come per le relazioni.

Metriche sulla rete:

- diametro;
- densità;
- modularità;
- calcolo delle componenti connesse e numero;
- coefficiente di clustering;

Metriche sui nodi:

- degree centrality;
- betweenness centrality;
- closeness centrality;
- Eigenvector centrality;
- PageRank;

- HITS (importanza di una pagina web all'interno di una rete, variante della centralità di Eigenvector);

Metriche sulle relazioni:

- lunghezza del cammino medio;

Metriche dinamiche (dynamic metrics): alcune metriche possono essere calcolate nel tempo: il numero dei nodi, il numero delle relazioni, il degree medio e il coefficiente di clustering.

Filtri Una delle caratteristiche chiave di Gephi è l'utilizzo di filtri che si applicano in tempo reale sul grafo. E' possibile dicotomizzare il grafo in base agli attributi dei nodi e delle relazioni, oppure utilizzando i classici attributi di un grafo presentati dalla letteratura:

- ego network;
- range del degree (in e out-degree);
- componente gigante;
- k-core;
- degree range reciproco;
- reti dei vicini;

Capitolo 5

Dati

5.1 Fonti e Limiti

I dati utilizzati in questo elaborato provengono da Mint Italy, soluzione a pagamento della Bureau van Dijk, basata su Aida. Aida è uno strumento software che consente una analisi immediata delle aziende italiane, fornendo tutti i dati economico-finanziari di imprese di grandi, medie e piccole dimensioni, con diverso grado di dettaglio.

Nel nostro caso si utilizzeranno i dati dell'anno 2007 relativi al settore manifatturiero indicato tramite codice *ATECO 2007 (Attività Economiche)*. Si tratta di una classificazione alfa-numerica con diversi gradi di dettaglio: le lettere indicano il macro-settore di attività economica mentre i *digit* (che vanno da due fino a sei cifre) rappresentano, con grado di dettaglio crescente, le articolazioni e le disaggregazioni dei settori stessi.

L'interrogazione del database Mint Italy offre come risultato, in prima istanza, i dati economico-finanziari (aggiornati al 31/12/2007) di tutte le articolazioni del settore manifatturiero (*settore "c"*)[27], compresa la lista di partecipazioni azionarie e azionisti per ogni soggetto. Da questa lista si accede ricorsivamente, per ogni elemento, ad una interrogazione successiva del database.

Il risultato finale è un'insieme di dati in formato tabellare rappresentante tutte le imprese del settore manifatturiero ed i relativi livelli di partecipazione inferiori e superiori. Le partecipazioni e gli azionisti possono essere esterni al settore di riferimento della ricerca.

Per legge (*art. 120 del testo unico sulla finanza*) l'obbligo di comunica-

Relazioni	Vertici
197.030	169.502

Tabella 5.2.1: Numero di relazioni e vertici del dataset 1

zione alla Consob delle partecipazioni possedute ricorre solo se, una persona fisica, società o ente, partecipano direttamente o indirettamente in una società per azioni quotata in misura superiore al 2% del capitale sociale di questa. Per tutte le partecipazioni inferiori non vige l'obbligo di comunicazione alla Consob.

Il dataset verrà utilizzato in seguito per la costruzione di due edgelist a seconda della presenza, o meno, della percentuale di partecipazione azionaria:

1. *dataset 1*: composto da edge-list ($vertice_1, vertice_2$) con relazione $vertice_1$ partecipa in $vertice_2$ quando verrà utilizzato come grafo non diretto.
2. *dataset 2*: composto da edge-list ($vertice_1, vertice_2, \%_{di\ partecipazione}$) con relazione $vertice_1$ partecipa in $vertice_2$ con $\%_{di\ partecipazione}$ quando verrà utilizzato come grafo diretto.

5.2 Dataset 1

Il primo dataset analizzato rappresenterà tutte le relazioni presenti nel dataset iniziale, vengono infatti trascurate le percentuali di partecipazione ottenendo una rete basata sul calcolo del degree.

Il dataset è sotto forma di "edge-list" composto da diadi ($Vertex\ 1, Vertex\ 2$), rappresentato in figura 4.3.1:

- $Vertex\ 1$ = nome del soggetto avente una relazione con $Vertex\ 2$.
- $Vertex\ 2$ = nome del soggetto avente una relazione con $Vertex\ 1$.

Grazie all'ambiente di lavoro NodeXL, l'edge-list in input produce automaticamente una vertex-list rappresentante tutti i vertici unici del dataset. Il numero di relazioni e vertici è presentato in tabella

Vertex	Degree	Degree	Degree
FIAT SPA	558	478	80
SOCIETA' INIZIATIVE AUTOSTRADALI E SERVIZI SPA	353	337	16
SOCIETA' PER AZIONI ISTITUTO FINANZIARIO INDUSTRIALE	347	321	26
FINMECCANICA SPA	344	292	52
TELECOM ITALIA SPA	324	217	107
FINCANTIERI CANTIERI NAVALI ITALIANI SPA	315	313	2
ENI SPA	306	217	89
CONSORZIO NAZIONALE IMBALLAGGI	304	1	303
PIRELLI & C. SPA	275	220	55
AUTOGRILL SPA	269	210	59
SOCIETA' ITALIANA PER CONDOTTE D'ACQUA SPA	263	148	115
PARMALAT SPA	257	136	121
ASTALDI SPA	252	202	50
ITALMOBILIARE SPA	249	230	19
TEJAS SPA	241	235	6
VIKA SRL	240	234	6
PROJEOS SRL	236	230	6
FUTURA SRL	219	89	130
ITALCEMENTI FABBRICHE RIUNITE CEMENTO SPA BERGAMO	218	191	27
PIRELLI & C. REAL ESTATE SPA	192	166	26
VALTUR S.P.A.	182	178	4

Figura 5.2.1: Dataset 1: vertex-list con vertici in ordine di degree

5.2.1 Caratterizzazione

La fase successiva è quella di dare un valore al legame presentato nella edge-list, caratterizzando così la vertex-list tramite il calcolo del degree, come presentato in figura 5.2.1. Il calcolo del degree può essere effettuato sul dataset inteso come grafo diretto e grafo non diretto: nel caso di grafo diretto, nella vertex-list verranno aggiunti i valori di in-degree e out-degree rappresentanti i legami entranti ed uscenti da un nodo; nel caso di grafo non diretto, nella vertex-list verrà aggiunto il solo valore di degree inteso come legame del nodo.

Tramite il calcolo del degree, effettuato sia su grafo diretto che non diretto, si ottiene una definizione dei vertici in base al degree, e, da questa lista è già possibile ottenere delle informazioni riguardanti la distribuzione, soprattutto verso una distribuzione power-law:

- esiste un elevato numero di vertici con degree uguale a 1 e un ristretto numero di vertici con degree elevato (che si potrebbe definire come la "coda lunga" di una power-law);
- dividendo il dataset in gruppi a seconda delle componenti connesse, si nota la presenza di un gruppo principale che contiene la maggior parte

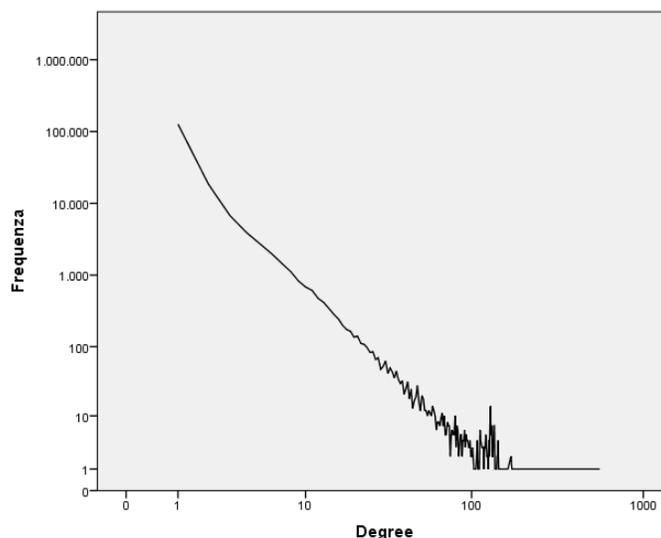


Figura 5.2.2: Dataset 1: distribuzione del degree

dei vertici del dataset, e un grandissimo numero di gruppi di piccole dimensioni.

Secondo [12] identificare un comportamento power-law in sistemi naturali o umani può creare difficoltà. La strategia standard prevede il disegno di un istogramma rappresentante le quantità della distribuzione. Una distribuzione power-law apparirà assimilabile ad una linea retta quando disegnata su doppia scala logaritmica.

Data una lista di vertici con un determinato degree assegnato, a seconda delle relazioni, a tale vertice, è possibile studiare la distribuzioni di frequenze basandosi sul degree, come presentato in figura 5.2.2. La figura 5.2.2 rappresenta la distribuzione del degree in un grafo non diretto, e verrà divisa in in-degree e out-degree nel caso di grafo diretto.

Si può immediatamente notare come il grafico rappresenti una linea retta, fino ad un certo punto, in cui la retta segue una trama "rumorosa" o, come definita da [12], "noisy", questo può indicare anche errori nel campionamento. La rumorosità della distribuzione power-law in queste regioni, indica che ciascuno "scomparto" ha solo pochi campioni in esso, se sono presenti. Così le fluttuazioni frazionarie dei conteggi degli scomparti sono grandi e questo appare come una curva dalla trama rumorosa. Un modo per affrontare questo problema sarebbe quello di, semplicemente, eliminare i dati dalla coda

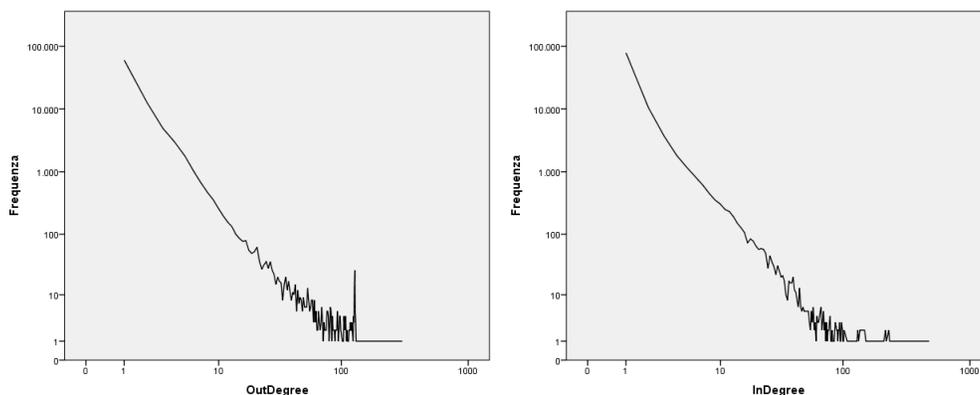


Figura 5.2.3: Dataset 1: distribuzione di frequenze in-degree e out-degree

della curva, ma c'è spesso informazione utile in tali dati e, è possibile che alcune distribuzioni seguano una power-law solo in coda, quindi è necessario mantenere tutti i dati della distribuzione.

Nel caso del dataset sotto forma di grafo diretto, le considerazioni precedenti sulla struttura della rete vengono confermate. Una power-law è l'unica distribuzione che rimane tale per qualsiasi scala osservata, supponendo di avere una distribuzione di probabilità $p(x)$ per una quantità x , e supponendo che segua una proprietà per cui

$$p(bx) = g(b)p(x),$$

per ogni b . Ovvero, effettuando una trasformazione sulla rete, incrementando o diminuendo in parte delle unità, la forma della distribuzione $p(x)$ rimane invariata, eccetto il caso in cui si utilizzi una costante moltiplicativa globale. Questa è definita come proprietà della distribuzione *scale-free*.

Dividendo quindi il degree in in-degree e out-degree in due parti non uguali, ci attendiamo che la distribuzione rimanga una power-law. In figura 6.2.1 vengono mostrate le due distribuzioni di frequenza a confronto, ovviamente ignorando i valori nulli.

Entrambe rispettano una distribuzione power-law, è, tuttavia, troppo presto per affermare che il nostro dataset segua una distribuzione power-law.

In tabella 5.2.2 si analizzano nel dettaglio i dati della distribuzione, calcolando la *verosimiglianza* dei dati al modello distributivo power-law. I dati presentati in tabella mostrano come le distribuzioni non rappresentino delle "pure" power-law, come si possono studiare con modelli randomizzati, ma di-

Grafo	α	x_{min}	$p - value$	power-law ?
Non diretto	2,44	10	0,0010	con cut-off ($x_{min} = 30$)
Diretto	2,30 / 2,27	1/2	0/0	con cut-off ($x_{min} = 6$) / con cut-off ($x_{min} = 13$)

Tabella 5.2.2: Dataset 1: Risultati dell'analisi sulla distribuzione power-law. ”/” indica la divisione tra in-degree/out-degree

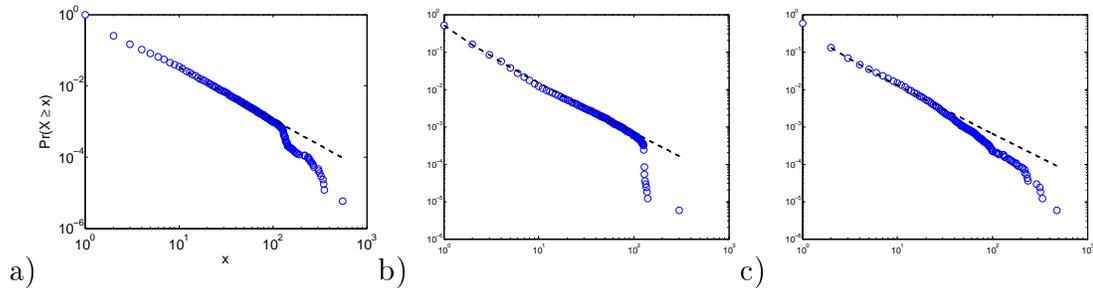


Figura 5.2.4: Dataset 1: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza

stribuzioni power-law con un livello di *cut-off* differente, come si può vedere anche dalla figura 5.2.4 che mostra la funzione di distribuzione cumulativa dei vari casi. La figura mostra la funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza, rispettivamente del: a) *grafo diretto*, b) *grafo non diretto (in-degree)*, c) *grafo non diretto (out-degree)*.

Una power-law con cut-off esponenziale è semplicemente una power-law moltiplicata per una funzione esponenziale $e^{\beta x}$ [12]:

$$f(x) \propto x^\alpha e^{\beta x}.$$

5.2.2 Rappresentazione della rete e analisi delle metriche

Arrivati a questo punto, è necessario utilizzare gli strumenti di analisi delle reti per studiare il comportamento del nostro grafo. La rete rappresentata in figura 4.1.7 è sicuramente di difficile comprensione, un altissimo numero di

Grafo	Componenti connesse	Coefficiente di clustering	Lunghezza media del cammino	Diametro
Diretto	debolmente: 8103 fortemente: 169017	0,018	8,428	28
Non Diretto	8076	0,018	65,481	28

Tabella 5.2.3: Dataset 1: risultati metriche

nodi e legami con la stessa misura. E' necessario quindi analizzare le metriche del grafo, per effettuare successive dicotomizzazioni, che possono permettere una rappresentazione più chiara e significativa.

Vediamo i risultati delle metriche applicati al nostro dataset, calcolati tramite lo strumento Gephi, presentati in tabella 5.2.3.

Come primo commento ai dati riguardanti le metriche e, contrapponendole alla tabella di Mitchell [29], presentata nel capitolo 2, si può notare come differiscano i livelli riguardanti il *coefficiente di clustering* e la *lunghezza media del cammino*. Rappresentando un modello empirico di rete basata sul mondo reale ci aspetteremmo, secondo la tabella di Mitchell, coefficiente di clustering alto e lunghezza media del cammino bassa; i risultati mostrano l'esatto contrario per quanto riguarda il coefficiente di clustering, mentre la lunghezza media dei cammini si può definire come bassa.

Si può obiettare che, essendo un grafo completo e rappresentante un alto numero di reti, questi dati possano essere distorti. Si attenderà quindi l'analisi dei gruppi, nel capitolo successivo, per trarre delle conclusioni in merito al paragone con la tabella di Mitchell.

E' necessario ricordare una delle proprietà fondamentali di una rete scale-free, secondo cui la distribuzione del coefficiente di clustering diminuisce all'aumentare del degree dei nodi e, inoltre, questa distribuzione seguirà, a sua volta, una power-law. Questo implica che i nodi di basso grado appartengono a sottografi molto densi che, a loro volta, sono collegati tra loro tramite degli *hub*. Ad esempio, se si considera una rete sociale in cui i nodi sono persone e i collegamenti sono relazioni di conoscenza tra le persone. E' facile notare come le persone tendono a formare comunità, cioè piccoli gruppi in cui tutti hanno una relazione con qualcuno. Inoltre, i membri di una comu-



Figura 5.2.5: Dataset 1: esempio di grafo

rità hanno anche relazioni con persone al di fuori di quella comunità, alcuni sono collegati ad un numero di comunità (e possono essere definiti come hub responsabili del fenomeno small-world).

Di questa proprietà si può estrapolare, come attinente al nostro contesto, solo la prima parte, per cui si può spiegare un coefficiente di clustering basso dato dai nodi con degree molto alto presenti nel dataset.

E' possibile utilizzare questi parametri per arrivare ad una visualizzazione grafica della rete, che altrimenti risulterebbe difficilmente leggibile. In questo caso (figura 5.2.5) il un grafo rappresenta i nodi con l'out-degree maggiore (quindi col maggior numero di partecipazioni esterne), il colore è su una scala cromatica basata sul degree che va dal giallo (degree più basso) al rosso (degree più alto). Il grafo è stato inizialmente dicotomizzato basandosi su una grandezza del degree x con $x > 90$, questa scelta è dovuta dalla densità di nodi che, altrimenti, non permetterebbe una corretta visualizzazione.

Da questo esempio si può notare come alcuni soggetti abbiamo un valore di degree maggiormente dato dall'in-degree, mentre altri hanno un valore

Relazioni	Vertici
158.565	145.138

Tabella 5.3.1: Numero di relazioni e vertici del dataset 1

basato soprattutto sull'out-degree.

5.3 Dataset 2

Il secondo dataset analizzato rappresenta tutte le relazioni contenenti una percentuale di partecipazione, vengono infatti trascurate le relazioni senza percentuale.

Il dataset è sotto forma di "edge-list" composto da triple ($Vertex\ 1$, $Vertex\ 2$, %), rappresentato in figura 4.3.1:

- $Vertex\ 1$ = nome della società partecipante in $Vertex\ 2$.
- $Vertex\ 2$ = nome della partecipata.
- % = percentuale di partecipazione.

Grazie all'ambiente di lavoro NodeXL, l'edge-list in input produce automaticamente una vertex-list rappresentante tutti i vertici unici del dataset. Il numero di relazioni e vertici è presentato in tabella 5.3.1.

5.3.1 Caratterizzazione

Le relazioni, così come i vertici, sono minori di numero rispetto al dataset contenente le sole relazioni senza percentuale di partecipazione. Si può ipotizzare quindi che la distribuzione non abbia significative variazioni rispetto a quella osservata nel primo dataset, se poi, successivamente, si ipotizza la presenza di proprietà scale-free si deve necessariamente avere la stessa distribuzione rispetto alla rete precedente.

La distribuzione basata su degree, come mostrato in figura 5.3.1, segue la forma della distribuzione riguardante il dataset 1, e si può sempre ipotizzare una distribuzione power-law.

Lo stesso vale per la distribuzione divisa in in-degree e out-degree che rispettano la forma del dataset 1, figura 5.3.2.

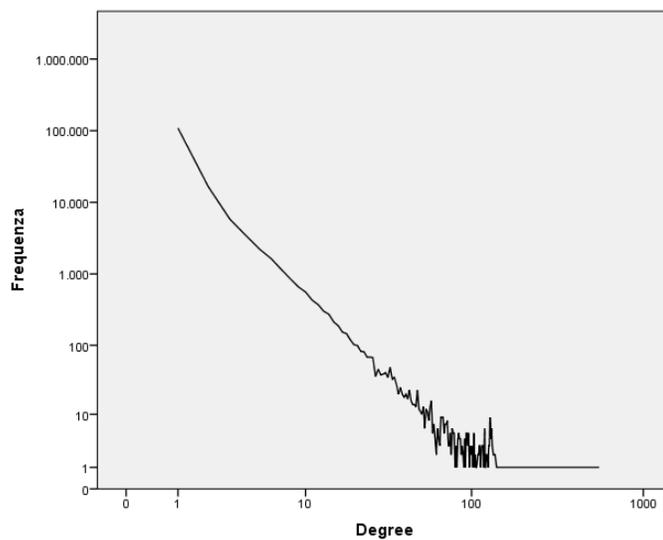


Figura 5.3.1: Dataset 2: distribuzione

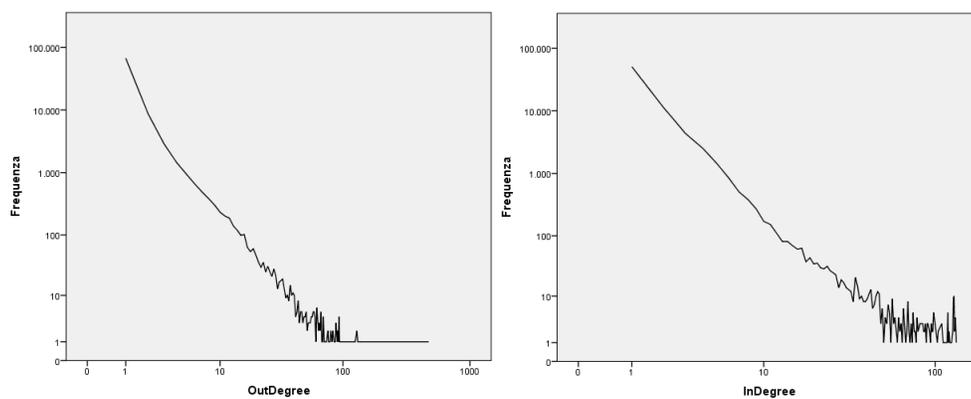
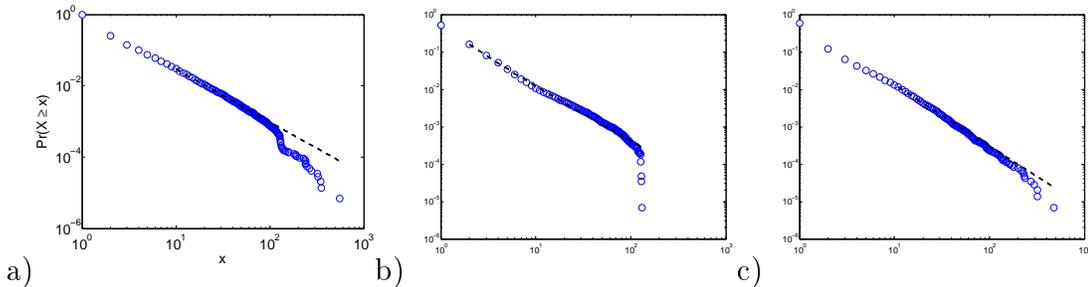


Figura 5.3.2: Dataset 2: distribuzione di frequenze in-degree e out-degree

Grafo	α	x_{min}	$p - value$	power-law ?
Non diretto	2,46	10	0,0020	con cut-off ($x_{min} = 18$)
Diretto	2,43 / 2,61	2/11	0/0,1210	con cut-off ($x_{min} = 9$)/si

Tabella 5.3.2: Dataset 2: risultati dell'analisi sulla distribuzione power-law

Figura 5.3.3: Dataset 2: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza

I due dataset hanno, quindi, le stesse caratteristiche di distribuzione, entrambi presentano inizialmente una linea retta e, avvicinandosi alla coda, aumenta il "rumore" e, diventano "noisy". La figura mostra la funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza, rispettivamente del: a) grafo diretto, b) grafo non diretto (*in-degree*), c) grafo non diretto (*out-degree*).

5.3.2 Rappresentazione della rete e analisi delle metriche

Vediamo ora i risultati delle metriche applicati al nostro dataset, presentati in tabella 5.3.3. Anche in questo caso si nota come le metriche siano in disaccordo con l'idea di power-law scale-free proposta da Mitchell, soprattutto per quanto riguarda il coefficiente di clustering basso.

Grafo	Componenti connesse	Coefficiente di clustering	Lunghezza media del cammino	Diametro
Diretto	debolmente: 9651 fortemente: 144908	0,021	7,553	25
Non Diretto	9674	0,042	37,8	30

Tabella 5.3.3: Dataset 2: risultati metriche

5.4 Dicotomizzazione basata sulla partecipazione azionaria

Partendo dal dataset iniziale, contenente vertici, relazioni e percentuali di partecipazione, è possibile dicotomizzare il dataset basandosi sulla percentuale di partecipazione azionaria. Per fare ciò è necessario innanzitutto stabilire un valore di dicotomizzazione per cui il valore minimo $\%_{min} \geq 0$ diventa $\% = \%_{min}$.

L'ipotesi iniziale si potrebbe soffermare sulla logica idea che "un soggetto A possiede una Società per Azioni B se la sua percentuale di partecipazione è $> 50\%$ ", tale ipotesi può essere smentita da alcuni esempi:

Esempio 1: L'Eni, ex Ente Nazionale Idrocarburi (ENI), è un'azienda multinazionale creata dallo Stato Italiano come ente pubblico. Eni S.p.A è partecipata al 30,31% dal Ministero dell'Economia e delle Finanze e al 9,99% da Cassa Depositi e Prestiti S.p.A. (il cui 70% è controllato dal Ministero dell'Economia e delle Finanze). Il Ministero dell'Economia e delle Finanze è il vero proprietario di Eni S.p.A. anche se possiede meno del 50% delle azioni, in realtà il 40,3 %.

Esempio 2: Fiat S.p.A. (Fiat Società per Azioni) è un gruppo industriale italiano. Fiat S.p.A. è partecipata al 33,65% dalla finanziaria Giovanni Agnelli e C. Sapa che è il vero proprietario della società anche se ne possiede il 33,65%

Relazioni	Vertici
85.349	92.460

Tabella 5.4.1: Numero di relazioni e vertici del dataset 2 dicotomizzato

Esempio 3: Telecom Italia è la principale azienda italiana di telecomunicazioni. La società è partecipata al 23,6% da Telco che ne detiene la maggioranza relativa.

Sono necessarie dunque due operazioni di dicotomizzazione:

1. Dicotomizzazione basata su $\%_{min} = 20\%$.
2. Dicotomizzazione basata su $degree > 0$, sul ricalcolo del degree nel nuovo dataset dicotomizzato (molti vertici possono avere $degree = 0$ dalla prima operazione).

Il nuovo grafo avrà la forma presentata in tabella 5.4.1.

Notevolmente ridotta rispetto al dataset iniziale. Le relazioni rappresentano il 53,44% del dataset iniziale, i vertici il 63,7%.

Vediamo ora come cambia la distribuzione di frequenza del degree, su grafo non diretto, dopo l'applicazione dell'operazione di dicotomizzazione (figura 5.4.1).

Come si può notare, il grafico mostra come la distribuzione non sia variata nonostante la dicotomizzazione. Vedendo come cambia la distribuzione rappresentando graficamente in-degree e out-degree, e quindi il grafo diretto, si nota come, anche in questo caso, la distribuzione rappresenti sempre una power-law (figura 5.4.2).

In questo caso si può notare come nel grafico rappresentante l'in-degree, dopo la dicotomizzazione, sparisce l'effetto "noise" nella coda della distribuzione, e sembrerebbe rappresentare sempre una distribuzione power-law. Per analizzare direttamente se questi dati seguono una distribuzione power-law si calcola la *log-likelihood* applicata al nostro dataset dicotomizzato, come si mostra in tabella 5.4.2.

Dalla tabella si nota il primo caso in cui la distribuzione non segue una power-law. Analizzando il caso si nota che l'in-degree massimo, su partecipazione dicotomizzata, è di 6 indicante qualche errore a livello dei dati all'interno del dataset. Dall'analisi è evidente che per quanto riguarda le relazioni entranti, si abbiano molti casi in cui l'in-degree sia basso (da 0 a 3),

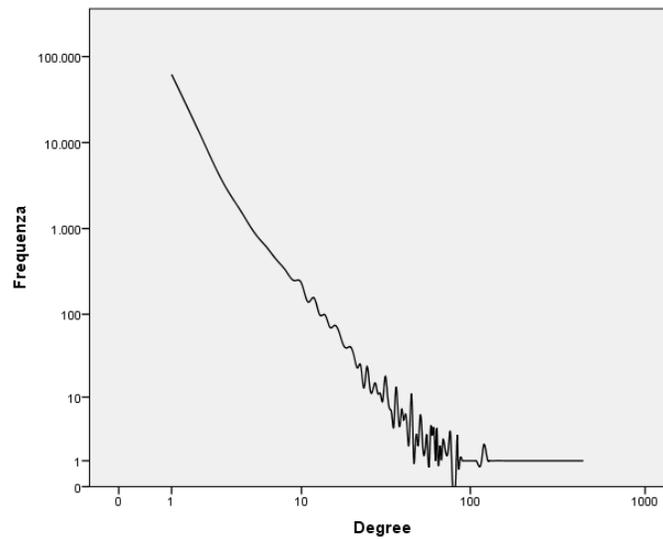


Figura 5.4.1: Dataset dicotomizzato: distribuzione degree su dataset dicotomizzato

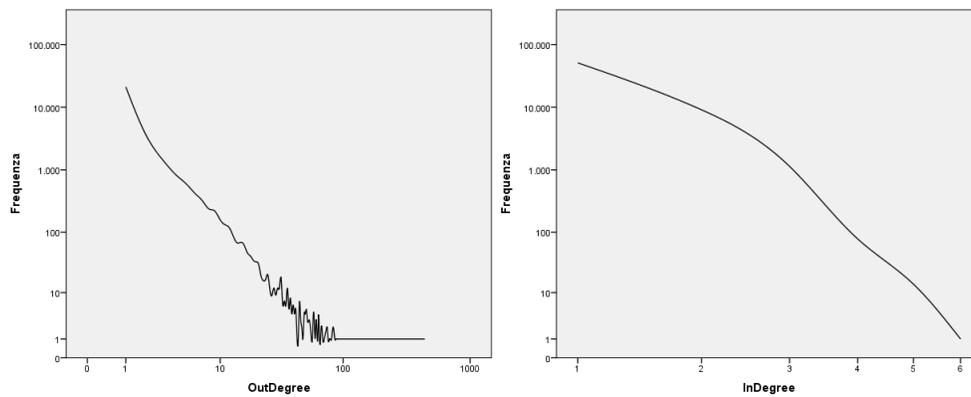


Figura 5.4.2: Dataset dicotomizzato: distribuzione di frequenze in-degree e out-degree dicotomizzati

Grafo	α	x_{min}	$p - value$	power-law ?
Non diretto	2,50	1	0,0020	con cut-off ($x_{min} = 2$)
Diretto	2,94 / 2,63	1/11	0/0,0280	no/con cut-off ($x_{min} = 15$)

Tabella 5.4.2: Dataset dicotomizzato: risultati dell'analisi sulla distribuzione power-law.

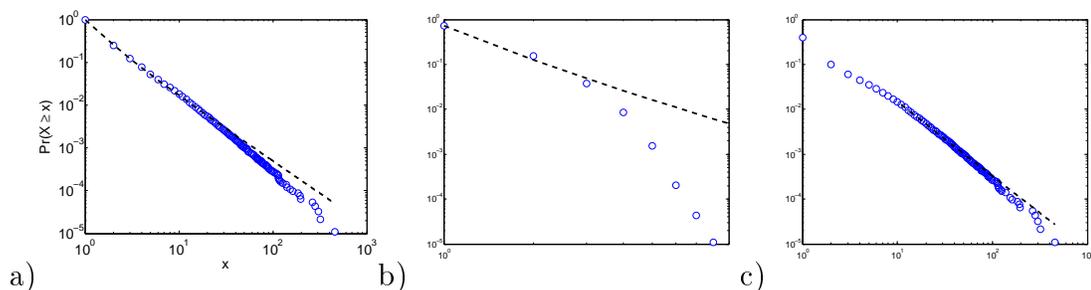


Figura 5.4.3: Dataset dicotomizzato: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza.

e in pochi casi sia alto. La figura 5.4.3 mostra la funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza, rispettivamente del: a) *grafo diretto*, b) *grafo non diretto (in-degree)*, c) *grafo non diretto (out-degree)*.

5.4.1 Rappresentazione della rete e analisi delle metriche

Vediamo l'analisi delle metriche dopo le operazioni di dicotomizzazione (tabella 5.4.3). I valori divergono di molto per quanto riguarda la lunghezza media del cammino che si dimezza di un valore superiore al 50%. Ovviamente le operazioni di dicotomizzazione ricadono sulle relazioni e, come si può vedere dal numero di componenti connesse, sui cluster della rete a dimostrazione della importante funzione di intermediazione svolta dai nodi, con partecipazioni inferiori del 20%, presenti all'interno della rete, e dall'assenza di un numero elevato di relazioni ridondanti tra i nodi.

In figura 5.4.4 si rappresenta la rete dicotomizzata che, per motivi di visibilità, viene ulteriormente dicotomizzata basandosi su un valore di betweenness che permette di escludere un elevato numero di nodi marginali.

La grandezza dei nodi è basata sul degree (dal nodo più piccolo dal degree più basso, al nodo più grande dal degree più alto). Il colore, invece, si basa sulla misura di betweenness e su una scala cromatica dal verde chiaro (nodo con valore meno alto betweenness centrality) al viola scuro (nodi con valore più alto di betweenness centrality).

Grafo	Componenti connesse	Coefficiente di clustering	Lunghezza media del cammino	Diametro
Diretto	debolmente: 13181 fortemente: 81614	0,02	1,61	8
Non Diretto	13201	0,039	15,895	55

Tabella 5.4.3: Dataset 2: risultati metriche con dicotomizzazione

E' facile notare come, nonostante i nodi con degree più alto abbiano un valore alto di betweenness centrality, il degree non sia necessariamente associabile ad un valore di betweenness centrality, come si vedrà in seguito. Si può quindi arrivare alla conclusione che non è importante il *numero* di connessioni ma la *qualità* delle stesse, si veda il caso della società "A2A S.p.A.", il cui degree è abbastanza basso, nei confronti del grafo e, nonostante ciò, possiede un elevato grado di betweenness che indica la sua funzione di "ponte" nei legami della rete.

Capitolo 6

Analisi su componenti connesse

In questo capitolo verranno scelti, per ogni dataset, i 10 gruppi divisi per componenti connesse, sulla base del numero decrescente di nodi contenuti al proprio interno. Questa operazione viene effettuata per vari motivi:

- Studiare la distribuzione all'interno dei vari gruppi e verificare se rispetta i valori di coefficiente di clustering e di lunghezza del cammino medio definiti da Mitchell [29].
- Verificare fino a quale punto i dati rispetteranno una distribuzione power-law.

Per quanto riguarda il dataset 1 (senza percentuale) verranno studiate la distribuzione e le metriche associate al gruppo. Per il dataset 2, in aggiunta a queste analisi, verrà effettuata una successiva dicotomizzazione basata su $\%_{min}$.

6.1 Considerazioni sui gruppi

Vediamo, in tabella, le statistiche riguardanti la divisione dei due dataset in 10 gruppi, rappresentanti il maggior numero di nodi al loro interno. Nelle seguenti tabelle vengono riassunti i numeri riguardanti i nodi e le relazioni delle 10 componenti del dataset. I numeri tra parentesi indicano la percentuale di nodi o relazioni sul dataset totale.

E' facile notare come il primo gruppo sia molto più significativo rispetto agli altri. Da questa tabella si può anche ipotizzare la presenza di una "legge

Numero	Nodi	Relazioni
1	131389 (77,52%)	165984 (84,24%)
2	76 (0,04%)	75 (0,04%)
3	68 (0,04%)	67 (0,03%)
4	58 (0,03%)	60 (0,03%)
5	54 (0,03%)	53 (0,03%)
6	49 (0,03%)	48 (0,02%)
7	47 (0,03%)	46 (0,02%)
8	44 (0,03%)	43 (0,02%)
9	44 (0,03%)	43 (0,02%)
10	43 (0,03%)	49 (0,02%)

Tabella 6.1.1: Suddivisione del dataset 1 in gruppi

Numero	Nodi	Relazioni
1	94685 (65,24%)	117388 (73,5%)
2	96 (0,07%)	103 (0,06%)
3	83 (0,06%)	82 (0,05%)
4	76 (0,05%)	75 (0,05%)
5	70 (0,05%)	72 (0,05%)
6	68 (0,05%)	67 (0,04%)
7	67 (0,05%)	73 (0,05%)
8	67 (0,05%)	69 (0,04%)
9	65 (0,04%)	73 (0,05%)
10	59 (0,04%)	58 (0,04%)

Tabella 6.1.2: Suddivisione del dataset 2 in gruppi

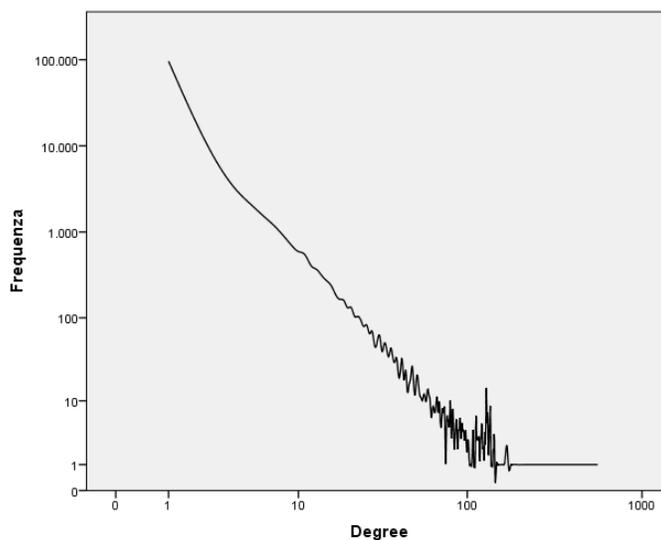


Figura 6.2.1: Dataset 1: distribuzione del degree nel gruppo principale

80/20” o principio di Pareto, secondo la quale “la maggior parte degli effetti è dovuta ad un numero ristretto di cause”, contestualizzandola nel nostro caso si può dire che *“il 20% dei gruppi possiede l’80% dei nodi del grafo”*.

6.2 Dataset 1

6.2.1 Componente gigante

Vediamo in figura 6.2.1 la distribuzione del degree riguardante il gruppo principale del nostro dataset.

Il gruppo principale del dataset contiene l’80% dei nodi totali ed è legittimo quindi aspettarsi, anche come in-degree e out-degree (rappresentati in figura 6.2.2), che la distribuzione non subisca modifiche notevoli rispetto alla distribuzione del dataset.

Anche per quanto riguarda l’analisi delle metriche, non si notano divergenze rilevanti rispetto al dataset totale, come si può vedere in tabella 6.2.1.

Graficamente, su scala log-log, si può intravedere una funzione rappresentante inizialmente una linea retta, che caratterizza una distribuzione power-law, seguita da varie distorsioni che rendono la coda della distribuzione “noi-

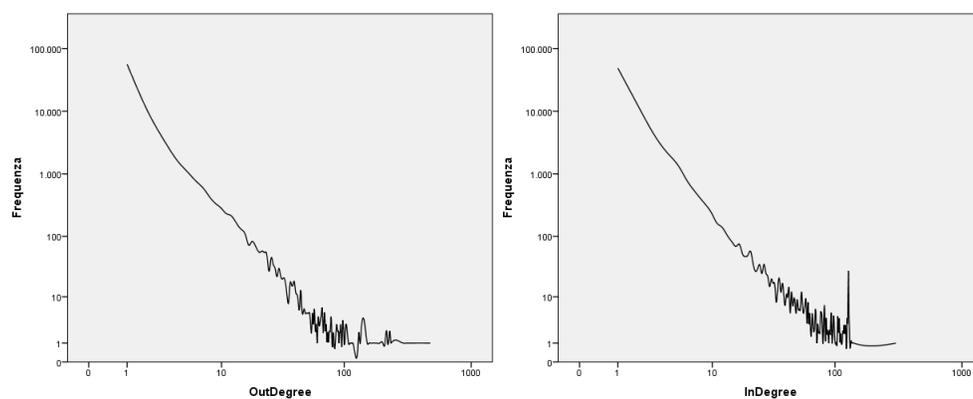


Figura 6.2.2: Dataset 1: distribuzione di in e out-degree nel gruppo principale

Degree medio	Coefficiente di clustering	Lunghezza media del cammino	Diametro
1,18	0,022	8,449	28

Tabella 6.2.1: Dataset 1: analisi delle metriche riguardanti il gruppo principale

Grafo	α	$p - value$	x_{min}	power-law?
Non diretto	2,30	0,0697	10	con cut-off ($x_{min} = 11$)
Diretto	2,30 / 2,21	0,5410 / 0,0190	1 / 1	si/con cut-off ($x_{min} = 2$)

Tabella 6.2.2: Dataset 1: Risultati dell'analisi sulla distribuzione power-law, componente gigante. ”/” indica la divisione tra in-degree/out-degree

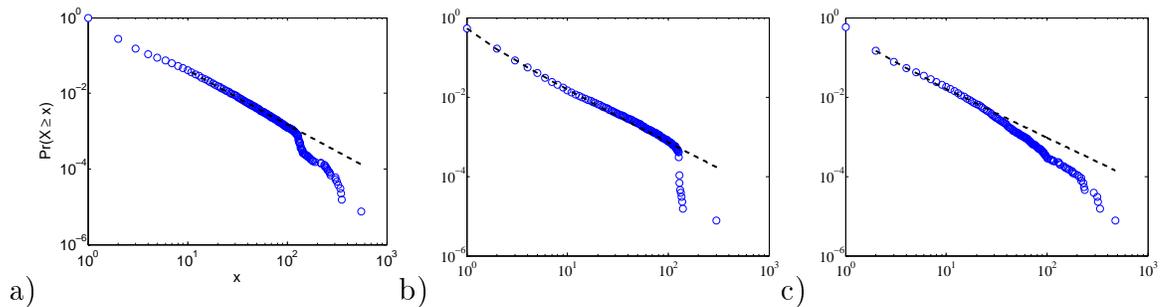


Figura 6.2.3: Dataset 1: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza. a) grafo diretto; b) grafo non diretto (in-degree); c) grafo non diretto (out-degree)

sy”. Si può fare un’ulteriore analisi della distribuzione utilizzando i metodi proposti da [12] per l’analisi di una distribuzione power-law su dati empirici.

Vediamo i risultati delle nostre analisi sul gruppo principale (tabella 6.2.2).

I $p - values$, indicanti il livello di significatività osservato, rappresentano valori al di sopra della nostra soglia di significatività del test, e cioè 0.1. Per questo motivo l’ipotesi di distribuzione power-law non può essere rigettata. E’ necessario notare, come si può vedere in figura 6.2.3, rappresentante la funzione di distribuzione cumulativa $P(x)$ e il massimo insieme di valori di verosimiglianza per il dataset, che la distribuzione, in questo caso, non indica una power-law ”pura”, bensì sempre una power-law ma con un livello di ”cut-off”.

Queste distribuzioni si possono mettere a confronto con un particolare esempio di distribuzione power-law con cut-off presente in letteratura, come si può vedere in figura 6.2.4. L’esempio mostra il numero di visite a siti web ricevute da parte di clienti del servizio American Online Internet in un giorno singolo. Il comportamento, come nei nostri tre casi, indica una

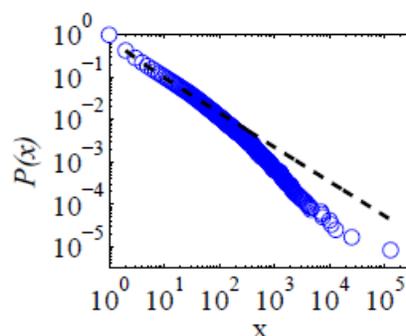


Figura 6.2.4: Esempio di power-law con cut-off

funzione che segue una distribuzione power-law fino ad un certo valore, in cui la distribuzione smette di seguire la linea retta.

6.2.2 Altre componenti

Le altre componenti, all'interno dei 10 gruppi col maggior numero di componenti connesse, rappresentano solamente lo 0,29% dei nodi e il 0,17% delle relazioni, figura 6.2.5.

Dal punto di vista topologico queste reti si possono definire come "nodi subordinati ad un basso numero di hub". Come vedremo in seguito, le reti si presenteranno come un insieme di nodi con un degree (in e out) basso, rappresentanti la maggioranza dei nodi, e un piccolo numero di nodi, a volte anche solo 1, con un degree molto alto (che si può definire come "centro di riferimento" del grafo).

Dal punto di vista economico, utilizzando il modello della "Solar System Enterprise" offerto da Butera [10], una *rete di imprese*, a seconda dell'esistenza o meno di un centro di riferimento può essere definita come:

1. *rete di imprese dotate di un centro di riferimento* attorno al quale orbitano le imprese della rete. L'istituto centrale diventa il coordinatore delle relazioni e di solito il pianificatore e controllore della strategia di rete;
2. *reti di imprese dotate di centri di riferimento multipli* in cui il sistema si muove attorno a più entità di riferimento secondo le relazioni d'influenza;

Componente	α	x_{min}	$p - value$	power-law ?
2	3,50	1	0	con cut-off ($x_{min} = 2$)
3	3,44	1	0,0040	con cut-off ($x_{min} = 2$)
4	2,71	1	0,0040	con cut-off ($x_{min} = 2$)
5	3,08	1	0,0130	con cut-off ($x_{min} = 2$)
6	3,50	1	0,0167	con cut-off ($x_{min} = 2$)
7	3,50	1	0,0214	con cut-off ($x_{min} = 2$)
8	3,15	1	0,0080	con cut-off ($x_{min} = 2$)
9	3,35	1	0,0779	con cut-off ($x_{min} = 2$)
10	2,15	1	0,6970	si

Tabella 6.2.3: Dataset 1: analisi della distribuzione di in-degree e out-degree nelle altre componenti di rete

3. *rete di imprese senza centro* di cui sono esempi i sistemi a base territoriale come i distretti e le fiere.

Nel nostro contesto si parla, naturalmente, di *reti proprietarie*, in cui il collegamento tra le imprese del network è garantito dal possesso di azioni. A questa categoria appartengono le holding, le joint venture corporation, i konzerne (società consociate) tipici delle realtà tedesche e anglo-americane e tutte quelle strutture assimilabili a quelle dei gruppi d'impresa.

La distribuzione del degree (grafo non diretto) mostra che la somma di in e out-degree porta sempre ad una distribuzione power-law, graficamente mostrata in figura 6.2.5. Lo studio di come cambia la distribuzione dopo la separazione del degree in in-degree e out-degree può portare a risultati interessanti, mostrati in figura 6.2.6.

Si nota immediatamente come, alcune rappresentazioni siano molto discordi dal tipo di funzione incontrata fino a questo momento. Di seguito si analizzeranno singolarmente le componenti che graficamente risultano discordanti rispetto ad una distribuzione power-law.

Gruppo 2: Il gruppo rappresenta la tipica forma di gruppo multinazionale, a cascata, in cui i nodi principali "*Rhodia Italia S.p.A.*" e "*Rhodia Participations*" sono nodi puramente finanziari e di controllo. L'in-degree è rappresentato da soli valori 0 e 1, questo indica che il gruppo è formato da un nodo che "*comanda*" tutti i sottonodi, ed è partecipato da un solo nodo, che è il nodo con il maggiore potere all'interno della rete.

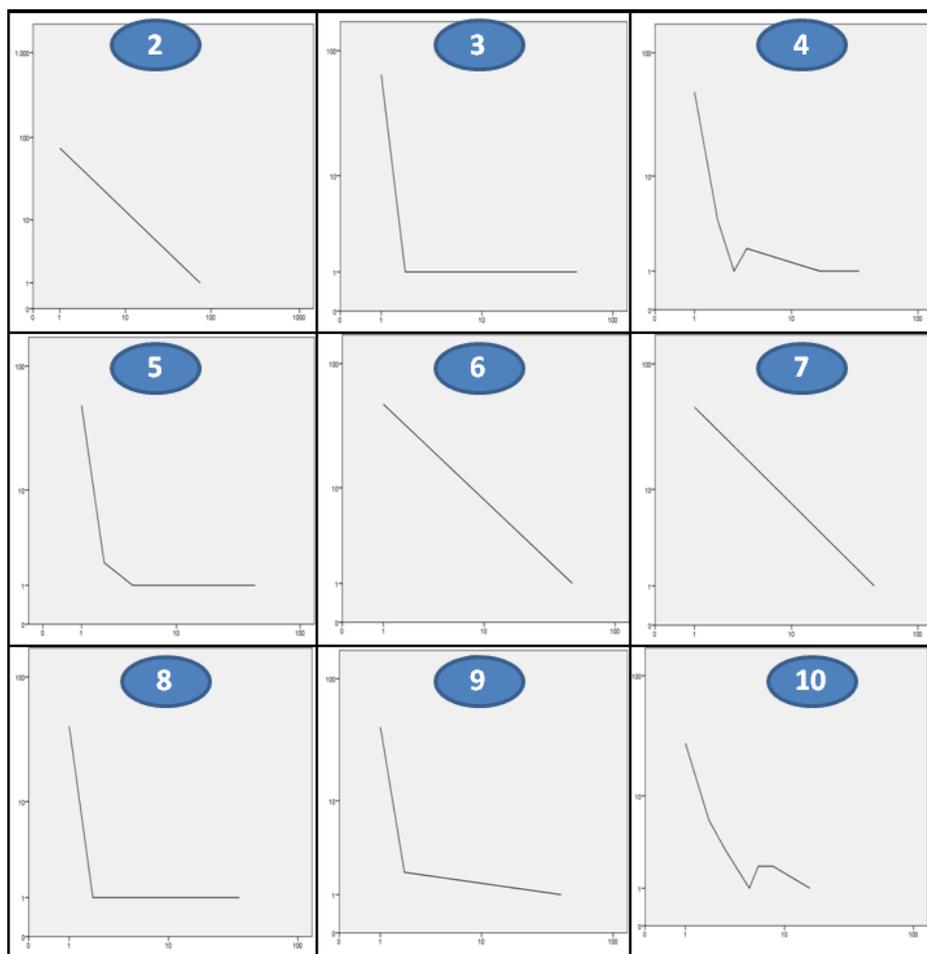


Figura 6.2.5: Dataset 1: distribuzione del degree nelle altre componenti di rete



Figura 6.2.6: Dataset 1: distribuzione di in e out-degree

Componente	α	x_{min}	$p - value$	power-law ?
2	1,50/1,50	0/1	0/0,2500	no/si
3	1,50/3,50	1/1	0,2222/0,0130	si/con cut-off ($x_{min} = 2$)
4	3,10/2,09	1/1	0,2440/0,3974	si/con cut-off ($x_{min} = 2$)
5	1,64/3,50	1/1	0,9048/0,0470	si/no
6	3,50/2,23	1/1	0,0140/0,0870	con cut-off ($x_{min} = 2$)/con cut-off ($x_{min} = 2$)
7	3,06/3,40	1/1	0,0110/0,0065	con cut-off ($x_{min} = 2$)/con cut-off ($x_{min} = 2$)
8	2,25/3,50	1/1	0,0278/0,1739	con cut-off ($x_{min} = 2$)/si
9	1,50/1,51	0/1	0/0,5000	no/si
10	3,18/2,71	5/1	0,7500/0,0810	si/con cut-off ($x_{min} = 2$)

Tabella 6.2.4: Dataset 1: analisi della distribuzione dell'in e out-degree

Dal punto di vista economico, infatti, il nodo *"Rhodia Participations"*, nodo finanziario della holding, ha svariate partecipazioni in tutte le succursali del gruppo (nel nostro caso è visibile solo il nodo *"Rhodia Italia S.p.A."*). Vedendo l'immagine 6.2.7 si potrebbe, in un primo momento, classificare la rete di imprese come dotata di un solo centro di riferimento attorno al quale orbitano le imprese della rete. In realtà non è così, sono presenti più centri di riferimento, uno che controlla la succursale italiana, e l'altro che controlla il sottogruppo italiano; siamo quindi nel secondo caso delle reti di imprese.

Passando allo studio della distribuzione, è facile arrivare alla conclusione che la distribuzione non rappresenti una power-law, come mostrato nella tabella riassuntiva.

Gruppo 5: Il gruppo rappresenta un'insieme di nodi che partecipano all'interno di un soggetto (che può essere considerato come una cooperativa, consorzio, o gruppo d'acquisto). E' inoltre presente un piccolo gruppo che indica le partecipazioni singole di un soggetto partecipante anche nel gruppo principale. Dal punto di vista economico questo gruppo può essere definito del primo tipo, con un solo centro di riferimento e contestualizzato a livello territoriale (regione Campania).

La distribuzione dell'in-degree è formata da 3 linee rette che rappresentano la congiunzione tra il "consorzio" (coda della distribuzione) con gli altri hub che portano ai nodi singoli. Questa distribuzione può inizialmente sembrare come divergente da una power-law ma, calcolandone i valori, si nota come l'ipotesi di distribuzione power-law non sia da escludere.

Gruppo 9: la rete, rappresentata in figura 6.2.9, rappresenta la stessa forma del gruppo 2 (*"Rhodia"*). Entrambe le reti rappresentano una multinazionale, e si può quindi arrivare alla generalizzazione che le multinazionali seguano la forma descritta in questi due casi. Rispetto al gruppo due è presente una partecipazione aggiuntiva della holding (*"Demag Cranes & Components GMBH"*). Dal punto di vista economico la situazione è la stessa, sono presenti due centri di riferimento, quindi nel secondo caso delle reti di imprese.

La funzione di distribuzione segue le stesse caratteristiche del gruppo 2, compresi i valori di riferimento. La distribuzione dell'in-degree non segue una power-law.

Degree medio	Coefficiente di clustering	Lunghezza media del cammino	Diametro
1,10	0,021	7,533	25

Tabella 6.3.1: Dataset 2: analisi delle metriche riguardanti il gruppo principale

Grafo	α	$p - value$	x_{min}	power-law?
Non diretto	2,33	0,1090	1	con cut-off ($x_{min} = 2$)
Diretto	2,32 / 2,79	0,5790 / 0,0450	1 / 10	si/con cut-off ($x_{min} = 25$)

Tabella 6.3.2: Dataset 2: Risultati dell'analisi sulla distribuzione power-law, componente gigante.

6.3 Dataset 2

6.3.1 Componente gigante

Vediamo in figura 6.3.1 la distribuzione del degree riguardante il gruppo principale del nostro dataset.

Il gruppo principale del dataset contiene il 65% dei nodi totali, inferiore rispetto alla percentuale del dataset 1. Questa diminuzione di presenza è data dal numero di nodi e relazioni inferiori rispetto al dataset 1. Rappresentando più della metà dei nodi e delle relazioni ci si attende che anche la distribuzione dell'in-degree e out-degree (rappresentati in figura 6.3.2) abbiano al forma di una power law, e che quindi questa prima divisione non porti a grandi differenze, così come per il dataset 1.

Rispetto al dataset 1 le metriche subiscono una diminuzione, soprattutto nei valori riguardanti la lunghezza media del cammino e il diametro, come si può vedere in tabella 6.3.1.

Anche in questo caso, si può intravedere una funzione rappresentante inizialmente una linea retta, così come nel dataset 1, che caratterizza una distribuzione power-law. Seguono una serie di distorsioni nella coda della distribuzione che la rendono "noisy".

Vediamo i risultati delle nostre analisi sul gruppo principale:

Anche in questo caso si nota, come si può vedere in figura 6.3.3, rappresentante la funzione di distribuzione cumulativa $P(x)$ e il massimo insieme

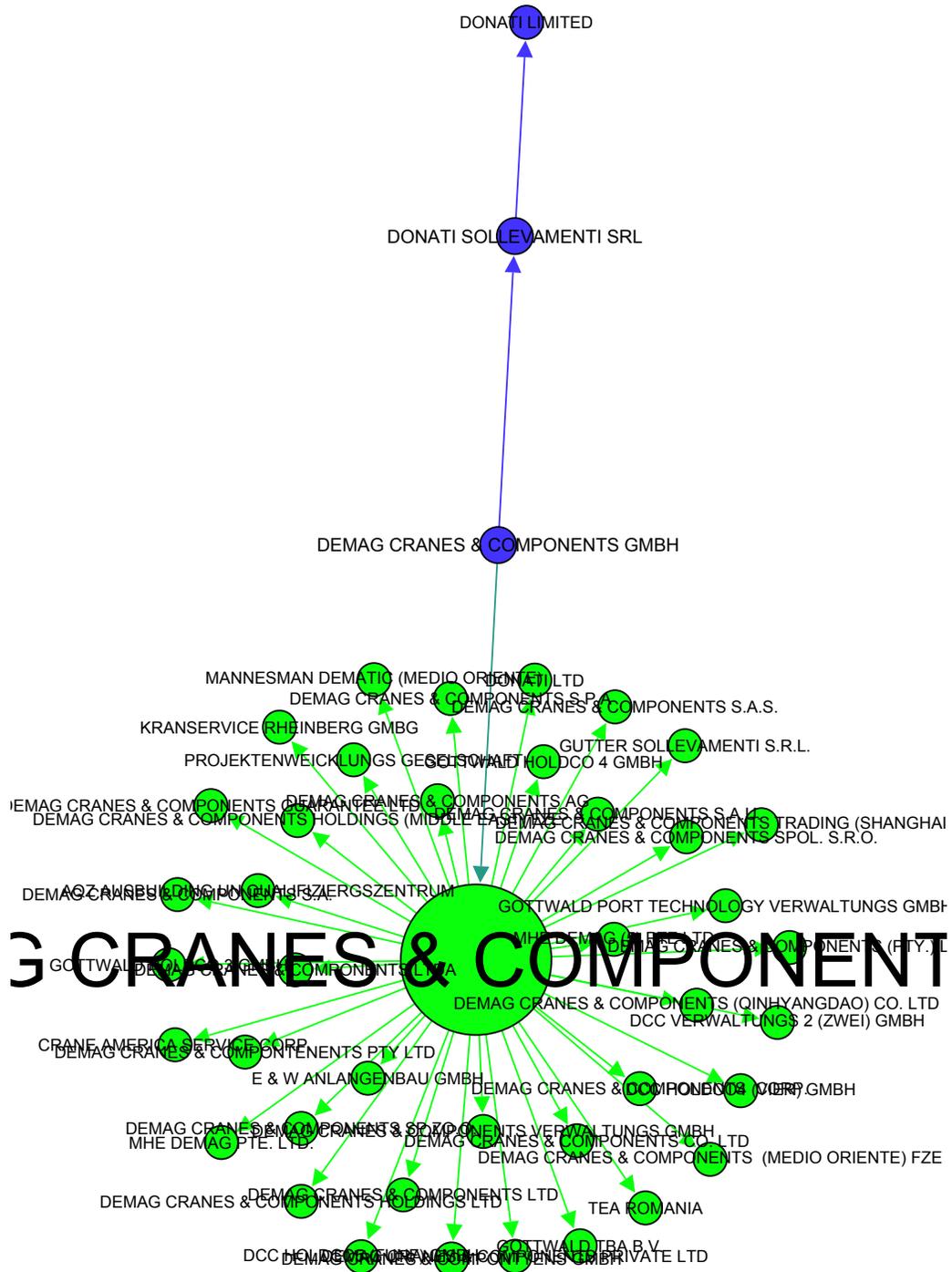


Figura 6.2.9: Gruppo 9 (Demag Cranes & Components GMBH)

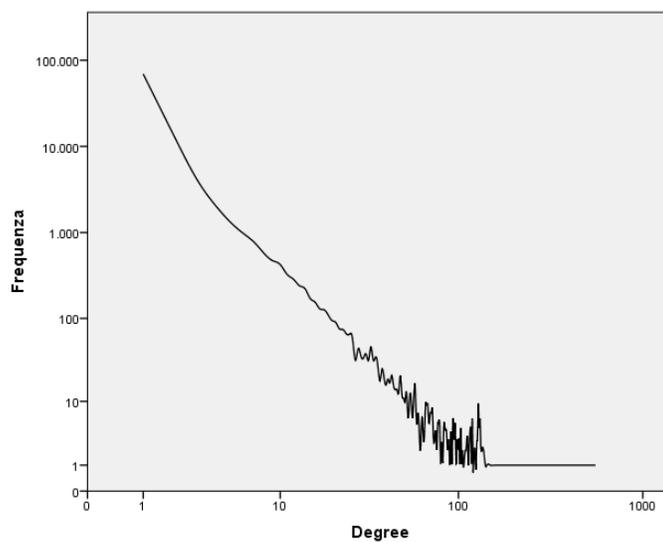


Figura 6.3.1: Dataset 2: distribuzione del degree nel gruppo principale

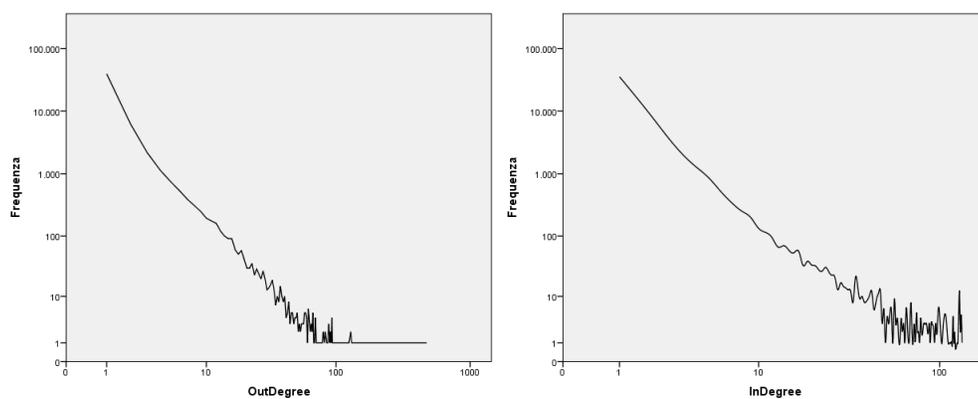


Figura 6.3.2: Dataset 2: distribuzione di in e out-degree nel gruppo principale

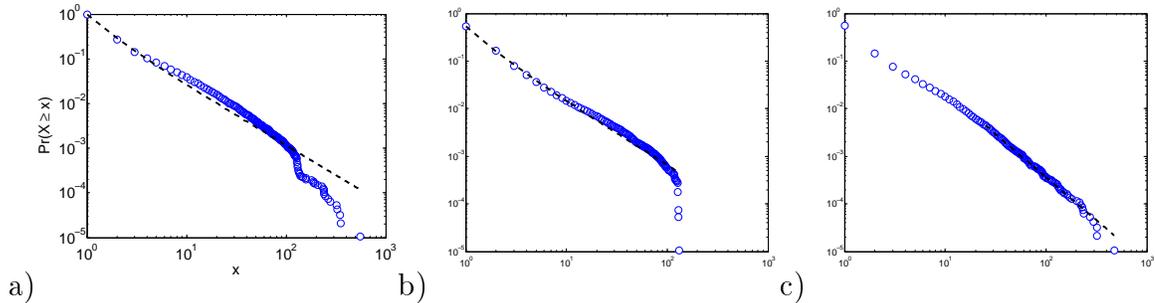


Figura 6.3.3: Dataset 2: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza. a) grafo diretto; b) grafo non diretto (in-degree); c) grafo non diretto (out-degree)

di valori di verosimiglianza per il dataset, che la distribuzione indica una power-law con "cut-off". Soprattutto nel caso *c*) si può notare questo tipo di distribuzione, dove il cut-off indica la parte non impressa da una linea tratteggiata.

6.3.2 Altre componenti

Le altre componenti, all'interno dei 10 gruppi col maggior numero di componenti connesse, rappresentano lo 0,43% dei nodi e lo 0,46% delle relazioni, figura 6.3.4. Una percentuale lievemente superiore rispetto al dataset 1, e che può sempre portare alla regola "80/20" di Pareto, anche all'interno di questo dataset.

Si può affermare che la distribuzione all'interno dei gruppi con il maggior numero di componenti connesse, segua una power-law. Da notare che alcune distribuzioni hanno delle piccole salite seguite immediatamente da discese, solitamente verso la coda della distribuzione, che potrebbe rappresentare l'effetto "noise" che si può vedere meglio nella distribuzione totale dei degree.

Studiando nel dettaglio la distribuzione, si ottengono i seguenti risultati, presentati in tabella 6.3.1.

Lo studio di come cambia la distribuzione dopo la separazione del degree in in-degree e out-degree porta, anche in questo caso a situazioni di distribuzione in cui non si segue un comportamento power-law, minori rispetto al dataset 1, come si può vedere in figura 6.3.5.

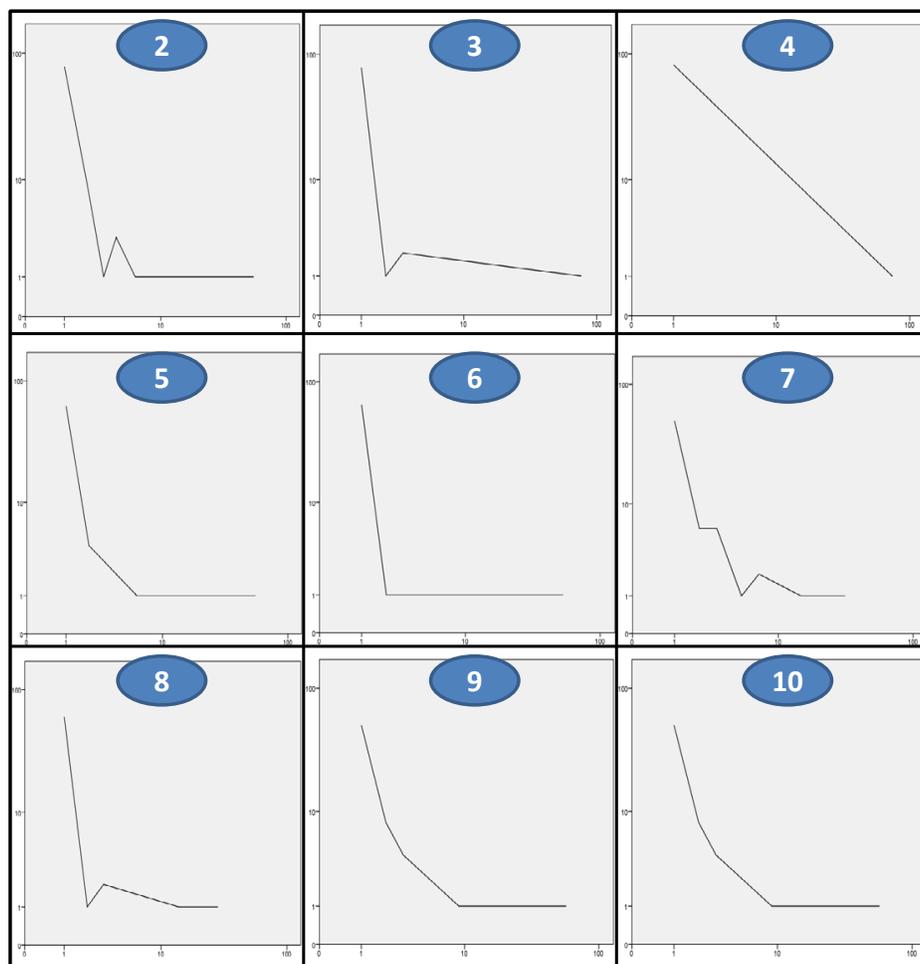


Figura 6.3.4: Dataset 2: distribuzione del degree nelle altre componenti di rete

Componente	α	x_{min}	$p - value$	power-law ?
2	2,67	1	0,0610	con cut-off ($x_{min} = 2$)
3	3,50	1	0,0430	con cut-off ($x_{min} = 2$)
4	3,50	1	0,0020	con cut-off ($x_{min} = 2$)
5	3,05	1	0,0060	con cut-off ($x_{min} = 2$)
6	3,44	1	0,0140	con cut-off ($x_{min} = 2$)
7	2,38	1	0,6600	si
8	2,91	1	0,0000	con cut-off ($x_{min} = 2$)
9	2,62	1	0,4340	si
10	3,50	1	0,0291	con cut-off ($x_{min} = 2$)

Tabella 6.3.3: Dataset 2: analisi della distribuzione del degree, altre componenti

Lo studio della distribuzione nel dettaglio, presentata in tabella 6.3.4, mostra alcuni casi in cui la distribuzione di in e out-degree non seguono una power-law e alcuni casi in cui il comportamento è anomalo solo per quanto riguarda l'in-degree. Questi casi verranno approfonditi in seguito.

Gruppo 4: Questo gruppo rappresenta la holding ("*Rhodia*") già incontrata nel gruppo 2 del dataset 1. Le caratteristiche sono le stesse, così come la distribuzione.

Gruppo 6: Il gruppo può rappresentare l'unione di due consorzi ("*Poggiorsini Acque S.p.A.*" e "*Vuelta International S.p.A.*") aventi settori differenti e uniti da un nodo che partecipa in entrambi. Rappresentando due consorzi si otterrà lo stesso risultato ottenuto per altri casi simili (figura 6.3.6) in cui la distribuzione del degree seguirà una power-law mentre l'in e out-degree non seguono questa distribuzione. Dal punto di vista economico la rete di imprese fa parte del primo tipo, e cioè rappresenta due reti con un centro di coordinamento.

Gruppo 7: il gruppo mostra principalmente due gruppi che agiscono nello stesso settore ("*D.A.EM S.p.A.*" e "*Gruppo Argenta S.p.A.*") e quindi in un sistema basato sulla concorrenza ma, avendo un rapporto di partnership commerciale a lungo termine ("*Universo Vending S.r.l.*"). Le due aziende operano nel campo delle macchine per la ristorazione, entrambe partecipano ("*Gruppo Argenta*" al 50%) in "*Universo Vending*



Figura 6.3.5: Dataset 2: distribuzione di in e out-degree nelle altre componenti di rete

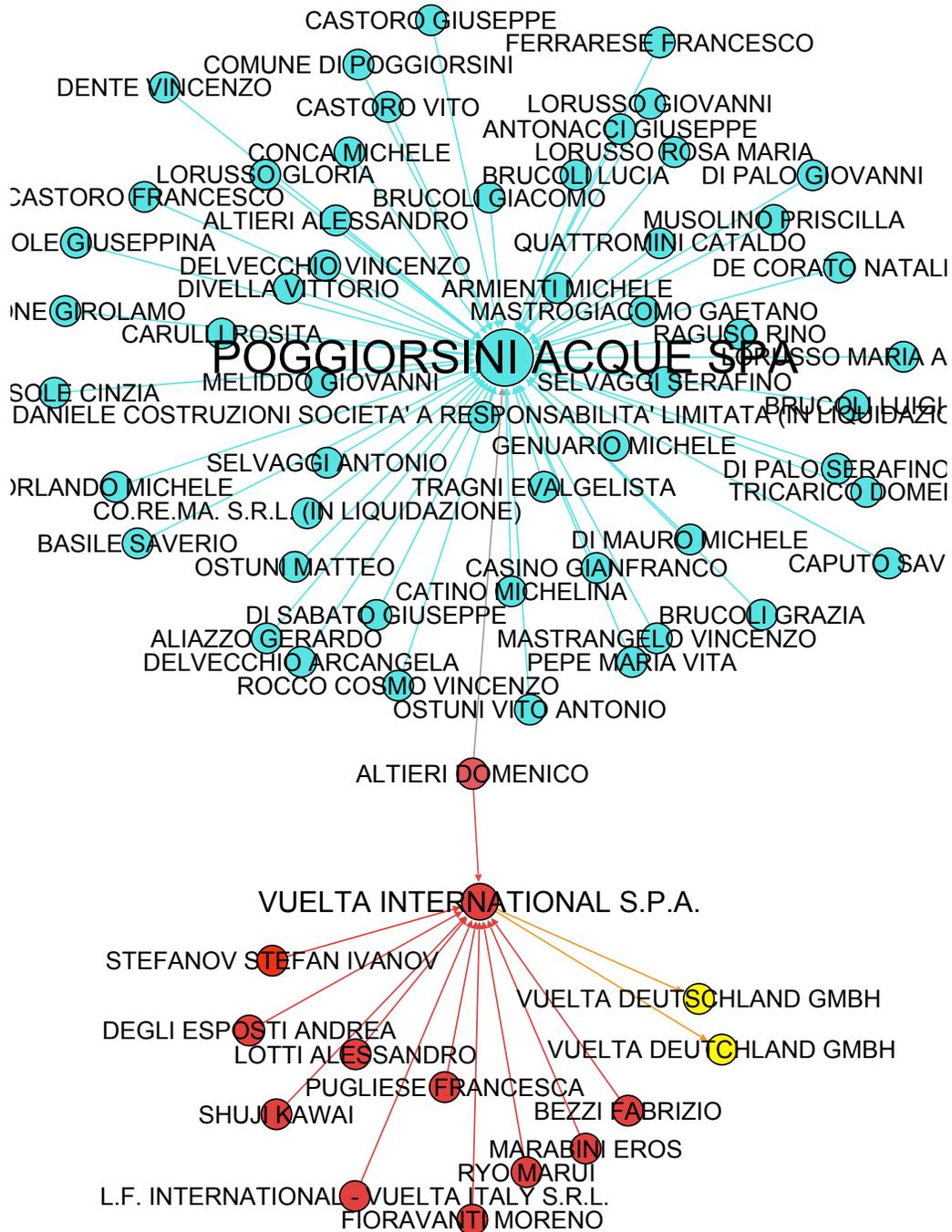


Figura 6.3.6: Gruppo 6 (Poggiorsini)

Componente	α	x_{min}	$p - value$	power-law ?
2	3,11/2,12	1/1	0,6790/0,5460	si/si
3	3,50/1,77	1/1	0,0020/0,6250	con cut-off ($x_{min} = 2$)/si
4	1,50/1,50	0/1	0/0	no/no
5	1,81/3,50	1/1	0,8589/0,3571	si/si
6	1,50/3,50	1/1	0/0,0090	no/no
7	2,38/1,99	1/1	0,6550/0,6780	si/si
8	1,72/3,50	1/1	0,5089/0,1990	si/si
9	3,50/1,61	1/1	0,3260/0,6222	si/si
10	1,50/1,55	0/1	0,0010/0,6667	no/si

Tabella 6.3.4: Dataset 2: analisi della distribuzione di in-degree e out-degree, altre componenti

S.r.l.”, azienda che opera nella riparazione delle macchine per la ristorazione. Come si può capire da questa descrizione, siamo nel campo in cui sono presenti due centri di riferimento, e quindi nella seconda categoria di reti di imprese. Nella seconda parte del grafo vediamo la presenza di 3 hub con una stretta rete di relazioni tra di loro, contestualizzati all’interno di uno specifico campo (vendita di auto). Si potrebbe definire come un piccolo distretto. Anche in questo, il gruppo presenta le caratteristiche di rete di impresa del primo tipo, cioè con un nodo che svolge il ruolo di coordinatore del gruppo, il nodo centrale. Ogni cluster della rete segue questa caratteristica.

La distribuzione dell’in-degree è rappresentata da un’iniziale forma power-law seguita da un disturbo, che sommato ai vari segnali di ”noise” presenti negli altri gruppi, e soprattutto nel primo, porta alla coda distorta presente della distribuzione del dataset.

6.4 Dicotomizzazione su Dataset 2

6.4.1 Componente gigante

Analizziamo ora come la distribuzione possa modificarsi dopo un’operazione di dicotomizzazione all’interno dei gruppi scelti per componente connesso.

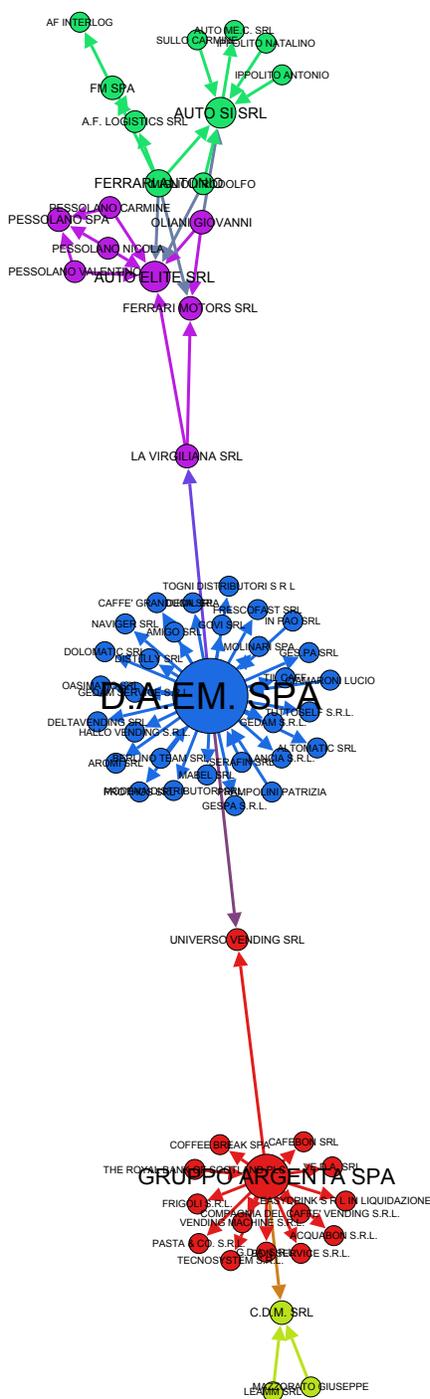


Figura 6.3.7: Gruppo 7 (D.A.EM)

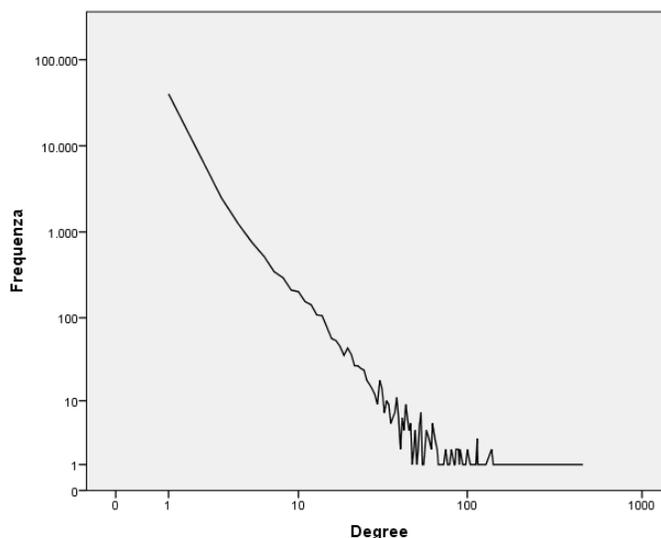


Figura 6.4.1: Dataset dicotomizzato: distribuzione del degree nel gruppo principale

Questa operazione porterà il livello minimo di partecipazione azionaria al 20%, e tutti i legami $< 20\%$ saranno esclusi.

Si ottiene quindi una rete filtrata, con la presenza di soggetti che possiedono una quota significativa dell'altra entità.

Studiando come si modifica la distribuzione del gruppo principale, contenente il maggior numero di elementi, si nota subito che la distribuzione, come grafo non diretto, si comporta come la distribuzione dicotomizzata presentata nel capitolo precedente (figura 6.4.1). La distribuzione nel caso di grafo diretto, invece, presenta delle differenze da analizzare nel dettaglio (figure 6.4.2).

La distribuzione dell'out-degree, graficamente, è molto simile alla distribuzione vista per il dataset 2: entrambe seguono una linea retta all'inizio e successivamente la funzione diventa "noisy". Per quanto riguarda la distribuzione dell'in-degree è ovvio aspettarsi una forma diversa rispetto a quelle viste in precedenza: la dicotomizzazione porta ad una riduzione drastica soprattutto all'interno delle partecipazioni entranti. Portando la percentuale minima al 20% ci si aspetta che il numero massimo di in-degree sia uguale a 5, in figura si vede come all'interno del dataset originale siano presenti, poche, incongruenze in cui la percentuale totale supera il 100%.

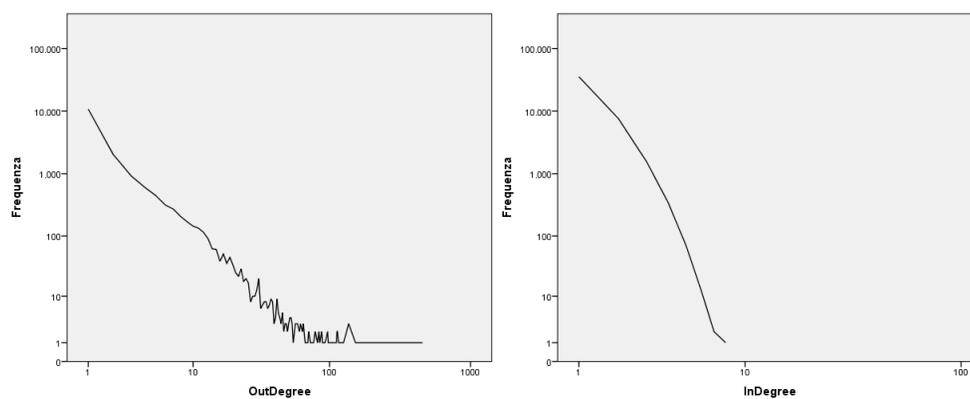


Figura 6.4.2: Dataset dicotomizzato: distribuzione di in e out-degree nel gruppo principale

Degree medio	Coefficiente di clustering	Lunghezza media del cammino	Diametro
1,10	0,038	7,528	25

Tabella 6.4.1: Dataset dicotomizzato: analisi delle metriche riguardanti il gruppo principale

Grafo	α	x_{min}	$p - value$	power-law ?
Non diretto	2,50	1	0,0050	con cut-off ($x_{min} = 2$)
Diretto	2,94 / 2,763	1 / 11	0 / 0,240	No / con cut-off ($x_{min} = 15$)

Tabella 6.4.2: Dataset dicotomizzato: Risultati dell'analisi sulla distribuzione power-law, componente principale

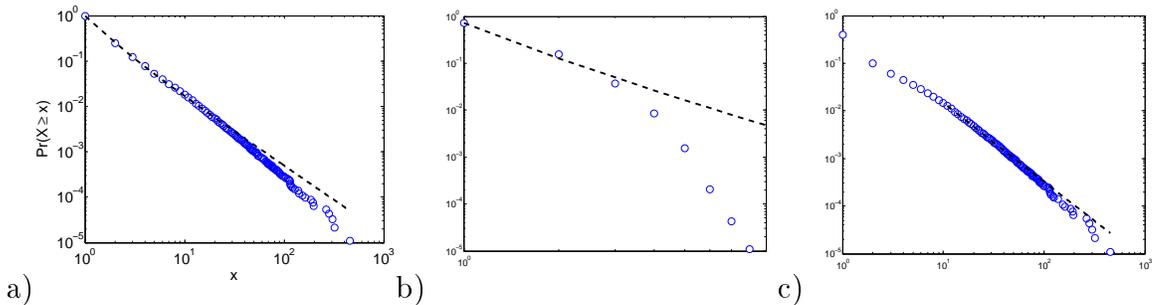


Figura 6.4.3: Dataset dicotomizzato: funzione di distribuzione cumulativa $P(x)$ e del massimo insieme di valori di verosimiglianza. a) grafo non diretto; b) grafo diretto (in-degree); c) grafo diretto (out-degree)

Anche in questo caso, il coefficiente di clustering rimane basso, rispetto ai casi presentati in letteratura, la lunghezza media del cammino e il diametro della rete sono molto simili a quelli osservati per il dataset 2, indicando che la dicotomizzazione, a livello globale, non influisce sui percorsi geodetici.

Vediamo i risultati delle nostre analisi sulla distribuzione riguardante il gruppo principale (tabella 6.4.2).

Anche in questo caso si nota, come si può vedere in figura 6.4.3, rappresentante la funzione di distribuzione cumulativa $P(x)$ e il massimo insieme di valori di verosimiglianza per il dataset, che la distribuzione indica una power-law con "cut-off", così come nello studio della distribuzione dell'intero dataset con dicotomizzazione del capitolo precedente. Trattandosi di "componente gigante" era preventivabile attendersi una distribuzione simile.

6.4.2 Altre componenti

In figura 6.4.4 si mostra la distribuzione delle altre componenti del dataset dicotomizzato.

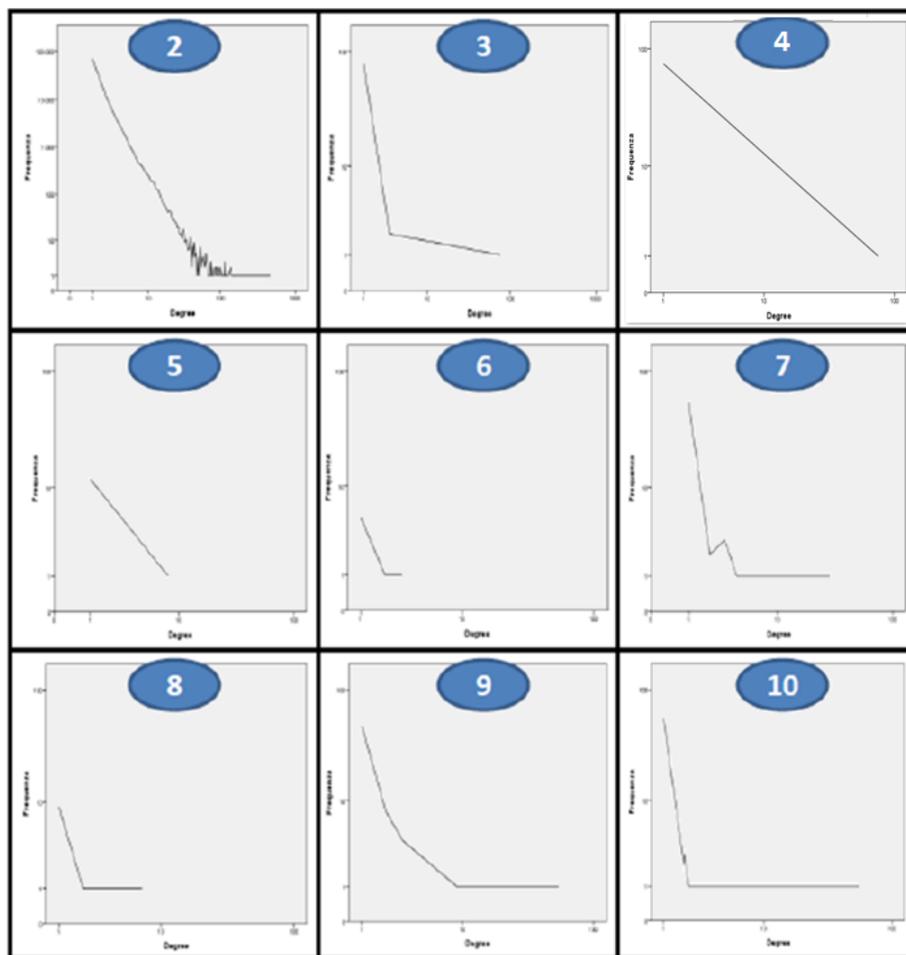


Figura 6.4.4: Dataset dicotomizzato: distribuzione del degree, altre componenti

Componente	α	x_{min}	$p - value$	power-law ?
2	3,50	1	0,0120	con cut-off ($x_{min} = 2$)
3	2,60	1	0,5310	si
4	3,50	1	0,0010	con cut-off ($x_{min} = 2$)
5	2,74	1	0,0130	con cut-off ($x_{min} = 2$)
6	2,59	1	0,4470	si
7	3,50	1	0,0373	con cut-off ($x_{min} = 2$)
8	2,59	1	0,3410	si
9	3,50	1	0,0450	con cut-off ($x_{min} = 2$)
10	3,50	1	0,0462	con cut-off ($x_{min} = 2$)

Tabella 6.4.3: Dataset dicotomizzato: analisi della distribuzione del degree, altre componenti

Si può affermare che la distribuzione all'interno dei gruppi rappresentati, nella maggior parte dei casi, segua una power-law. Alcuni gruppi, ad esempio il gruppo 7, hanno una forma che potrebbe indicare una distribuzione differente, ma, come visto in precedenza, la distribuzione segue comunque una power-law, con cut-off.

In tabella 6.4.3 vengono mostrati i risultati del test di verosimiglianza sulla distribuzione power-law del dataset dicotomizzato.

Lo studio di come cambia la distribuzione dopo la separazione del degree in in-degree e out-degree porta, anche in questo caso a situazioni di distribuzione anomale, minori rispetto al dataset 1, come si può vedere in figura 6.4.5. I dati vengono analizzati successivamente e mostrati in tabella 6.4.4.

Vedendo i gruppi che possono rappresentare una distribuzione differente da quella power-law, si può notare che in alcuni casi la distribuzione segue una forma già descritta in precedenza, come nel caso del *gruppo 4*.

Gruppo 10: il gruppo mostra la costellazione di partecipazioni della "B.S.E. Business Service Express S.r.l." società finanziaria con varie partecipazioni nel campo dei discount. Come si può chiaramente vedere dalla figura 6.4.6, è presente un nodo centrale che partecipa in tutti gli altri nodi e quindi, dal punto di vista di rete di imprese, del primo tipo con la presenza di un nodo che svolge la funzione di coordinatore. Questa forma rappresenta la tipica situazione delle società finanziarie, in cui il gruppo è formato da un nodo che comanda tutti gli altri, diverso dalla multinazionale, in cui è presente un nodo al livello superiore.



Figura 6.4.5: Dataset dicotomizzato: distribuzione di in e out-degree nelle altre componenti di rete

Componente	α	x_{min}	$p - value$	power-law ?
2	3,50/1,68	1/1	0,0010/0,6667	con cut-off ($x_{min} = 2$)/si
3	3,49/2	1/2	0,1870/0,6833	si/si
4	1,50/1,50	0/1	0/0	no/no
5	3,36/2,07	1/1	0,1700/0,4820	si/si
6	3,48/1,54	1/1	0,2220/0,3889	si/si
7	1,50/1,55	0/1	0,6511/0,3750	si/si
8	3,50/1,74	1/1	0,4286/0,6598	si/si
9	3,50/1,58	1/1	0,0272/0,6250	no/si
10	1,50/1,50	0/0	0/0	no/no

Tabella 6.4.4: Dataset dicotomizzato: analisi della distribuzione di in-degree e out-degree, altre componenti

Come si poteva notare anche dalla figura rappresentante l'in e out-degree del gruppo, l'in-degree non segue una power-law, tutti i nodi hanno $in-degree = 1$ a parte il nodo principale che ha $in-degree = 0$. L'ipotesi power-law viene quindi scartata.

Dall'analisi della distribuzione dei gruppi all'interno dei 3 dataset si può arrivare ad una serie di conclusioni sulla forma della distribuzione e sulle reti di imprese analizzata:

- i dataset seguono una power-law con cut-off;
- tutti i dataset studiati hanno un coefficiente di clustering basso, diversamente da quanto definito in letteratura;
- i gruppi all'interno dei vari dataset, se considerati come grafo non diretto, seguono una power-law;
- alcuni gruppi, se considerati come grafo diretto, non seguono un modello di distribuzione power-law, alcuni solo nella distribuzioni in-degree, altri anche nell'out-degree;
- la maggior parte dei gruppi che non seguono una power law, nella distribuzione dell'in e/o out-degree, rappresentano una rete di imprese con un centro di riferimento;

Capitolo 7

Centralità e potere

Il potere è una delle proprietà fondamentali di un sistema economico e sociale, quale ne sia la sua struttura. Non sono presenti, in sociologia ed economia, definizioni condivise di potere, nè per conseguenza, vi sono metodi univoci per caratterizzarlo dal punto di vista distributivo (*che lo ha*), della significatività (*quanto ne ha*), delle cause e delle conseguenze.

Un aspetto comune a varie definizioni è il concetto di "*asimmetria*" (informativa, della capacità di influenzamento, ecc...): il potere è infatti strutturalmente una *condizione di disequilibrio* di possibilità, di dipendenza (se lo si subisce) o di dominanza (nella misura in cui lo si esercita).

In questo capitolo si applicheranno vari metodi di centralità all'interno della componente gigante del dataset dicotomizzato al 20% di partecipazione azionaria; si è scelto questo gruppo per la sua significatività rispetto alle altre componenti del dataset, in cui si può facilmente riconoscere chi sono i nodi centrali, o maggiormente importanti/influenti. Il motivo è principalmente quello di ottenere delle informazioni da un'insieme di dati complessi; si cercherà quindi di contestualizzare ed applicare i vari concetti di centralità all'interno del dataset.

Si è scelto di utilizzare il dataset dicotomizzato anche per cercare di ridurre il numero di "*falsi positivi*", cioè nodi che hanno un grande numero di partecipazioni non rilevanti. Portando il livello di "*rilevanza*" di una partecipazione azionaria al 20%; non si può tuttavia assicurare l'assenza di falsi positivi, che vengono comunque ridotti in gran parte.

Secondo Soda [36] un "*contesto relazionale*" è composto da attori e da relazioni o legami. Lo studio del contenuto, delle entità e della natura delle relazioni, non aiuta complessivamente nella comprensione della complessità

di un network; non è possibile, in modo intuitivo, differenziare le posizioni occupate dagli attori al suo interno, e, allo stesso modo, non è possibile essere in grado di identificare i gruppi di attori maggiormente coesi o simili.

In un approccio di rete, si tratta di individuare le misure applicabili che possono presentarsi come migliore *proxy* del concetto. Esse sono individuate dalla letteratura in primis nel concetto di centralità (degree, closeness e betweenness), con particolare attenzione all'approccio detto dei "*buchi strutturali*" di Burt [9].

Il "*network thinking*" ha contribuito allo studio del potere sociale, sottolineando che il potere è *intrinsecamente relazionale* [23]. Una entità non ha potere in *astratto*, ha potere perchè può *dominare* altre entità, il potere dell'ego è la *dipendenza dai cambiamenti*. Poichè il potere è una conseguenza del modello dei rapporti, la quantità di potere nelle strutture sociali può, a sua volta, variare. Se un sistema è debolmente accoppiato (a bassa densità) non si può esercitare molto potere e, al contrario, in sistemi ad alta densità c'è la possibilità di esercitare potere superiore.

Il potere ha sia proprietà sistemiche (*macro*) che relazionali (*micro*). La quantità di potere in un sistema e la sua distribuzione fra gli attori sono collegate, ma non indicano la stessa cosa. Due sistemi possono avere la stessa quantità di potere, ma può essere distribuita equamente all'interno del primo e non esserlo nel secondo. Il potere nelle reti sociali può essere visto come una proprietà micro (ad esempio nel descrivere le relazioni tra entità) o una proprietà macro (ad esempio nel descrivere un'intera popolazione), e, così come in altri concetti sociologici, le proprietà macro e micro sono strettamente connesse nel network thinking.

Spesso si descrive il modo in cui un'entità è integrata all'interno di una rete relazionale come un'imposizione di vincoli o come offerta di opportunità per l'attore. Gli attori che incontrano meno vincoli, e hanno quindi più opportunità rispetto ad altri, sono in posizioni strutturali favorevoli. Avere una posizione favorita significa che l'attore può estrarre vantaggi negli scambi, avere maggiore influenza, e fungere da *punto di riferimento* per deferenza e attenzione di coloro che occupano posizioni svantaggiate.

Non esiste una singola definizione per rispondere a cosa significhi avere una "*posizione favorita*" in un network, così come il significato di "*avere più opportunità*" o "*meno vincoli*". L'analisi delle reti ha dato un contributo importante nel fornire definizioni precise e misure concrete riguardanti diversi approcci alla nozione di *potere*, che si attribuisce alle posizioni nelle strutture di relazioni.

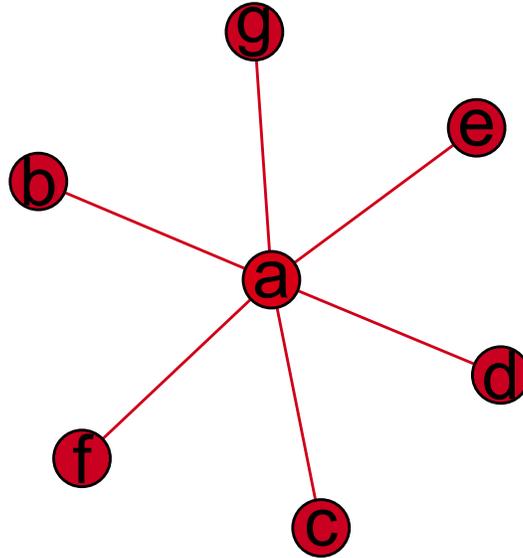


Figura 7.0.1: Esempio di rete a stella

Prendendo come esempio la figura 7.0.1 rappresentante una rete a stella, è facile supporre che l'attore "a" abbia una posizione favorita nella struttura di rete. Questo è un esempio molto semplice, come si può calcolare in generale una posizione di vantaggio all'interno di una rete?

E' necessario riflettere sul perchè la posizione strutturale possa essere vantaggiosa o svantaggiosa per gli attori. Ritornando all'esempio, concentriamo l'attenzione su perchè l'attore "a" è così in vantaggio nella rete a stella.

Degree: nella rete a stella, l'attore "a" ha più opportunità e alternative rispetto agli altri attori. Se l'attore "b" sceglie, ad esempio, di non fornire risorse ad "a", l'entità centrale ha comunque altri sbocchi comunicativi, mentre se l'attore "b" decide di non comunicare con "a" non avrà altri sbocchi comunicativi. Quindi si può arrivare alla conclusione che il potere dipenda dal maggior numero di legami che un nodo ha. Se aumenta il numero di legami, di conseguenza ne aumenta il potere di quel nodo nei confronti della rete. Nell'esempio l'attore "a" ha $degree = 6$ mentre gli altri nodi hanno $degree = 1$. Questa logica è alla base delle misure di centralità e potere basate sul degree dei nodi. Gli attori che hanno più legami avranno maggiori opportunità date dal fatto che possiedono

più scelte. Questa autonomia li rende meno dipendenti da ogni altro attore e quindi più potenti.

Closeness: il secondo motivo per cui l'attore "a" ha più opportunità rispetto ad altri nodi, nell'esempio, riguarda la vicinanza dell'attore ad altri, rispetto agli altri nodi. Il potere può essere esercitato dalla contrattazione diretta e lo scambio, ma anche dalla capacità di un nodo di agire come "*punto di riferimento*" tramite il quale gli attori possono comunicare. Questo concetto si può applicare, in questo contesto, alla teoria basata sull'economia dei costi di transazione per cui, gli attori che riescono a raggiungere altri nodi con una distanza geodetica minore rispetto agli altri ne saranno avvantaggiati. Nella rete a stella, l'attore "a" è ad una distanza geodetica di uno da tutti gli altri operatori.

Betweenness: la terza ragione per cui l'attore "a" è avvantaggiato nella rete a stella è dato dal fatto che l'attore "a" risiede tra altre paia di nodi (da cui "betweenness"), e nessun altro attore si trova tra "a" e gli altri soggetti. Se "a" vuole contattare "f" può semplicemente farlo; se "f" vuole contattare "c", deve passare forzatamente per "a". Questo dà all'attore "a" la capacità di fungere da "*mediatore*" tra altri attori, e quindi anche la capacità di isolare e prevenire contatti tra altri nodi.

In letteratura si preferisce utilizzare, al termine "*potere*", il termine di centralità. Ognuno dei tre approcci presentati descrive la posizione degli individui in termini di quanto sono vicino al "*centro*" dell'azione nella rete, anche se sono presenti differenze su come si determini il centro di una rete. E' più corretto definire questi approcci come misure di centralità, ma ci sono diversi motivi per cui le posizioni centrali tendono ad essere posizioni di potere.

Nel nostro modello i legami, nel dataset 2, avranno un valore, da aggiungere al degree. Mentre per il calcolo del degree ogni relazione avrà valore di 1, per quanto riguarda le partecipazioni azionarie si deve tenere conto della percentuale di partecipazione che lega due soggetti e quindi del loro rapporto di dipendenza. Quindi applicando misure di centralità e varie dicotomizzazioni basate sulla percentuale di partecipazione e sui dati finanziari legati ai nodi, sarà possibile arrivare ad una ipotetica "*mappe del potere*".

7.1 Potere e influenza nelle organizzazioni

In letteratura, le definizioni di potere e influenza all'interno di una organizzazione sono state lungamente dibattute. Brass [8] propone un riassunto di tutte queste teorie, arrivando alla definizione di potere come l'inverso della dipendenza e derivante dal controllo di risorse rilevanti; il controllo di un attore implica che, all'interno della relazione sociale, gli altri soggetti abbiano poche fonti alternative da cui acquisire le risorse, e questo porta gli altri ad avere una *dipendenza* dal soggetto potente. La complessità della *relazione tra dipendenza e potere* aumenta quando si considera la moltitudine di varietà di risorse che si possono considerare rilevanti in organizzazioni complesse; un attore *a* potrebbe controllare una particolare risorsa che può essere ritenuta rilevante da *b*, mentre *b* può detenere una risorsa ritenuta rilevante dall'attore *a*. In questi casi, focalizzarsi su una sola risorsa può portare a risultati non accurati.

Brass conclude che al fine di incrementare il potere sono necessarie *due condizioni*: un attore deve diminuire la sua dipendenza dagli altri e aumentare la dipendenza verso di lui.

Nella letteratura riguardante il potere, gli attori o le unità occupanti posizioni centrali all'interno di una rete sono considerate come possibili potenti grazie alla loro facilità di *accesso*, e possibilità di *controllo*, di risorse rilevanti. Secondo Astley e Fombrun [1] la centralità in una rete di comunicazioni è relativa all'influenza attribuita al soggetto, e l'effetto è relativamente minore quando viene comparata con la posizione formale, il rank e lo status dello stesso.

I vari studi riguardanti la centralità presenti in letteratura conclusero individuando una generale nozione di centralità, secondo cui le persone in posizioni centrali tendono ad emergere come leader del gruppo[8]. Sono emerse una serie di variazioni a questa nozione con il passare degli anni. Il lavoro di Freeman [19] ha tentato un chiarimento nel concetto di centralità in relazione a misure particolari. Freeman divise le misure di centralità in 3 categorie: *degree* (o numero di contatti), *betwenness* e *closeness* (o prossimità).

7.2 Degree centrality

Gli attori che hanno molte relazioni con altri attori possono avere una posizione vantaggiosa. La misura di centralità basata su degree è la più semplice,

ma spesso più efficace, delle misure di centralità di un'attore. Il possibile potere è il loro degree.

In grafi non diretti, gli attori differiscono tra di loro solo per quante connessioni posseggono. Con i grafi diretti, invece, può essere importante distinguere la centralità basata sull'in-degree dalla centralità basata sull'out-degree.

Se un'attore riceve molti collegamenti, si può definire come "*prominente*" o avere alto "*prestigio*", e questo può indicare la sua importanza. Gli attori che hanno un'elevato out-degree sono soggetti che sono in grado di effettuare scambi con molti altri.

7.2.1 Degree centrality come coinvolgimento nella rete

L'idea alla base delle teorie di Freeman [19] è che all'interno di una struttura relazionale, l'importanza di un'attore possa essere determinata da fattori molteplici, riconducibili essenzialmente alla posizione occupata dallo stesso rispetto agli altri. Se un attore ha molti legami all'interno di un network, si può definire più "centrale" rispetto ad altri nella rete.

Tornando al nostro dataset, si mostra in figura 7.2.1 la rappresentazione dei nodi che posseggono un out-degree superiore a 20. Questa scelta è stata fatta per permettere una rappresentazione migliore del risultato, in questo caso i legami non sono importanti per la comprensione della rete, e si è deciso quindi di escluderli. La dimensione dei nodi è stata scelta in base al valore dell'out-degree, come la scala cromatica dal rosa (out-degree basso) al viola (out-degree alto).

Questa classificazione può essere interpretata come il livello di coinvolgimento dei nodi all'interno della rete, la centralità si basa sul numero di linee che sono incidenti ad un nodo rispetto agli altri [33].

Tuttavia questa misura non si può ritenere completa. Il numero di relazioni che convergono su un attore non è necessariamente una spiegazione diretta del potere di cui dispone l'attore. I nodi di grandi dimensioni, come quelli rappresentati nella figura precedente, rappresentano gruppi di società che possono raggiungere un elevato livello di complessità, come nei gruppi gerarchici, o "*gruppi di gruppi*", presentati in figura 7.2.2.

I gruppi gerarchici si formano solitamente quando le unità controllate operano in diversi settori di attività. Queste divisioni mostrano anche che, se un'azienda opera in più campi in modo così elevato da creare società divise, tale azienda si può definire di grandi dimensioni. In conclusione la degree centrality può servire a spiegare tre caratteristiche del nostro dataset:

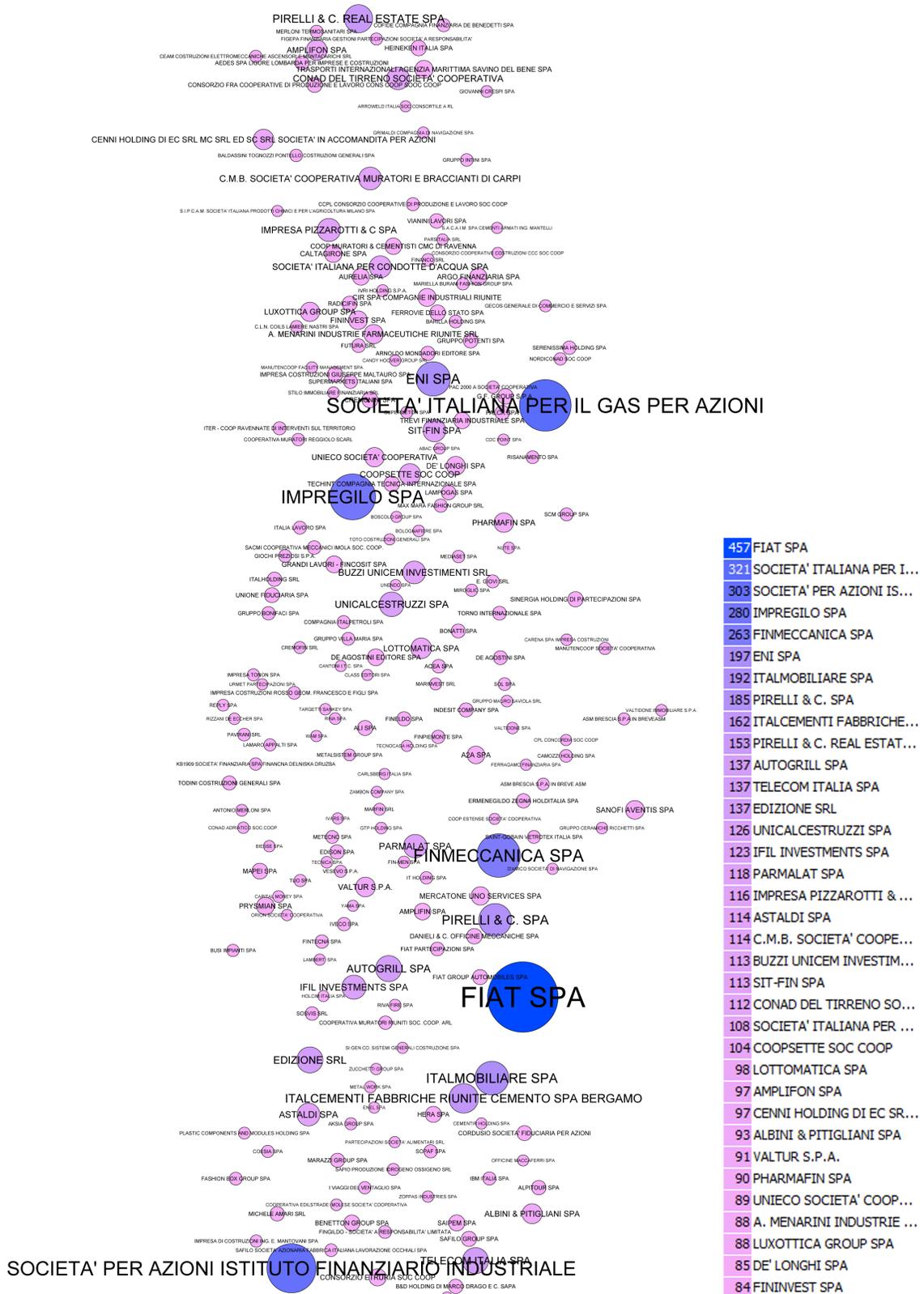


Figura 7.2.1: Degree centrality basata su out-degree (rete di influenza)

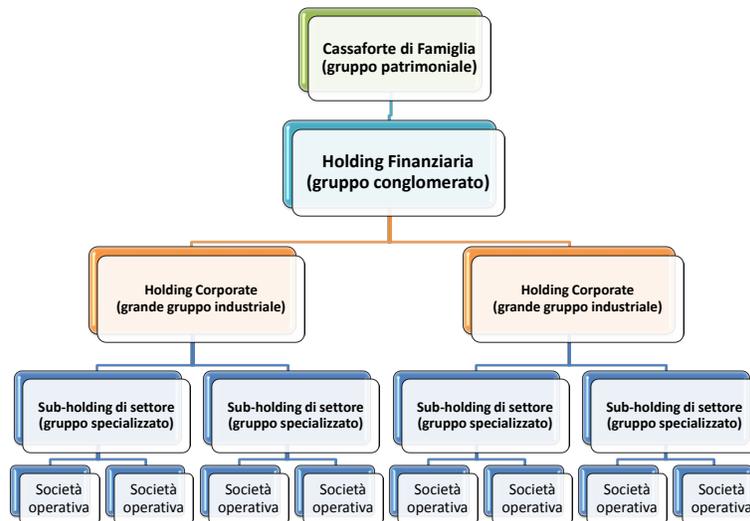


Figura 7.2.2: Gruppi di gruppi [6]

- aiutare nell'individuazione dei grandi gruppi industriali/finanziari;
- individuare il numero di connessioni dirette delle entità;
- individuare il coinvolgimento di un nodo all'interno della rete.

Il caso Fiat S.p.A.: Prendiamo il gruppo *"Fiat S.p.A"* controllato dalla famiglia Agnelli. La cassaforte di famiglia è rappresentata dal gruppo *"Giovanni Agnelli & C. Sapa"*, che controlla il 53% del capitale votante della holding finanziaria quotata del gruppo Agnelli, *"IFI S.p.A."* (Istituto Finanziario Industriale). Quest'ultima controlla a sua volta l'altra holding finanziaria quotata *"IFIL S.p.A."* con il 64,98% del capitale. *"IFIL S.p.A."* controlla di fatto, l'holding industriale *"Fiat S.p.A."*, che controlla a sua volta *"Fiat Auto S.p.A."*, *"Iveco S.p.A."*, *"Maserati S.p.A."* e altre (figura 7.2.3).

Questo esempio dimostra come la degree centrality non sia sufficiente per dimostrare le eventuali posizioni di potere all'interno di una rete, *"Fiat S.p.A."* ha un out-degree di 457, molte delle quali sono società del gruppo Fiat sui cui l'influenza della holding centrale è diretta.

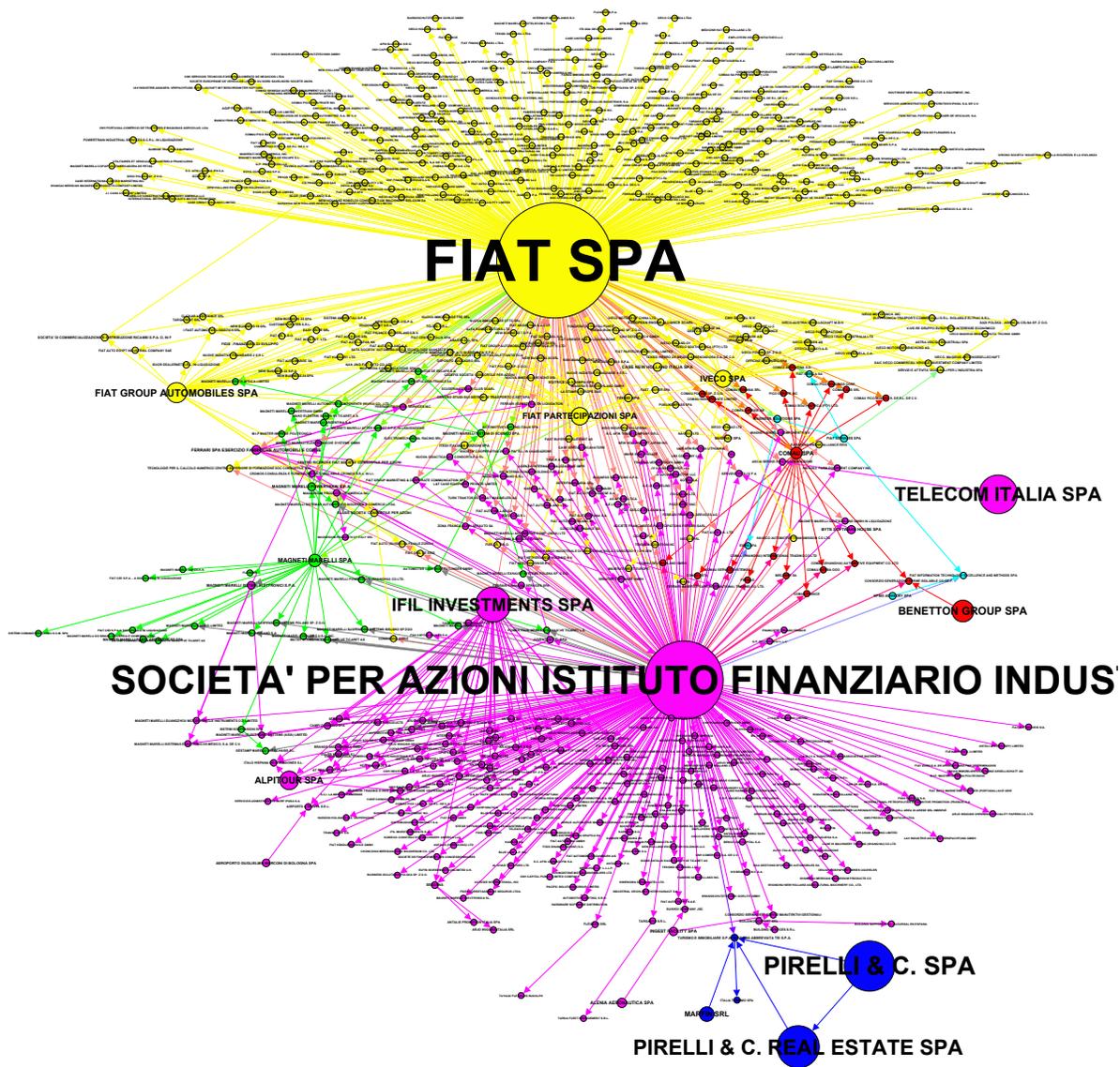


Figura 7.2.3: Ego network "Fiat S.p.A."

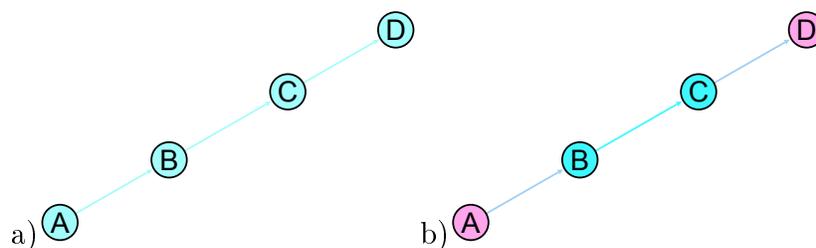


Figura 7.3.1: Esempio di rete lineare. a) senza misura di betweenness; b) con misura di betweenness

7.3 Betweenness centrality

La misura di centralità basata su betweenness (cioè l'abilità nello stare al centro) è la misura della centralità di un nodo all'interno di una rete; l'equivalente del numero di cammini minimi che partono da un vertice, verso tutti gli altri che passano per un nodo. La betweenness centrality dà un'idea del volume dei flussi informativi tra due nodi, che passano da un intermediario, il nodo centrale. Un nodo quindi si può definire centrale se controlla una grande parte dei flussi informativi all'interno di una rete. Rispetto alla degree centrality si può subito notare che come misura può essere più indicativa nel rappresentare l'importanza di un nodo, da tale importanza è possibile anche attribuire la probabilità che ha quel nodo di influenzare i suoi nodi "vicini", data dalla sua funzione di intermediazione tra due punti del grafo.

Ipoteticamente, se all'interno di una rete di relazioni sociali, un'attore a volesse influenzare l'opinione dei suoi conoscenti, cioè dei suoi nodi uscenti, avrà più possibilità di un'attore b nel caso in cui la sua rete di relazioni sia superiore.

Vediamo l'esempio in figura 7.3.1, mostrante un semplice grafo diretto e lineare. Rispetto alla rete a stella presentata precedentemente, in cui è evidente il nodo con più alta betweenness (a), in questo caso rappresentante 4 nodi con legami diretti, i nodi con maggiore betweenness all'interno del grafo sono, come dice il nome, quelli centrali; si può arrivare alla conclusione che la misura di betweenness non cambia da grafo diretto a non diretto, questo la rende una misura di centralità, in termini di potere, poco efficace, ma sicuramente migliore rispetto alla misura di degree centrality.

Una *sfera di influenza* si può definire come un'area intorno alla quale un'entità, nel nostro caso un'impresa o soggetto, esercita una dominazione

culturale indiretta, economica o politica. *L'influenza*, come il potere, non si può concretizzare. Non si può dire con certezza che il soggetto b ha più probabilità di influenzare i suoi conoscenti rispetto al soggetto a . Diciamo che, utilizzando un termine informatico, ha un "*framework*" migliore rispetto agli altri. Con *framework* si può indicare, in questo caso, una struttura logica di supporto (la rete di relazioni) per ottenere uno scopo.

Cercando di unire i flussi informativi con il "potere" (che in questo caso sarebbe più appropriato definire come influenza) è necessario ritornare ai motivi per cui un mercato si definisce *imperfetto*. Uno dei motivi dell'imperfezione dei mercati è *l'asimmetria informativa* [28], definita come la condizione di mercato in cui alcuni soggetti dispongono di un'informazione che altri non hanno.

Contestualizzando questi concetti all'interno del nostro modello, risulta chiaro che un nodo ottiene flussi informativi in base alle sue relazioni con gli altri nodi all'interno della rete, qualunque sia la sua percentuale di partecipazione. La misura di *betweenness centrality* può essere rappresentata in due modi nel nostro contesto:

- come misura del *grado di attrazione dei flussi informativi*, e quindi senza l'utilizzo di dicotomizzazione sulla percentuale di partecipazione;
- come misura di *influenza* che un nodo può generare all'interno del grafo, rappresentato da legami dicotomizzati al 20% di partecipazione

7.3.1 Betweenness centrality come capacità di intermediazione

Vediamo, in figura 7.3.2 i nodi con il valore di *betweenness centrality* maggiore. In questo caso si è scelto di mantenere i legami nella rappresentazione per indicare che i legami con *betweenness* maggiore possiedono un cammino che li può collegare, a tutti gli altri nodi della rete. I risultati di questa misura di centralità sono molto differenti rispetto alla misura precedente. Il motivo è spiegato direttamente dalla definizione di *betweenness centrality*, secondo cui la distanza geodesica, che misura i cammini minimi tra due punti, ha la capacità di analizzare l'intero diametro della rete, mentre la *degree centrality* si ferma ai punti collegati direttamente.

Il risultato interessante è notare come ai primi posti si posizionino 3 soggetti di tipologia diversa: "*Rizzoli Corriere della Sera mediagroup S.p.A.*"

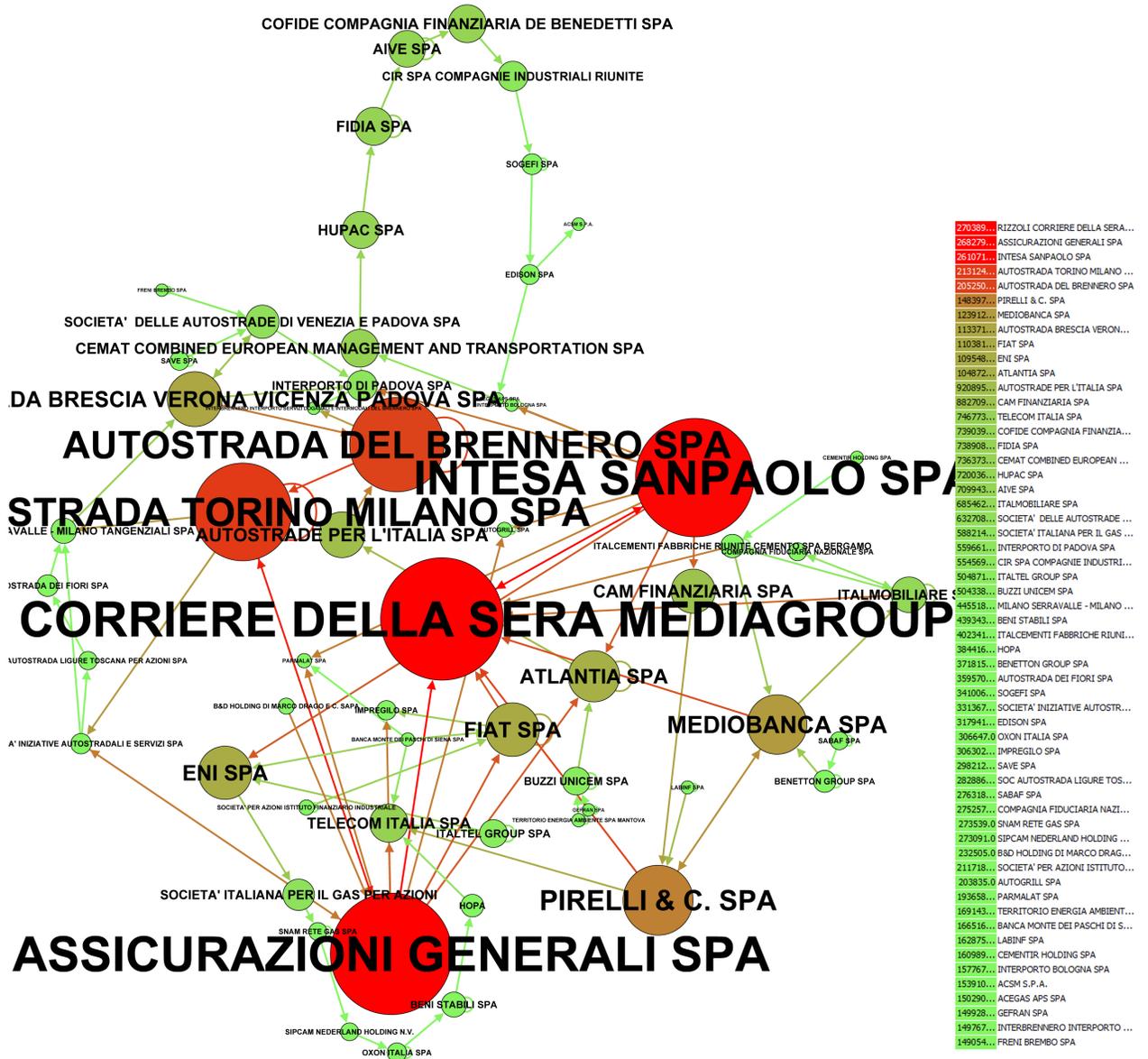


Figura 7.3.2: Betweenness centrality (rete di influenza)

rappresenta il gruppo editoriale più importante d'Italia, "Assicurazioni Generali S.p.A." rappresenta una impresa di assicurazioni e "Intesa San Paolo S.p.A." rappresenta una banca. Non si possono accomunare tutti i casi e affermare semplicisticamente che questi soggetti detengono grandi flussi informativi e, di conseguenza detengono, potere, non è così. E' necessario distinguere anche altri motivi, oltre a quelli informativi, per cui un nodo, nel nostro caso, può avere alta betweenness:

- un soggetto che partecipa in molte aziende (e quindi con out-degree alto) può avere un'attitudine maggiore alla *diversificazione degli investimenti*;
- un soggetto può ricevere molte partecipazioni in base alla *renumerosità dell'investimento*;
- un soggetto può ricevere molte partecipazioni per *motivi politici*, nel caso in cui il nodo sia considerato come nevralgico nel suo contesto e diversi soggetti si contendano il comando della stesso (ad esempio "Rizzoli Corriere della Sera mediagroup S.p.A.");

Come si potrà vedere nel paragrafo successivo, in cui la dicotomizzazione porta alla rottura in vari gruppi della componente gigante, i nodi bancari svolgono il ruolo che Granovetter definì di "local bridge" nel suo studio sulla forza dei legami deboli ("*Strenght of weak ties*") [20]. In una rete sociale, le relazioni che svolgono il ruolo di "ponte" trasmettono informazioni da un gruppo ad un altro, e quindi la diffusione delle informazioni dipende fortemente dal numero e dalla connessione dei ponti a disposizione del network.

Il caso Rizzoli Corriere della Sera mediagroup S.p.A.: "Rizzoli Corriere della Sera mediagroup S.p.A." è uno dei principali gruppi editoriali italiani, attivo a livello nazionale e internazionale nei seguenti mercati: quotidiani, libri, periodici, radio, nella televisione e inoltre sul web e nella raccolta pubblicitaria. Come si può notare dalla figura 7.3.3, rappresentante la ego network del nodo "Rizzoli Corriere della Sera mediagroup S.p.A." con grandezza dei nodi in base alla betweenness centrality e colori in base alla modularità, la maggior parte dei gruppi industriali della borsa italiana partecipano in questo nodo (con maggioranza relativa del 20% da parte di "Fiat partecipazioni S.p.A."). E' inoltre interessante vedere le attività al di fuori del nodo alla base della

ego network; i nodi rappresentanti banche ed assicurazioni agiscono con svariate partecipazioni in molti dei gruppi industriali rappresentati in questo grafo. Le partecipazioni di questi soggetti sono sempre con percentuali basse, non dominanti, ma molto variegata e, di fatto, si può arrivare a concludere che i nodi bancari e assicurativi siano i maggiori intermediari all'interno di questa rete.

7.3.2 Betweenness centrality come influenza nella rete

Vediamo, in figura 7.3.4 i nodi con il valore di betweenness centrality maggiore.

Dall'immagine si nota come la classifica, in ordine di betweenness centrality, sul componente gigante dicotomizzato, sia differente rispetto alla misura di centralità precedente. Dicotomizzando la componente gigante al 20% di partecipazione, si escludono, quasi totalmente, i nodi bancari ed assicurativi che, come detto nel paragrafo precedente, partecipano in un grande numero di nodi della rete, con un livello di partecipazione basso.

Analizzando i dati si può arrivare a varie conclusioni, assumendo che un valore di betweenness centrality alto indica un nodo che è molto presente nei cammini minimi di buona parte della rete sottostante:

- la betweenness centrality, su grafo dicotomizzato, indica che un nodo può avere più possibilità, rispetto ad altri, di influire direttamente nei confronti dei nodi collegati ad esso;
- un nodo con alto valore di betweenness centrality può rivelarsi come eventuale "punto di rottura" di un sistema. La caduta di un nodo può provocare un'effetto a cascata negativo nei confronti dell'intero network;
- se riprendiamo la definizione di betweenness data in precedenza nel contesto dei rapporti sociali, la possibilità di fungere da ponte tra vari nodi può dare la possibilità di raggiungere il massimo numero di risorse, soprattutto informazioni, che la rete ha da offrire. Questa idea può portare alla conclusione che un nodo con alta betweenness possa essere influente sia all'interno che all'esterno del sistema stesso.

Il caso Impregilo S.p.A.: "Impregilo S.p.A." è il principale gruppo italiano nel settore delle costruzioni e dell'ingegneria per dimensioni e ricavi,

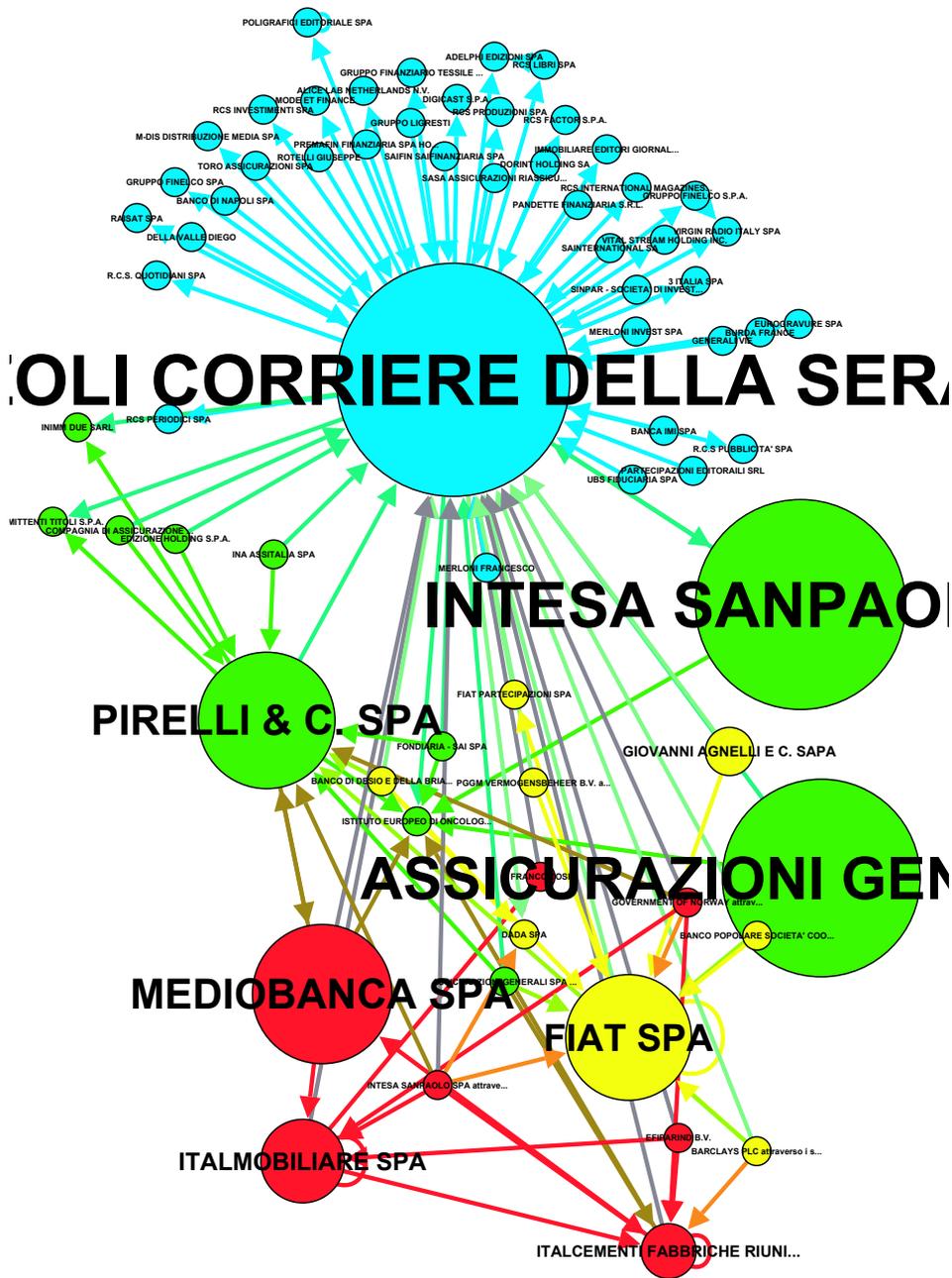


Figura 7.3.3: Ego network "Rizzoli Corriere della Sera mediagroup S.p.A."

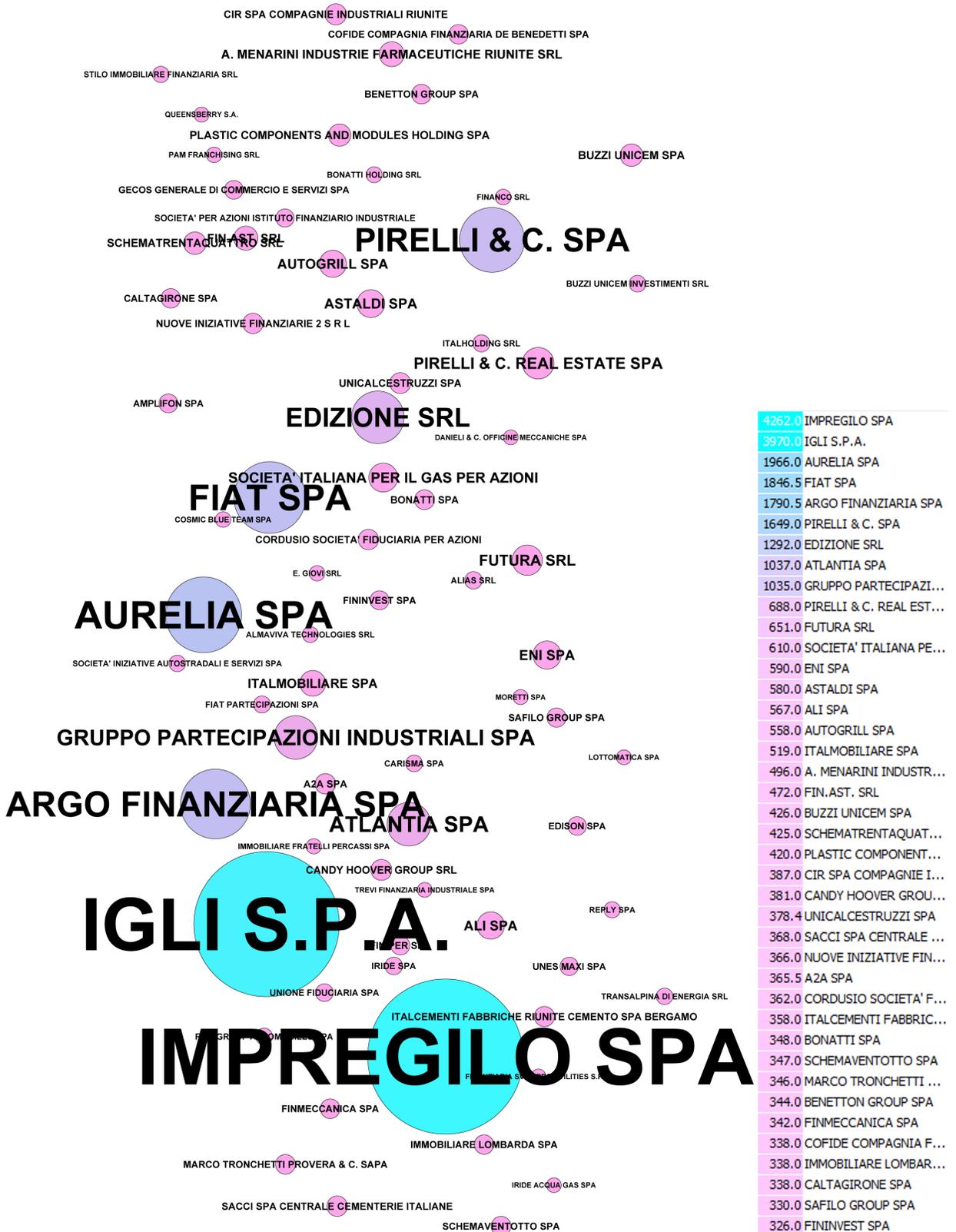


Figura 7.3.4: Betwiness centrality (rete di influenza)

nel 2007. L'assetto azionario di maggioranza ("*Igli S.p.A*") è composto da un patto di sindacato in modo paritetico tra le società "*Argo Finanziaria S.p.A.*" (gruppo Gavio), "*Autostrade per l'Italia S.p.A.*" (gruppo Benetton), e da "*Immobiliare Lombarda S.p.A.*" (gruppo Ligresti). In figura 7.3.5 si rappresenta l'ego network, a 2 livelli di profondità per motivi di chiarezza; l'ego network riduce il dato visualizzabile, perchè essendo un nodo con betweenness alta, e quindi con molte diramazioni, aggiungere altri livelli porterebbe a visualizzare tutta la rete.

Dalla figura (in ordine di grandezza del nodo basata su betweenness centrality, e di colore in base alla modularità, per motivi di chiarezza) si nota come gran parte dei nodi con alta betweenness partecipano nella ego network "*Impregilo S.p.A.*", alcuni di questi nodi sono presenti anche nei primi posti della classificazione secondo degree centrality.

Da queste due osservazioni si può arrivare alla semplice conclusione che i nodi importanti, in termini di degree e betweenness centrality, alimentino a vicenda il proprio grado di centralità nei confronti della rete.

7.4 Eccentricità

Una definizione di eccentricità viene proposta da Hage and Harary in [22], secondo cui l'*eccentricità*, $e(v)$, di un nodo v in un grafo connesso G è la massima distanza $d(v, u)$ per tutte le u . Il diametro di un grafo G è l'eccentricità massima di un nodo, ad esempio la distanza massima tra due nodi di G .

7.4.1 Eccentricità come diametro delle partecipazioni

Nel nostro contesto, di grafo diretto, l'eccentricità rappresenta il diametro di partecipazione massima di un nodo, come si può vedere in figura 7.4.1 rappresentante i nodi in base al valore di eccentricità.

Da questa immagine si può comprendere la *complessità* del nostro contesto, derivante dalla difficoltà che si può avere nel ricondurre un nodo terminale al suo vero proprietario (se tutte le partecipazioni sono di maggioranza relativa o assoluta). Ad esempio, il primo nodo in classifica ("*Tronchetti Provera Marco*") possiede un valore di eccentricità di 9, questo vuol dire che dal primo nodo all'ultimo ci sono altre 7 partecipazioni, che equivalgono a

7 società diverse, questo meccanismo viene definito, in campo finanziario, "scatole cinesi".

Gruppi piramidali e scatole cinesi

Per aumentare il potere di controllo sul gruppo, il soggetto economico, a parità di mezzi investiti, cioè quindi per espandere le dimensioni del gruppo senza perdere il controllo, può utilizzare strumenti derivanti dallo sviluppo piramidale del gruppo (cosiddetto "pyramiding"). Il gruppo piramidale è composto da una holding a monte che, attraverso partecipazioni dirette e indirette (tramite sub-holding), *controlla a cascata* un numero, a volte elevato, di società operative. *Si separa così il controllo di un proprietario-controllante dalla proprietà degli azionisti non controllanti.* Questo tipo di struttura determina il fenomeno dei "controlli di minoranza" che consentono di diluire il capitale del gruppo; solo al vertice della piramide il capitale è concentrato nelle mani di pochi (il soggetto economico).

L'obiettivo di fondo nella creazione della struttura piramidale è quello di "separare il controllo dalla proprietà del capitale" (il cosiddetto "capitalismo senza capitali")[15], attraverso la costituzione di "scatole cinesi", cioè società in cui l'attivo è rappresentato in misura prevalente dalla partecipazione, di solito di controllo, in altre società quotate, mentre nelle fonti di finanziamento una quota consistente è costituita dal capitale di proprietà conferito da minoranze azionarie.

La loro redditività è associata all'incasso dei dividendi prodotti dalle società operative agli ultimi posti della gerarchia. Quanto più è elevato il numero di scatole cinesi tanto maggiore è la dispersione dei dividendi, e da qui la tecnica da parte degli azionisti controllanti, di sfruttare una serie di canali attraverso i quali estrarre benefici privati, a danno, soprattutto, degli azionisti di minoranza delle controllate a cascata.

Il caso Marco Tronchetti Provera & C. Sapa: "Marco Tronchetti Provera & C. Sapa" è la società cassaforte della famiglia Tronchetti Provera, una società in accomandita per azioni che ne contiene il patrimonio. La società possiede (61%) di "Gruppo Partecipazioni Industriali S.p.A." e, attraverso la holding controlla "Cam Finanziaria S.p.A." con il 52,09%, la quale a sua volta detiene la maggioranza relativa (25,51%) delle azioni di "Pirelli & C. S.p.A.". A questo punto ogni partecipazione di "Pirelli & C. S.p.A." se di maggioranza relativa o assoluta sarà,

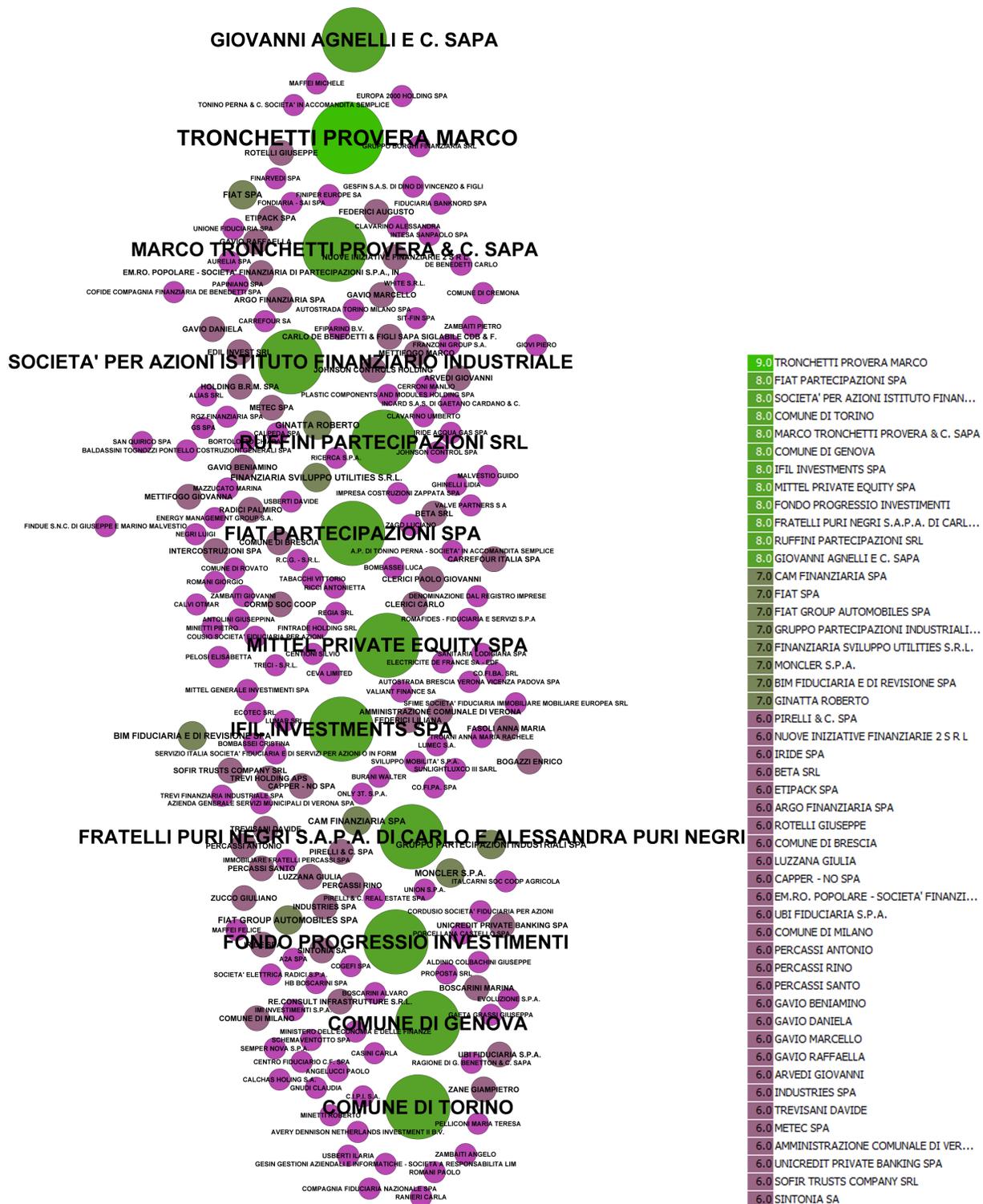


Figura 7.4.1: Eccentricità

ricorsivamente del nodo iniziale. In sostanza, la "Marco Tronchetti Provera & C. Sapa" è una società che fa capo a una complessa struttura piramidale, che include un centinaio di società industriali e finanziarie (holding e sub-holding). È grazie a questa struttura di "partecipazioni a cascata" che Marco Tronchetti Provera è in grado di controllare Pirelli e, in precedenza, anche Telecom Italia (figura 7.4.2).

L'immagine è in ordine di grandezza, rappresentante l'eccentricità, e in scala di colore per modularità, per motivi di chiarezza nella rappresentazione.

La misura di eccentricità è quindi utile nell'individuare i possibili meccanismi di "scatole cinesi" all'interno della rete, ovviamente non tutte le reti di partecipazione con diametro alto saranno individuate con tale meccanismo. Una volta calcolato il diametro è facile, utilizzando il modello proposto, risalire ai nodi che formano questa rete di partecipazione.

7.5 Closeness centrality

La misura di centralità basata su degree rappresenta solamente le relazioni dirette che un nodo possiede nei confronti degli altri nodi a lui vicini, non mostrando quindi le relazioni che si possono svolgere tra nodi a livello di profondità successivi. Ai nostri fini questa misura non è dunque sufficiente.

Nei grafi connessi la diffusione di informazioni, in generale, è modellata dall'uso dei cammini minimi. L'approccio alla closeness centrality enfatizza la distanza di un attore verso tutti gli altri presenti nella rete, focalizzandosi sulla distanza geodetica di ogni attore verso tutti gli altri.

Si può definire la closeness centrality come "la distanza media da un nodo di partenza verso tutti gli altri nodi nella rete". Introduciamo un esempio che servirà a spiegare meglio il concetto di closeness in un grafo diretto (figura 7.5.1).

Mentre in un grafo non diretto i nodi con closeness più alta sarebbero b e c , intesi come i nodi con la vicinanza maggiore all'interno del grafo, per quanto riguarda una rete diretta, un nodo acquisisce, e somma, il grado di closeness da tutti i nodi a cui è connesso (tramite out-degree ovviamente). Questo nodo può rappresentare il soggetto generatore della rete di relazioni.

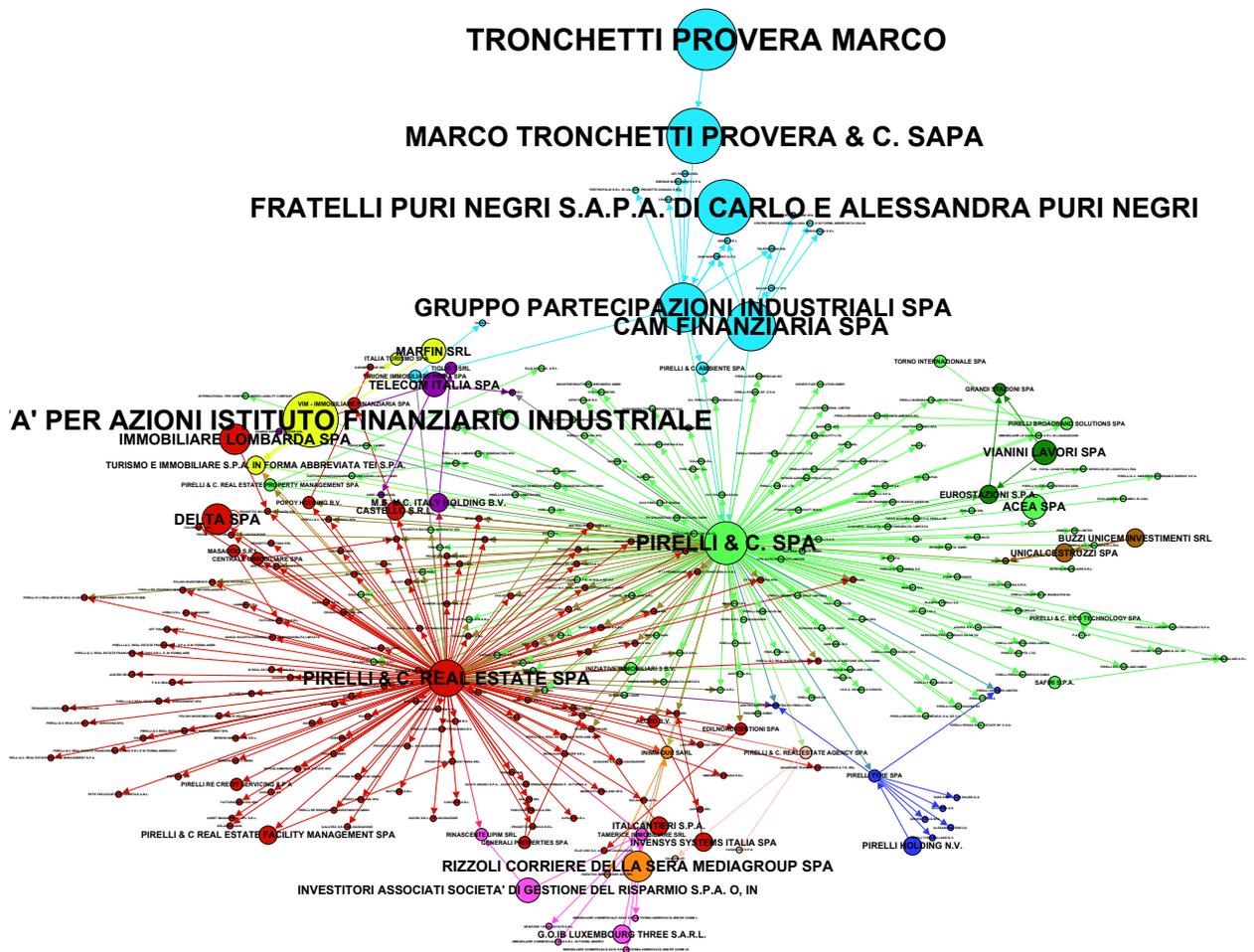


Figura 7.4.2: Ego network "Marco Tronchetti Provera & C. Sapa"

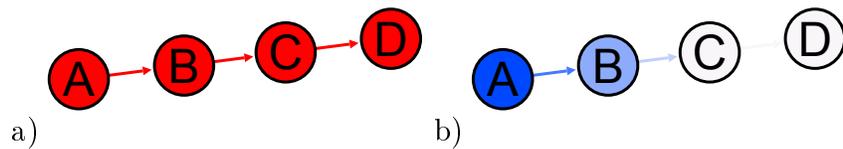


Figura 7.5.1: Esempio di rete lineare. a) senza misura di closeness; b) con misura di closeness

Per capire al meglio questa misura di centralità è necessario presentare le differenze tra la closeness centrality e la betweenness centrality presentata precedentemente.

La misura di betweenness centrality, offre un calcolo della centralità in base alla figura da "mediatore" di un nodo. Questa tecnica si rivela fondamentale nel caso in cui si vogliono studiare i flussi di informazione all'interno di un grafo, per cui un nodo accede ad un flusso di informazioni solo esclusivamente con la mera presenza di una relazione, senza basarsi sulla percentuale di partecipazione al nodo con cui si collega.

Le relazioni di partecipazione azionaria rappresentano un insieme di gerarchie sociali (connesse e non), a livello macroeconomico, in cui ogni relazione è asimmetrica e indica una forma di controllo nei confronti di un altro.

Se all'interno di una rete, si volessero studiare eventuali rapporti di dipendenza, la sola relazione non sarà sufficiente a descrivere tale comportamento. La misura di betweenness centrality calcolata sulla dicotomizzazione al 20% può essere interpretata come la capacità di un determinato nodo di influire direttamente nei confronti della rete, data la sua posizione centrale, ma, tale nodo rappresenta solitamente una holding finanziaria e avrà necessariamente, come si è visto nei gruppi gerarchici presentati in precedenza, delle relazioni entranti (in-degree), cioè di dipendenza. La betweenness centrality non può perciò rappresentare rapporti di controllo diretto all'interno di un grafo, perchè non è una misura transitiva.

Prendiamo, ad esempio, la figura a) dell'esempio mostrato in alto e associamo ad ogni relazione un rapporto di controllo diretto. Se utilizzassimo la misura di betweenness centrality, i nodi con il valore di centralità maggiore sarebbero "b" e "c" anche se, nel nostro contesto, dipendono direttamente da "a". La misura di betweenness centrality non varia da grafo diretto a non diretto.

Per studiare la "rete di potere" di un nodo si utilizzerà quindi la misura di closeness centrality che, come mostrato in figura b), gode di proprietà transitiva e di direzionalità del legame. I nodi con il maggior valore di closeness centrality saranno i nodi con la rete di potere maggiormente estesa e, si può anche ipotizzare che siano, di conseguenza, i nodi con il potere maggiore all'interno della rete.

7.5.1 Closeness centrality come misura di potere

Dal punto di vista del nostro dataset, e quindi una rete di partecipazioni azionarie rilevanti al 20%, la closeness centrality sarà interpretata come:

- la rappresentazione delle "cassaforti di famiglia" delle holding finanziarie, cioè i nodi che generano una rete di legami a cascata successivi;
- la misura dell'ampiezza di una rete di potere.

Tornando all'esempio precedente, è facile interpretare, nel nostro contesto, la figura *b*) in questo modo: il nodo *a* rappresenta la cassaforte di famiglia e gli altri nodi rappresentano i gruppi industriali, le sub-holding e le società operativa facenti capo ad *a*.

Vediamo i risultati che emergono dallo studio della closeness centrality nel nostro dataset (figura 7.5.2).

Si nota immediatamente come molti dei soggetti ai primi posti nella centralità basata su degree in questa classifica non sono presenti, o almeno non nella stessa posizione. Il numero di società ai primi posti della classifica sia molto basso, e questo si spiega riprendendo l'esempio di rete precedente: utilizzando la closeness centrality diretta si arriva al soggetto generatore della rete, colui che fa partire tutto; nel nostro caso quindi, indica il soggetto che detiene il controllo delle società a cui è diretto e, di conseguenza, influisce su tutti i legami successivi.

E' possibile ipotizzare perciò che questa rappresentazione indichi, al netto di eventuali falsi positivi, la rete di potere (o rete di controllo) di un soggetto, o famiglia. In questa immagine, infatti, sono presenti buona parte dei soggetti, e delle famiglie, più importanti del panorama finanziario italiano.

La misura di closeness centrality, al netto dei risultati, sembra essere la misura più appropriata per studiare il vero e proprio "potere" di un soggetto all'interno di una rete, le decisioni delle società dipendono direttamente, da questi soggetti. E' importante notare come molti di questi soggetti abbiano, molto spesso, relazioni indirette comuni in altre società; questo lo si può dedurre dal fatto che il diametro delle partecipazioni sia molto ampio, e che quindi ci sia una buona probabilità di incrociarsi con altri soggetti con closeness alta, derivante dalla legge 80/20 di Pareto.

Il caso Sintonia S.A.: Sintonia S.A. è la finanziaria controllata dalla holding *Edizione* della famiglia Benetton che si occupa principalmente di infrastrutture e servizi per la mobilità e la comunicazione. Vediamo

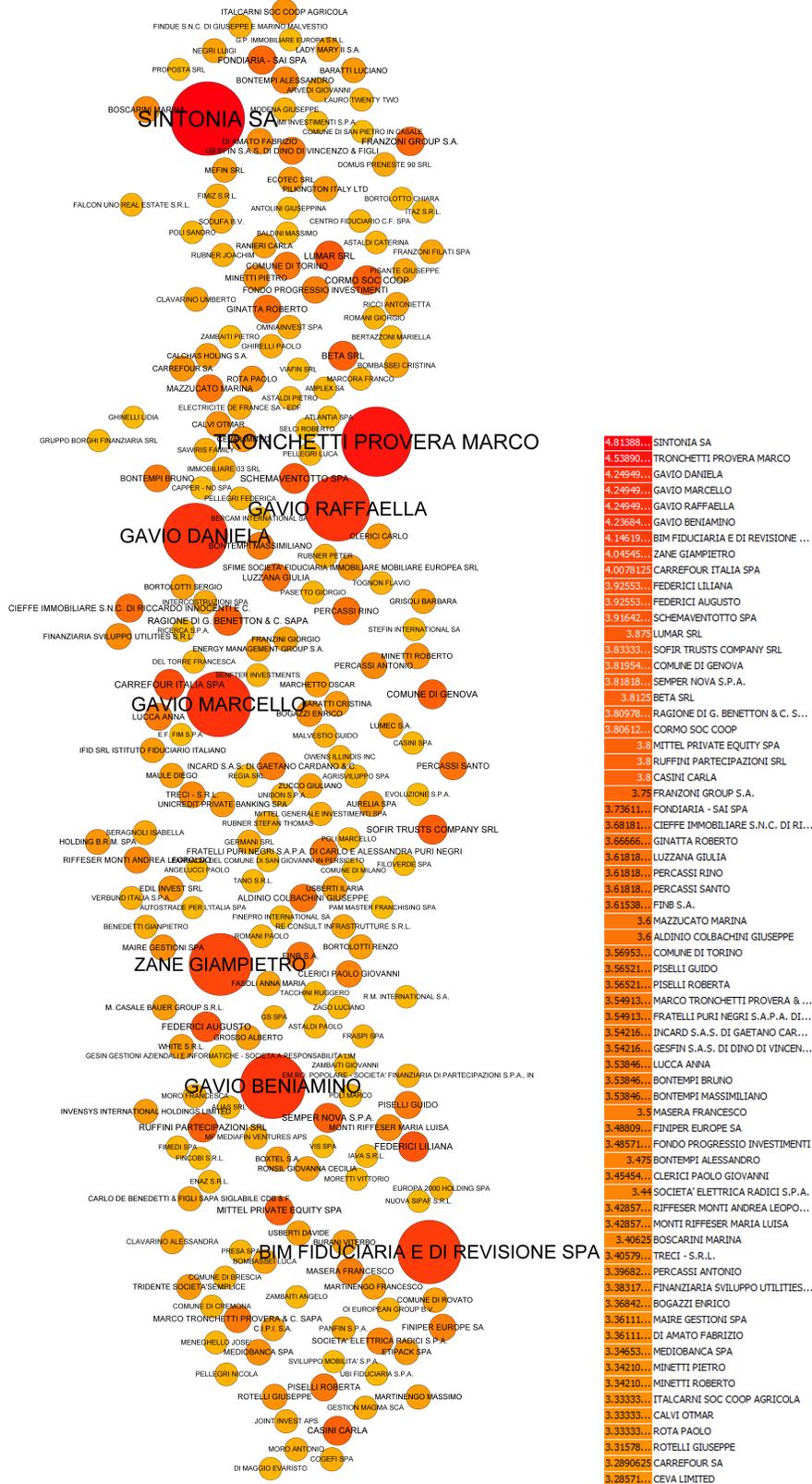


Figura 7.5.2: Closeness centrality (rete di potere)

tramite l'utilizzo di una ego network a 3 livelli di profondità, che ci servirà per rendere chiara la rappresentazione e la rete di partecipazioni che rendono questo soggetto, teoricamente, il più "potente" della rete (figura 7.5.3). Si utilizza una ego network a 3 livelli di profondità, i nodi hanno grandezze differenti a seconda della closeness centrality assegnata ad ognuno di loro, i colori sono scelti, per motivi di comprensibilità, in base alla modularità (formazione di sottogruppi in base alla vicinanza). I settori di partecipazioni sono multipli, come ci si può aspettare dalle grandi holding finanziarie, e per ogni categoria di investimento è presente una sub-holding di riferimento ("*Benetton Group S.p.A.*", "*Autogrill S.p.A.*", ecc...). Il vero e proprio centro di questa rete è il nodo "*Edizione S.r.l.*", sempre holding finanziaria della famiglia Benetton¹ da cui partono la maggior parte delle partecipazioni. Nonostante l'utilizzo di 3 livelli di profondità nella ego network si può comunque vedere dove sono presenti punti di contatto tra "*Sintonia S.A.*" e i nodi con alta closeness centrality presentati in precedenza; vediamo alcuni:

- "*Sintonia S.A.*" e "*Comune di Torino*" partecipano in "*Società Azionario Gestione Aeroporto Torino S.p.A.*", società partecipata con azionariato di maggioranza detenuto dal Comune di Torino;
- "*Edizione Holding*", facente parte del gruppo "*Edizione S.r.l.*", partecipa in "*Il Gazzettino S.p.A.*" insieme a "*Caltagirone Editore S.p.A.*" e "*Caltagirone S.p.A.*";
- "*Benetton Group*" e "*Istituto Finanziario Industriale S.p.A.*", proprietari del gruppo "*Fiat S.p.A.*", partecipano in "*Consorzio Generazione Forme*";
- e molte altre...

Il caso Aurelia S.p.A.: Aurelia S.p.A. è una holding che lavora soprattutto nel settore dei trasporti e della logistica, fa capo alla famiglia Gavio che ne detiene la maggioranza assoluta delle azioni. Le partecipazioni principali di questa famiglia sono presenti nel settore autostradale e delle infrastrutture, la principale partecipazione riguarda "*Impregilo*"

¹nel 2007, anno che si sta descrivendo, "*Sintonia S.A.*" e "*Edizione S.r.l.*" erano due entità separate che, ora, sono riunite sotto il nome di "*Edizione S.p.A.*"

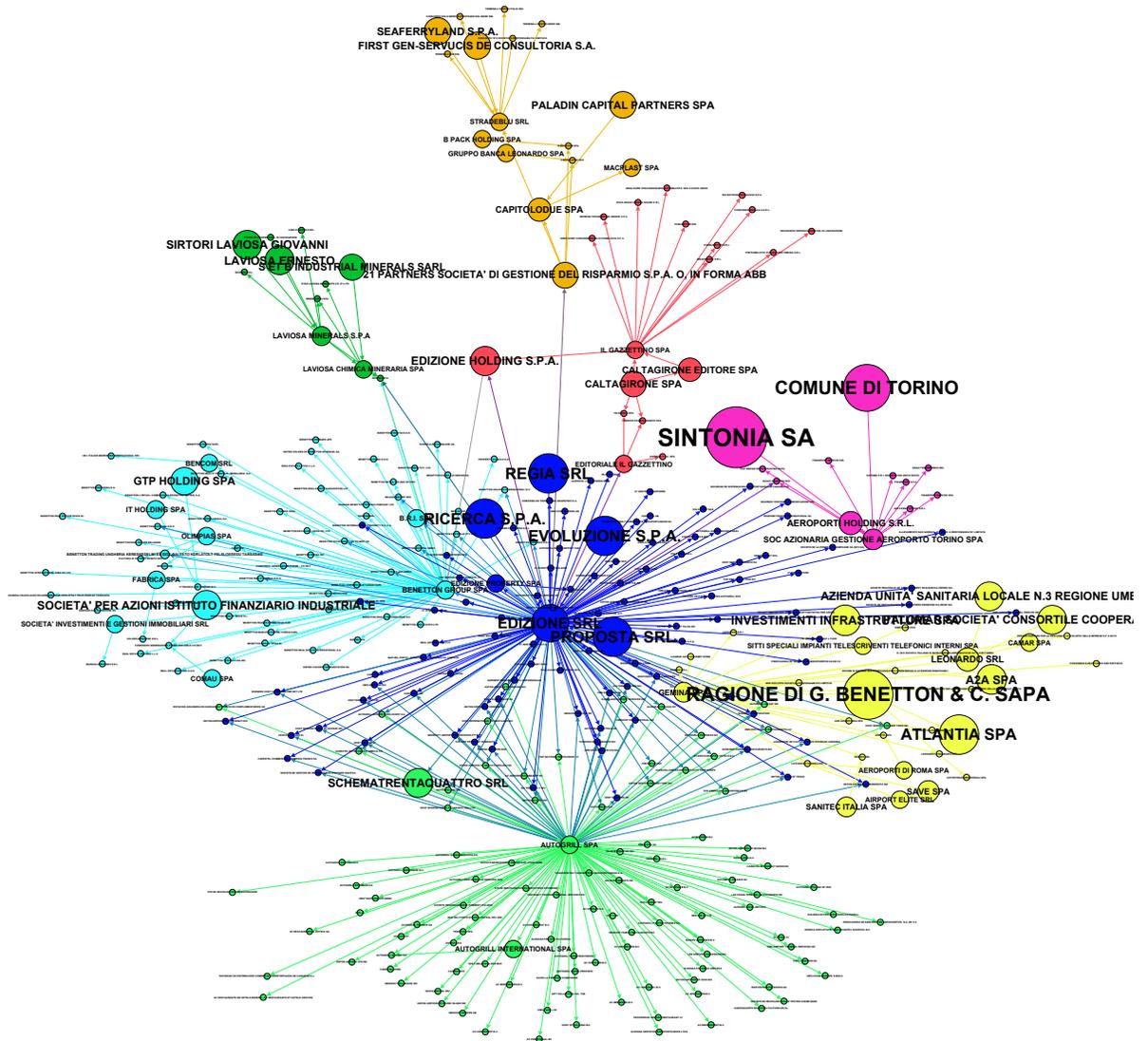


Figura 7.5.3: Ego network "Sintonia S.A."



Figura 7.5.4: Ego network "Aurelia S.p.A."

S.p.A.". In figura 7.5.4 si mostra la rete di partecipazioni della famiglia Gavio, all'interno di una ego network basata sul nodo "Aurelia S.p.A." a due livelli di profondità, per garantirne una maggiore comprensione; la grandezza dei nodi è in base alla closeness centrality, il colore in base alla modularità.

Le principali partecipazioni all'interno di "Impregilo S.p.A." sono detenute dalla società "Igli S.p.A." (società partecipata da "Autostrade per l'Italia S.p.A." (famiglia Benetton), "Argo Finanziaria S.p.A." (famiglia Gavio), e "Immobiliare Lombarda S.p.A." (famiglia Ligresti). Nella nostra rete "Impregilo S.p.A." è il nodo con il valore più alto di betweenness centrality, ad indicare che i nodi più potenti di un mercato possiedono nodi con alta betweenness.

Il caso Bim Fiduciaria e di Revisione S.p.A.: il nodo in questione partecipa all'interno del gruppo "*Carlo Benedetti & Figli Sapa*" facente parte della famiglia De Benedetti. Il gruppo De Benedetti si occupa soprattutto di editoria, industria ed energia. Vedendo l'immagine 7.5.5 si nota che la rete è più chiara rispetto alle precedenti, divisa in blocchi societari non troppo complicati da interpretare, rispetto ai precedenti, e, soprattutto, non presenta connessioni con altri nodi dotati di alta closeness centrality.

Il nodo rappresenta un sistema a sè, che può essere ricollegato ad altri nodi "potenti" della rete, ma solo aggiungendo molti livelli di profondità.

7.6 Risultati finali

Dai dati emersi da queste ego network, così come i dati emersi dalle misure di centralità precedenti, si può arrivare ad una serie di conclusioni sulla relazione tra centralità e potere:

- la centralità basata su out-degree mostra il coinvolgimento di un soggetto all'interno della rete;
- la centralità basata su out-degree aiuta nell'individuazione dei maggiori gruppi industriali, chi ha un degree maggiore, nel nostro caso, rappresenta un gruppo industriale maggiore. La natura del valore dell'out-degree può dipendere sia da una scelta organizzativa (dove la proprietà, o il soggetto di controllo, ha deciso di articolare il gruppo in una pluralità di soggetti specializzati, tutti controllati, disposti secondo una logica a stella) sia un possibile modello di alleanza fra diversi soggetti (ognuno con la autonomia strategica), che mantengono un nodo comune, di natura coalizionale;
- la betweenness centrality è la misura più sensibile alla variazione della topologia, esprimendo interdipendenza sistemica fra le sue componenti;
- la probabilità di trovare partecipazioni di grandi nodi industriali all'interno di un nodo con alta betweenness è molto frequente;

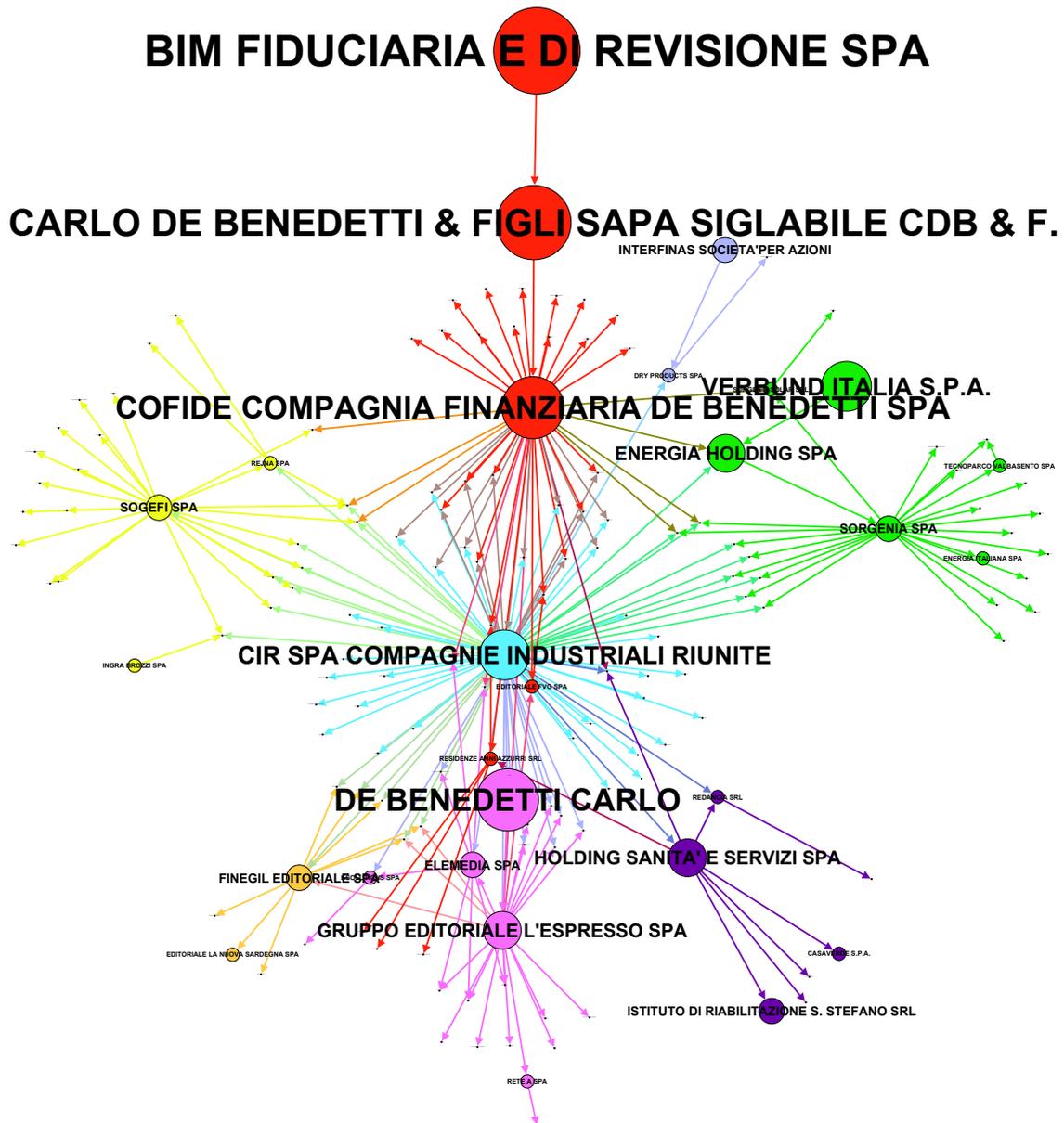


Figura 7.5.5: Ego network "Bim Fiduciaria e di Revisione S.p.A."

- i nodi bancari ed assicurativi svolgono la funzione di intermediari nella rete. Tramite questi nodi si ha la possibilità di avere una componente gigante nel grafo, sono i veri e propri "local bridge" della rete, come definito da Granovetter [20].
- un nodo con betweenness alta, nel dataset dicotomizzato, tramite la sua densa rete di relazioni e partecipazioni, può essere definito come avente maggiori possibilità di influenzare il sistema;
- un nodo con eccentricità alta avrà una rete di partecipazioni con un diametro alto;
- la centralità basata su eccentricità può dimostrare il meccanismo delle "scatole cinesi", risalendo al diametro del nodo si può riconoscere fino a che punto si spinge tale meccanismo;
- la closeness centrality è la misura più appropriata per definire il "potere" all'interno del grafo, rappresenta il nodo generatore della rete di relazioni, nel nostro caso la "cassaforte di famiglia";
- i nodi con alta closeness centrality hanno probabilità maggiori di partecipare insieme all'interno della rete, e quindi di alimentare a vicenda il loro valore di centralità;
- I nodi con alta closeness spesso sono i controllanti di nodi con alta betweenness.

Tornando alla categorizzazione di Freeman [19], si trova una relazione tra le misure di centralità basate su betweenness e closeness secondo cui, diminuendo la dipendenza di un soggetto da altri corrisponde ad aumentare l'accesso indipendente alle risorse, mentre aumentare la dipendenza di altri da sé ne aumenta il controllo.

In figura si mostra un riassunto dei principali gruppi di "potere" rilevati dalle nostre analisi. La dimensione rappresenta il grado di closeness centrality. La figura rappresenta alcuni tra i nodi con il maggior valore di closeness centralità, con relative relazioni, da questa rappresentazione si può aggiungere alla definizione di distribuzione 80/20 di Pareto secondo cui "rich get richer", la definizione "powerful get more powerful", i potenti alimentano il loro potere a vicenda.

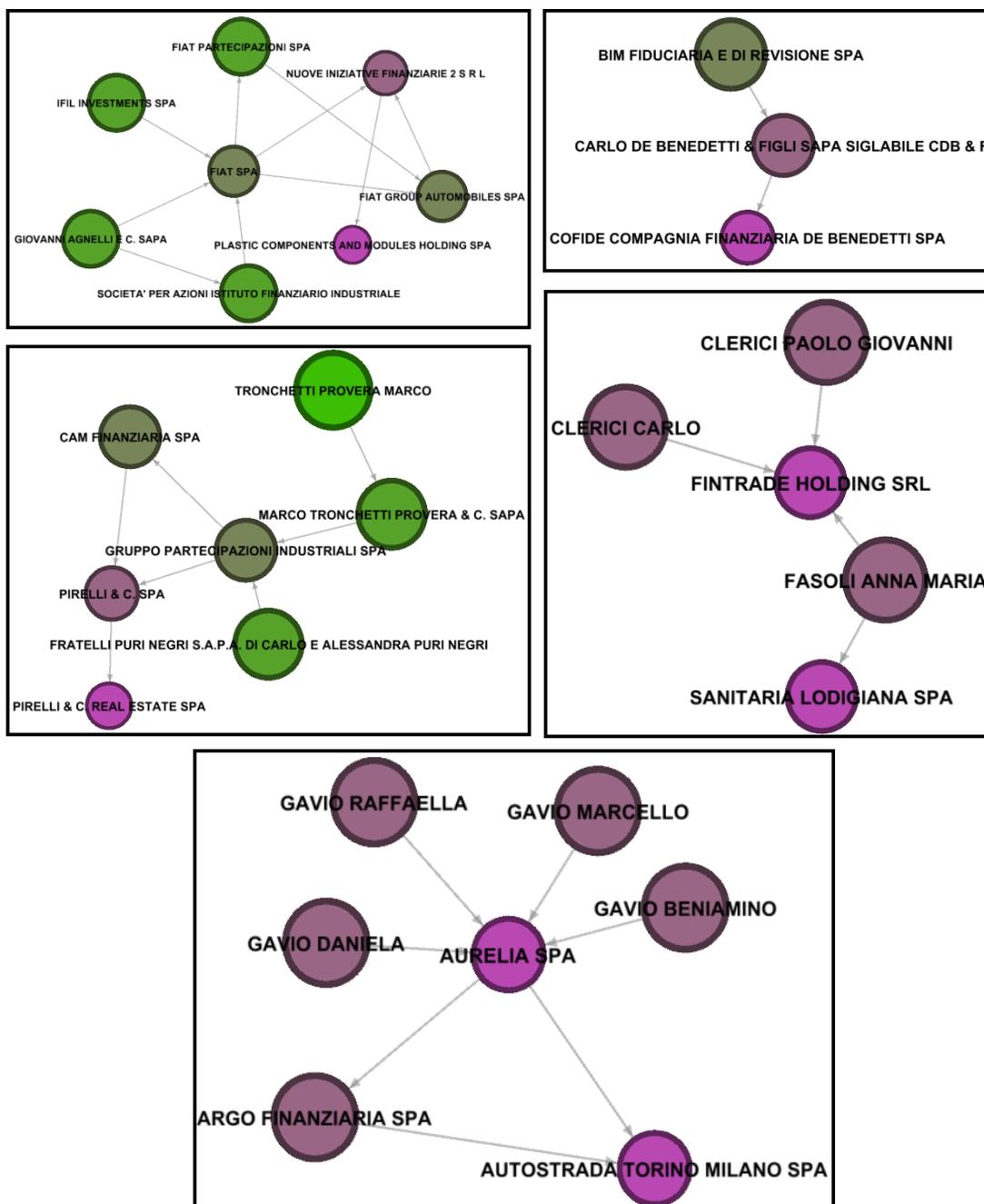


Figura 7.6.1: Riassunto dei principali gruppi di potere

Capitolo 8

Conclusioni e lavori futuri

Il mio lavoro di tesi si è principalmente concentrato sullo studio della distribuzione in reti complesse rappresentanti una porzione del mondo reale, nel nostro caso una parte delle partecipazioni azionarie di società per azioni. Per studiare la distribuzione si è deciso di utilizzare gli strumenti offerti dalla analisi delle reti sociali e quindi di trasformare il dataset in formato reticolare, cioè in un insieme di nodi e relazioni.

Si è proceduto alla costruzione di un modello in cui dai calcoli delle metriche e attributi derivanti dalle relazioni del grafo, si è costruita una vertex-list con l'aggiunta di dati finanziari, contenuti nel dataset originale, ai nodi del nuovo dataset. Il risultato ottenuto è dunque un modello che contiene tutte le informazioni riguardanti la rete sociale con l'aggiunta dei dati finanziari riguardanti i nodi.

I dataset sono di due tipi, per raccogliere tutte le informazioni disponibili dalle relazioni:

1. Dataset composto da edge-list ($vertice_1, vertice_2$) con relazione $vertice_1$ partecipa in $vertice_2$ quando verrà utilizzato come grafo diretto;
2. Dataset composto da edge-list ($vertice_1, vertice_2, \%_{di\ partecipazione}$) con relazione $vertice_1$ partecipa in $vertice_2$ con $\%_{di\ partecipazione}$ quando verrà utilizzato come grafo diretto;
3. Dataset 2 dicotomizzato al 20% di partecipazione.

Per ogni dataset viene studiata la distribuzione su:

1. dataset;

2. componente gigante;
3. prime 10 componenti connesse.

Dal calcolo dei valori riguardanti il coefficiente di clustering e la lunghezza del cammino medio si ottengono risultati discordanti rispetto ai valori, definiti in letteratura, riguardanti la distribuzione power-law scale-free su dati empirici: il coefficiente di clustering è sempre basso, mentre il valore della lunghezza media del cammino è, in alcuni casi, alto.

Dall'analisi della distribuzione si rileva che questi dataset seguono un comportamento basato sull'invarianza di scala, definito come power law. Il risultato era atteso e ampiamente descritto in letteratura, in cui si indica che molte reti complesse rappresentanti il mondo reale seguono una power law.

Molte delle componenti, comprese le componenti giganti, non seguono una power law pura, bensì una power law con un livello di cut-off, che viene, di volta in volta, indicato in tabella. In letteratura sono pochi i casi in cui dati empirici seguono una power law pura.

Da questi modelli, cognitivamente più semplici, si applicano le misure di centralità che permettono di estrarre della conoscenza, che con altri strumenti sarebbe difficilmente realizzabile. Le misure di centralità applicate al dataset 2 sono le seguenti:

Senza dicotomizzazione:

- betweenness centrality.

Con dicotomizzazione:

- degree centrality;
- betweenness centrality;
- eccentricity;
- closeness centrality.

Da queste misure si ottengono varie informazioni e si ipotizza che questi dati rappresentino misure di coinvolgimento, influenza, diametro delle partecipazioni, intermediazione, e potere all'interno della rete.

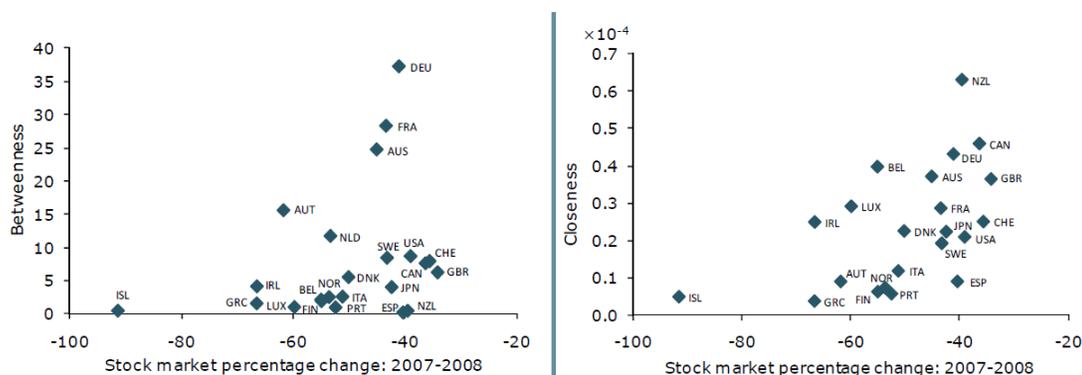


Figura 8.0.1: Paesi con betweenness e closeness media più alta hanno sofferto meno durante la crisi economica[11]

Nel mio lavoro propongo la misura di closeness centrality, su dataset dicotomizzato, come la più accurata per rappresentare un'eventuale misura di potere all'interno di un dataset contenente partecipazioni azionarie.

Dallo studio della betweenness, nel dataset 2 non dicotomizzato, si studiano i nodi con la maggiore probabilità di fungere da intermediari all'interno del grafo. Caballero, Candelaria e Hale in [11] propongono un'interessante analisi sulla crisi economica, dal lato delle banche, che inizia nel 2007-2008 (il dataset rappresenta l'anno 2007), e soprattutto, analizzano la relazione tra stabilità durante la crisi economica e il grado di betweenness tra di loro. La ricerca suggerisce che, analizzando i dati aggregati a livello di nazione, la misura in cui la banca di un determinato paese ha svolto un ruolo più o meno di intermediazione, è stata un indicatore della severità della crisi economica (figura 8.0.1). Lo studio si basa sulle fluttuazioni finanziarie e al grado di betweenness e closeness di queste banche.

Nel nostro contesto, si trova che i nodi con il maggior grado di betweenness sono le società bancarie e assicurative, che agiscono nel mercato con numerose partecipazioni, soprattutto di percentuale bassa. Oltre al ruolo da intermediari, questi nodi svolgono anche il ruolo che Granovetter [20] definisce di "local bridge", cioè di ponte locale tra gruppi. Se si escludono, all'interno della componente gigante, questi nodi, ad esempio applicando una dicotomizzazione al 20%, si nota che la rete si scompone. L'interpretazione data dall'analisi della betweenness e la closeness, e lo studio successivo all'interno del nostro dataset possono portare ad ottenere una possibile dimostrazione di come i sistemi bancari e assicurativi siano riusciti a *mantenere meglio la*

stabilità rispetto ad altri paesi, ad esempio Grecia e Islanda.

Dal modello proposto in questo lavoro di tesi è possibile estrarre una quantità di informazione elevata. Data la semplicità si riduce la difficoltà nell'analizzare graficamente i dati, e applicare calcoli complessi in tempo reale.

Il dataset rappresenta una porzione significativa delle società per azioni italiane, quotate e non, dell'anno 2007, e quindi pre-crisi economica. Un futuro lavoro potrebbe studiare, utilizzando lo stesso modello, e quindi con tempi molto rapidi, come cambia la distribuzione. E' legittimo aspettarsi che un dataset rappresentante la stessa porzione di mercato nel 2014 seguirà sempre una power law, come si può intuire dalla letteratura.

Potrebbe essere interessante studiare come cambiano i risultati delle misure di centralità 7 anni dopo, e con l'avvento della crisi economica del 2008.

Un altro perfezionamento del modello riguarda la dicotomizzazione del modello basandosi sulla leva azionaria di ogni nodo, e quindi data dal rapporto tra la quota azionaria (di capitale sociale), detenuta da una società capogruppo, nei confronti delle società operative (controllate) alla base della piramide delle partecipazioni societarie e il possesso integrato. La leva azionario rappresenta il rapporto tra il ROE (*"Return on Equity"*) della capogruppo ed il ROE del gruppo sottostante (partecipato tramite quote azionarie), ed esprime la capacità della capogruppo di poter usufruire a suo vantaggio del capitale delle controllate.

Bibliografia

- [1] ASTLEY, W. G., AND FOMBRUN, C. J. Collective strategy: Social ecology of organizational environments. *The Academy of Management Review* 8, 4 (1983), pp. 576–587.
- [2] BAK, P., TANG, C., AND WIESENFELD, K. Self-organized criticality. *Physical Review A* 38, 1 (July 1988), 364–374.
- [3] BARABASI, A.-L. *Linked: How Everything Is Connected to Everything Else and What It Means*, reissue ed. Plume, Apr. 2003.
- [4] BARABASI, A.-L., AND ALBERT, R. Emergence of scaling in random networks. *Science* 286, 5439 (1999), 509–512.
- [5] BASTIAN, M., HEYMANN, S., AND JACOMY, M. Gephi: An open source software for exploring and manipulating networks. In *ICWSM (2009)*, E. Adar, M. Hurst, T. Finin, N. S. Glance, N. Nicolov, and B. L. Tseng, Eds., The AAAI Press.
- [6] BERTINETTI, G. *La Finanza dei grandi gruppi aziendali italiani*. Scritti di economia aziendale. EGEA, 1994.
- [7] BORGATTI, S. P., AND FOSTER, P. C. The network paradigm in organizational research: A review and typology. *Journal of Management* 29, 6 (December 2003), 991–1013.
- [8] BRASS, D. J. Being in the right place: A structural analysis of individual influence in an organization. *Administrative Science Quarterly* 29, 4 (1984), pp. 518–539.
- [9] BURT, R. *STRUCTURAL HOLES*. Harvard University Press, 1995.

- [10] BUTERA, F. *Il castello e la rete. Impresa, organizzazioni e professioni nell'Europa degli anni '90*. Tecnologia, organizzazione e persone. Franco Angeli, 2005.
- [11] CABALLERO, J., CANDELARIA, C., AND HALE, G. Bank relationships and the depth of the current economic crisis. *FRBSF Economic Letter*, dec14 (2009).
- [12] CLAUSET, A., SHALIZI, C. R., AND NEWMAN, M. E. J. Power-law distributions in empirical data. *SIAM Rev.* 51, 4 (Nov. 2009), 661–703.
- [13] DAVID, E., AND JON, K. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.
- [14] D'ERRICO, M., GRASSI, R., STEFANI, S., AND TORRIERO, A. Shareholding networks and centrality: An application to the italian financial market. In *Networks, Topology and Dynamics*, A. Naimzada, S. Stefani, and A. Torriero, Eds., vol. 613 of *Lecture Notes in Economics and Mathematical Systems*. Springer Berlin Heidelberg, 2009, pp. 215–228.
- [15] DI CARLO, E. *I gruppi aziendali tra economia e diritto*. Univ.Tor Vergata-Dip. studi impresa.Quad. Giappichelli, 2009.
- [16] ERDOS, P., AND RENYI, A. On the evolution of random graphs. *Publ. Math. Inst. Hungary. Acad. Sci.* 5 (1960), 17–61.
- [17] ESTOUP, J. B. *Gammes Sténographiques: méthode & exercices pour l'acquisition de la vitesse (French) [Stenographic ranges: method and exercises for achieving speed]*, fourth ed. Institut Sténographique de France, Paris, France, 1916.
- [18] FREEMAN, L. *The Development Of Social Network Analysis: A Study In The Sociology Of Science*. Empirical Press, BookSurge, 2004.
- [19] FREEMAN, L. C. A set of measures of centrality based on betweenness. *Sociometry* 40, 1 (1977), pp. 35–41.
- [20] GRANOVETTER, M. The Strength of Weak Ties. *The American Journal of Sociology* 78, 6 (1973), 1360–1380.

- [21] HACKETT, A. *70 Years of Best Sellers: 1895-1965*. R. R. Bowker Co., 1967.
- [22] HAGE, P., AND HARARY, F. Eccentricity and centrality in networks. *Social Networks* 17, 1 (1995), 57 – 63.
- [23] HANNEMAN, R. A., AND RIDDLE, M. *Introduction to social network methods*. University of California, Riverside, Riverside, CA, 2005.
- [24] HILL, B. M. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* 3, 5 (1975), pp. 1163–1174.
- [25] HUBERMAN, B., AND ADAMIC, L. Information dynamics in the networked world. In *Complex Networks*, E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, Eds., vol. 650. Springer, 2004, pp. 371–398.
- [26] HUBERMAN, B. A., AND ADAMIC, L. A. The nature of markets in the world wide web. *Quarterly Journal of Economic Commerce* (Jan. 2000). Power laws & long tails in web merchants.
- [27] ISTAT. Classificazione delle attività economiche ateco 2007, derivata dalla nace rev. 2.
- [28] LOFGREN, K. G., PERSSON, T., AND WEIBULL, J. W. Markets with asymmetric information: The contributions of george akerlof, michael spence and joseph stiglitz. *Scandinavian Journal of Economics* 104, 2 (2002), 195–211.
- [29] MITCHELL, M. Complex systems: Network thinking. *Artificial Intelligence* 170 (2006), 1194–1212.
- [30] MORENO, J. *Sociometry, experimental method and the science of society: an approach to a new political orientation*. Sociometry monographs. Beacon House, 1951.
- [31] NEWMAN, M. E. J. Power laws, pareto distributions and zipf’s law. *Contemporary Physics* 46 (December 2005), 323–351.
- [32] NEWMAN, M. E. J., BARABÁSI, A. L., AND WATTS, D. J., Eds. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

- [33] NIEMINEN, U. On the centrality in a directed graph. *Social Science Research* 2, 4 (1973), 371 – 378.
- [34] PENROSE, E. *The Theory of the Growth of the Firm*. Oxford scholarship online. Oxford University Press, 1995.
- [35] SALVINI, A. *L'analisi delle reti sociali: risorse e meccanismi*. Didattica e ricerca: Manuali. PLUS-Pisa University Press, 2005.
- [36] SODA, G. *Reti tra imprese. Modelli e prospettive per una teoria del coordinamento tra imprese*. Biblioteca di testi e studi. Carocci, 1998.
- [37] TARJAN, R. Depth first search and linear graph algorithms. *SIAM Journal on Computing* (1972).
- [38] WATTS, D., AND STROGATZ, S. Collective dynamics of 'small-world' networks. *Nature*, 393 (1998), 440–442.
- [39] WATTS, D. J. *Six Degrees: The Science of a Connected Age*. W. W. Norton & Company, February 2004.