



Università degli studi della Calabria

Dottorato di ricerca in Psicologia della Programmazione
e Intelligenza Artificiale

Scuola di Dottorato "Archimede" in Scienze, Comunicazione e Tecnologie

Tesi Di Dottorato

Modellazione di linguaggi naturali e artificiali attraverso la scienza delle Reti

Per il Coordinatore del

Dottorato

Prof. ssa Eleonora Bilotta

Supervisore

Prof. Pietro Pantano

Candidata:

Dott.ssa Francesca Bertacchini

ANNO ACCADEMICO 2011-2012

Sommario

Introduzione	6
Capitolo 1. Aspetti teorici e formali dei linguaggi e delle lingue	12
1.1. Introduzione.....	12
1.2. Evoluzione della facoltà del linguaggio e sviluppo della cognizione negli umani	13
1.2.1. Il linguaggio è innato o appreso?	13
1.2.2. Né innato né appreso, ma tanti modelli cognitivi contribuiscono a determinarlo	16
1.3. La struttura organizzativa della lingua.....	18
1.4. Lingua e emozioni.....	24
1.5. Linguaggi formali	26
Capitolo 2. Aspetti formali delle reti complesse	31
2.1. Introduzione.....	31
2.2. Aspetti formali: Grafi e Reti	31
2.3. Grandezze Statistiche	35
2.4. Modelli di rete.....	37
2.5. Motivi.....	43
Capitolo 3. Modellazione della Musica	48
3.1. Introduzione.....	48
3.2. Modellazione Matematica delle Melodie Musicali	51
3.3. Simmetrie e complessità.....	55
3.4. Le reti melodiche	59
Capitolo 4. Modellazione di reti cerebrali.....	64
4.1. Introduzione.....	64
4.2. Strutture cerebrali	65
4.3. Circuiti cerebrali cognitivi	67
4.4. Modellazione matematica: dalla scansione di Risonanza Magnetica alle Reti Cerebrali	69
Capitolo 5. Modellazione delle lingue naturali e del significato	79
5.1. Background di riferimento sulle reti del linguaggio umano	79
5.2. Modellazione di WordNet come rete	83
5.3. Creazione di un vocabolario.....	84
5.4. Il vocabolario delle emozioni	89
5.5. Crescita, simmetrie e particelle emergenti in una rete semantica	91
5.6. Reti con nodi non omogenei e particelle emergenti	93

5.7.	I sistemi software sviluppati.....	100
Capitolo 6. Reti musicali in Bach.....		105
6.1.	Introduzione.....	105
6.2.	L'entropia nelle melodie di Bach.....	105
6.3.	Analisi delle reti.....	109
6.3.1.	Statistica a livello della singola traccia.....	111
6.3.2.	Strutture di Piccolo Mondo, a livello delle singole collezioni.....	113
6.3.3.	Reti con Invarianza di Scala.....	114
6.4.	Statistiche di rete a livello delle Collezioni.....	115
6.4.1.	Strutture di piccolo mondo.....	118
6.4.2.	Reti ad Invarianza di scala.....	119
6.5.	Statistiche di rete a livello dell'intera collezione di Bach.....	121
6.6.	Motivi e Grammatiche musicali emergenti.....	123
6.7.	Cicli e crescita delle reti musicali.....	128
6.8.	Conclusioni.....	132
Capitolo 7. Reti cerebrali in soggetti sani e con malattie neurodegenerative.....		133
7.1.	Introduzione.....	133
7.2.	Background di riferimento sul Parkinson e il Parkinsonismi.....	133
7.3.	Materiali e metodi.....	136
7.4.	Risultati.....	137
7.5.	Analisi dei risultati.....	138
7.5.1.	Differenze statistiche fra i gruppi.....	141
7.5.2.	Il grado (k) come strumento di analisi della degenerazione neurale.....	143
7.6.	Conclusioni.....	151
Capitolo 8. Crescita di reti emozionali nel sistema WordNet.....		154
8.1.	Introduzione.....	154
8.2.	Il contesto di riferimento psicolinguistico.....	156
8.3.	Risultati sperimentali per la rete WordNet.....	161
8.4.	Crescita dei termini emozionali, metodo di indagine.....	165
8.4.1.	Evoluzione e analisi qualitativa delle reti.....	166
8.4.2.	Statistiche di rete per i termini emozionali.....	178
8.5.	Particelle emergenti.....	183
8.5.1.	Grafi Perfetti.....	184
8.5.2.	Rotture di simmetrie.....	186
8.5.3.	Altre Simmetrie.....	187

8.5.4. Composizione di Particelle	188
8.6. Conclusioni	190
Conclusioni Generali	191
Bibliografia	196
Appendice: Particelle Emergenti	218

Introduzione

Così come altri fenomeni, un linguaggio (artificiale o biologico), non è un aggregato di elementi, senza alcuna relazione. Al contrario, come sostiene De Saussure (1916), una lingua naturale (e di conseguenza tutti i linguaggi che godono delle stesse caratteristiche strutturali), ha rapporti a diversi livelli gerarchici (fonetico-fonologico, morfologico-sintattico e semantico). Tali rapporti caratterizzano la significatività dei termini usati e, quindi, sono *portatori* di significato (*semiofori*). La lingua naturale è il sistema di gran lunga più complesso fra i linguaggi, in quanto incorpora una serie di elementi semantico-sintattici e contestuali, sociali e psicologici, molto difficili da comprendere fino in fondo. Anche la musica, essendo un linguaggio a tutti gli effetti, con le sue relazioni orizzontali e verticali, determina una serie di legami attraverso le leggi dell'armonia e della composizione. Lo stesso si può dire per il cervello biologico. Nelle Neuroscienze, è possibile pensare a diversi tipi di rapporti gerarchici tra le aree del cervello umano, dove le aree diventano i nodi della rete, mentre gli archi rappresentano le connessioni anatomiche. Il linguaggio delle connessioni fra aree fornisce l'insieme dei processi cognitivi della mente e dell'agire umani. Lo studio delle connessioni nelle lingue naturali, nel linguaggio della musica e nelle dinamiche cerebrali ci consente di capire queste relazioni, identificando le principali aggregazioni di gruppi di elementi, per svolgere determinate funzioni, parlare o scrivere testi, ascoltare o scrivere musica, intendere noi stessi e il mondo rispettivamente. Il principale obiettivo di questa tesi è quello di individuare e interpretare la natura di queste relazioni nei linguaggi, analizzando le interconnessioni fra differenti livelli. Nessuna ipotesi è fatta in modo aprioristico sulla natura di tali legami. L'impianto concettuale della Linguistica serve come base da cui partire, per avere un sistema strutturato su cui paragonare gli altri linguaggi. L'approccio è invece quello della Scienza della Complessità (Bedau, 2003; Bilotta & Pantano, 2010). Si analizzeranno l'emergere di relazioni esistenti fra gli elementi di un linguaggio, in modo da rilevare se, nonostante la diversità di mezzi attraverso i quali i linguaggi sono veicolati, essi abbiano strutture organizzative comuni. In altre parole, si cercano pattern condivisi fra i vari sistemi sotto esame, " *the fingerprints of complexity*" (Casti, 1994, 1996). Il problema di modellizzazione affrontato è il seguente. I linguaggi hanno un significato informativo che travalica il contenuto letterale dei termini. Tali significati sono rappresentati come nodi di una rete. Come si possono modellare i termini di un linguaggio cercando di estrarre tutte le informazioni possibili, per poi restituirle ad

una comprensione più profonda del sistema di significazione proprio di una lingua o, in generale, di un sistema di comunicazione? E' possibile associare altri attributi, di natura diversa, non strettamente inerenti il significato, nel caso della musica? E' possibile individuare il linguaggio delle funzioni nel caso del cervello biologico? Possono le relazioni di tipo pragmatico e socio-linguistico essere considerate insieme agli approcci più tradizionali per capire il funzionamento della lingua? Le reti complesse e le loro funzioni sono utilizzate a questo fine. File musicali, file provenienti da Risonanza Magnetica e campioni di lingua sono trasformati in reti, attraverso opportuna modellazione. Nei capitoli che seguono, le relazioni non solo strutturali ma anche funzionali dei nodi di reti musicali, cerebrali e linguistiche saranno analizzate. Dopo aver introdotto i principali concetti relativi alle reti complesse, si riportano gli aspetti formali di grafi e reti, le relative grandezze statistiche, i differenti modelli di rete che sono stati individuati e si specificano, come ulteriori strumenti di analisi delle reti i *motivi*, pattern strutturati di elementi che realizzano connessioni significative nei linguaggi esaminati. L'analisi e la comprensione di tali motivi sarà un metodo fondamentale per capire la struttura dei linguaggi.

Si ipotizza che i linguaggi musicali, le connessioni cerebrali, le lingue naturali siano rappresentabili come strutture di elementi, tra loro interrelati, con dinamiche di interazioni fra gli elementi stessi. La modellazione matematica e la Scienza delle Reti Complesse (Wilson et al., 1999; Appert-Rolland et al., 2011) permettono di analizzare tali fenomeni e processi di natura differente, arrivando ad una comprensione profonda delle leggi di organizzazione dei fenomeni biologici, fisici (Barabási, 2005a, b), cognitivi e artificiali (Ball et al., 2011). La principale caratteristica delle reti sono i legami tra gli elementi della rete, che individuano, attraverso opportuni metodi di analisi, le caratteristiche di milioni di connessioni di sistemi complessi come la rete Internet (Barabási et al., 2002,), le reti del DNA (Barabási, 2004) o le reti sociali (Granovetter, 1973; Milgram, 1967; Watts & Strogatz, 1998; Wasserman & Faust, 1994). Questi legami articolano la rete e forniscono evidenza sulla natura dei fenomeni che si stanno osservando. Un insieme di vertici (detti anche nodi) e di linee (dette anche legami) tra coppie di vertici è detto *grafo*, e rappresenta la struttura formale di una rete. Il vertice è l'unità minimale di una rete, una linea è invece un legame tra due vertici in una rete. Il legame considera il nodo o vertice che origina la connessione (emettitore del legame) e il nodo o vertice sul quale il legame si connette (ricevitore del legame). Questo avviene nel caso di grafi diretti, come si vedrà nei paragrafi successivi. Nei grafi indiretti, invece, il legame connette coppie non ordinate, senza direzione. Un loop è un tipo speciale di connessione, un legame che connette il vertice a se stesso. Si trovano di solito nei grafi diretti, e non nei grafi indiretti. Informazioni aggiuntive si scoprono nella rappresentazione della dimensione del legame, dimensione degli elementi, raffigurate di solito da spessori più pronunciati delle linee (dei legami) o della dimensione dei nodi. Di solito, tali caratteristiche indicano rispettivamente la forza di una connessione e l'importanza del nodo all'interno della rete. Il centro, la semiperiferia e la periferia, il grado di importanza di un nodo in questi intorni, il diametro di una rete sono statistiche importanti per determinare la struttura di una rete.

Altre caratteristiche riguardano speciali aggregazioni di nodi. Nella letteratura, le reti che possiedono tali organizzazioni sono dette di *piccolo mondo o Small World Networks* (SWN) (Watts e Strogatz, 1998), o reti che seguono una legge di potenza, a *invarianza di scala*, dette *Scale-Free* (SF) (Barabási, Albert e Jeong, 2000), in quanto i fenomeni organizzativi che si manifestano, sono indipendenti dalla scala, come si vedrà nei paragrafi successivi. Inoltre, le reti forniscono informazioni per modellare fenomeni come le crescite dei linguaggi. In modo particolare, sarà possibile ipotizzare una serie di modelli di crescita per i vari contesti analizzati. La modellazione consente l'individuazione di *motivi* (strutture organizzate di 3, 4, 5, ... n nodi connessi insieme), che emergono come simmetrie composite sia in ambito musicale, che neurologico e linguistico. Tali strutture presuppongono crescite di diverso tipo, aventi però alla base una grammatica generativa universale.

In questa tesi, sono descritti diversi tipi di reti, i modi attraverso i quali si possono analizzare numericamente gli ambiti oggetto di indagine e i sistemi di rappresentazione grafica per tali sistemi. Di particolare importanza è la formalizzazione matematica dei vari ambiti di applicazione, la trasformazione in rete per ogni ambito degli elementi da modellare, l'analisi delle reti, il ritorno alla realtà per comprenderla. Il processo realizzato è rappresentato in Figura 1.

La tesi si divide in due parti: una prima parte sulla modellizzazione matematica dei linguaggi della musica, delle dinamiche cerebrali e della lingua naturale e una parte sperimentale. Nella prima parte, dopo l'introduzione generale, il Capitolo 1 discute i vari problemi dei linguaggi e delle lingue, concetti alla base di tutta la trattazione, in quanto generali e generalizzabili, sia all'ambito musicale, sia ai processi cerebrali e, ancor di più, all'analisi delle lingue naturali. Il Capitolo 2, fornisce i concetti di base della formalizzazione matematica delle reti complesse. Numerosi problemi di modellizzazione sono stati affrontati e risolti in modo innovativo. Per esempio, nel Capitolo 3, file musicali in formato MIDI (il formato di interscambio che registra gli eventi musicali con un apposito codice) sono stati rappresentati con il codice MET. Tale codice si basa su un modello computazionale che trasforma il contenuto di file *.mid* (con problemi di individuazione non univoca degli eventi sonori, per cui agli stessi eventi possono correlarsi valori diversi per le note di uno spartito), in elementi matematici discreti, a diversi livelli di complessità. La modellazione matematica della musica ha consentito diversi tipi di rappresentazioni visive degli spazi delle melodie. Un successivo processo ha permesso di trasformare le composizioni musicali in grafi o reti. Per questo processo di modellizzazione, numerosi software di gestione della trasformazione delle composizioni musicali in reti sono stati sviluppati. Sono stati anche elaborati una serie di pacchetti software per l'analisi statistica delle reti, utilizzati per tutti gli ambiti di analisi che questa tesi affronta.

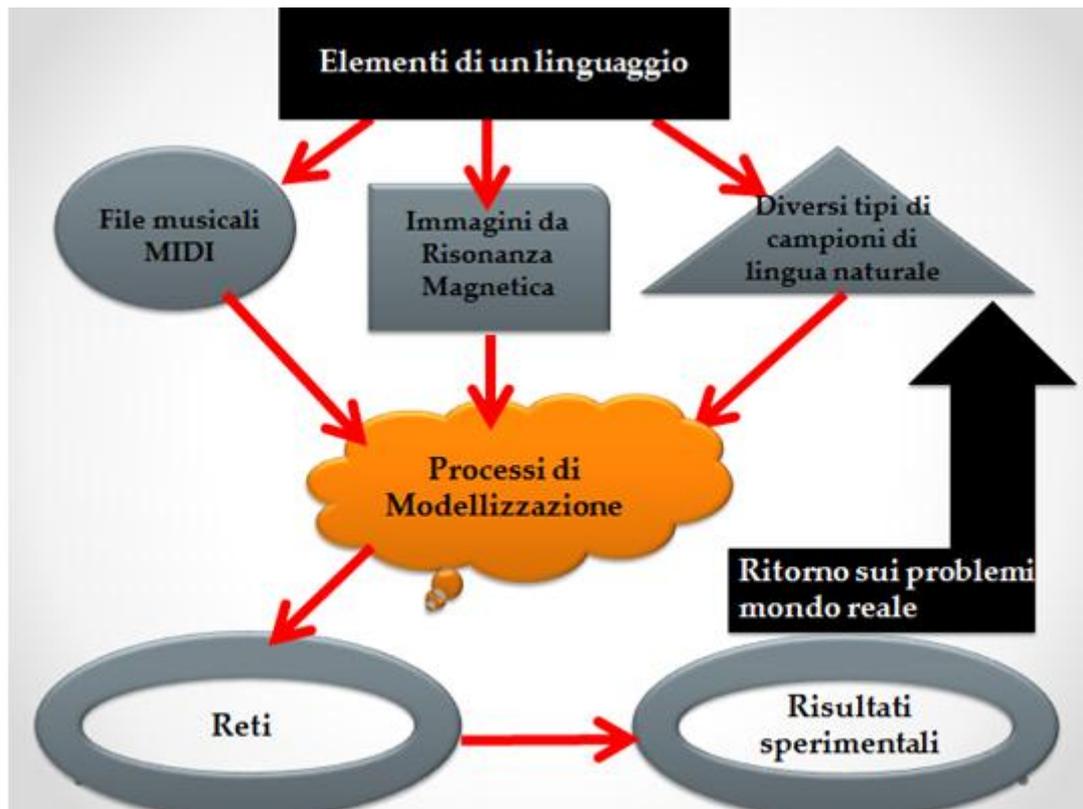


Figura 1. Vari tipi di linguaggi sono stati modellati matematicamente e trasformati in reti. Dopo l'analisi, i risultati sperimentali permettono di analizzare i problemi del mondo reale, con una maggiore comprensione dei fenomeni esaminati, per fornire nuove prospettive e soluzioni.

Per ottenere reti da scansioni cerebrali, Capitolo 4, è stata adottata una modellazione matematica più complessa. Immagini provenienti da Risonanza Magnetica Strutturale sono state trattate con i metodi della Connettistica, che combina la Risonanza Magnetica diffusiva e le metodologie della Trattografia per creare reti cerebrali. Per arrivare dalle scansioni di Immagini di Risonanza Magnetica alle reti, è necessario utilizzare file DICOM, (file di interscambio nella medicina). Il processo comporta un'articolata sequenza di passi:

- a. l'acquisizione di DTI, da Immagini di Risonanza Magnetica, attraverso T1 pesate ad alta risoluzione;
- b. la segmentazione della sostanza bianca del cervello;
- c. la Trattografia della Materia Bianca;
- d. la segmentazione in regioni di interesse (ROI) di tutto il cervello, nel file DICOM, l'output della Risonanza Magnetica; il processo è effettuato secondo atlanti differenti, che consentono di eseguire una segmentazione del cervello in molti modi diversi, ottenendo reti con un numero differente di nodi e connessioni.

Per realizzare i passi sopra indicati, sono stati utilizzati software open source. Il Capitolo 5, che chiude il processo di modellazione matematica, si occupa del complesso problema di come analizzare una lingua

naturale e, soprattutto di quali siano i metodi più idonei per formalizzare la complessità di questo fenomeno. Una qualsiasi lingua, infatti, oltre ad essere una struttura gerarchicamente organizzata in livelli, e con processi di interconnessione tra le parti, presenta il problema della semantica, che coinvolge fattori culturali, pragmatici, sociali ed evolutivi di enorme complessità. Nel capitolo sono forniti alcuni esempi di modellazione, soprattutto riferibili a WordNet, contenente una buona parte del lessico della lingua Inglese, costruito con un approccio psico-linguistico. Campioni di tale sistema sono stati analizzati e sono stati individuati i processi produttivi di crescita e di organizzazione delle simmetrie emergenti. Si vedrà nei capitoli dei risultati che il problema della crescita, unito alla differenza nella natura dei nodi (i nodi della rete non sono omogenei), rappresenta uno dei contributi innovativi e originali di questa tesi. Si profila, infatti, un nuovo modello di rete che porta anche a una profonda revisione dei meccanismi di attivazione di contenuti semantici in memoria. Anche per questo ambito di studio, vari pacchetti software, che permettono la trasformazione in reti di vasti campioni di linguaggio (testuale, da diversi database in lingua inglese, ma estendibili ad altre lingue) sono stati creati. I software incorporano diversi sistemi computazionali di interrogazione automatica, la creazione automatica di vocabolari finalizzati all'indagine di termini del lessico di una lingua, su specifici argomenti di interesse. La tesi analizza campioni di testo relativo alle parole emozionali nel database WordNet.

La seconda parte della tesi riguarda tre esperimenti, effettuati a valle del processo di modellazione e la trasformazione in reti dei linguaggi analizzati.

In particolare, nel Capitolo 6, è realizzato un piano sperimentale secondo il quale le composizioni di un repertorio di 241 melodie dell'opera di Bach, per un totale di 698, sono state trasformate da file MIDI in file MET, e poi trasformate in reti. Le reti sono state analizzate attraverso una funzione di entropia, per individuare una misura della quantità di informazione/disordine in ogni singolo file musicale, nelle collezioni e per tutto il repertorio. L'entropia permette di misurare le simmetrie presenti all'interno delle melodie. E' stata rilevata una debole correlazione (inversa) tra la lunghezza della melodia e l'entropia. L'entropia diminuisce, in media, seppur debolmente, all'aumentare del numero di note, mentre nessuna correlazione sembra esistere tra la dimensione dello spazio musicale e l'entropia. L'analisi delle reti, realizzata al livello della singola traccia, delle collezioni e dell'intera opera di Bach conferma che le reti sono in molti casi di piccolo mondo (Small World Networks), ma non perfettamente Scale Free. Le reti musicali si esprimono, invece che con una legge di potenza (secondo il modello di Barabási e Albert), con una legge di potenza troncata. Si individua, inoltre, una grammatica compositiva che realizza, nelle reti, motivi a differenti livelli di complessità, e cicli di note che ci offrono una visione dinamica e temporale delle reti musicali.

Nel Capitolo 7, attraverso la Connettistica e la Trattografia, modellazione descritta nel capitolo 4, 13 soggetti con diagnosi di Parkinson di circa 70 anni di età (in particolare, una variante del Parkinson, detta

Paralisi Sopranucleare Progressiva, in lingua Inglese *Progressive Supranuclear Palsy*, o PSP), 13 di controllo, della stessa età, sono stati messi a confronto per individuare quali aree cerebrali sono colpite in questa malattia neurodegenerativa. In particolare, oltre all'analisi statistica delle caratteristiche strutturali delle singole reti cerebrali, per tutti i soggetti, sono state rilevate caratteristiche di Piccolo mondo (Small World Networks) e di invarianza di scala (Scale Free), che non si manifestano per tutte le reti e non per tutti i gruppi con gli stessi valori. Inoltre, il parametro k , pensato come strumento di supporto alla diagnosi, è stato utilizzato per individuare i nodi della rete (e quindi le aree del cervello) che sono coinvolte nella neuro-degenerazione cerebrale, allargando i confini della diagnosi tradizionale e mettendo in evidenza che oltre alle aree usualmente interessate a queste malattie, anche altre aree cerebrali possono essere coinvolte nel processo di degenerazione. Nel Capitolo 8, un'investigazione sistematica di WordNet, per un gruppo di parole emozionali, ha fornito un quadro innovativo dei processi di aggregazione semantica, manifestando tutta la rete delle emozioni e la sua strutturazione. Anche di questo esperimento, dopo aver analizzato i dati statistici di rete (SWN e SF per la rete WordNet), sono riportate le straordinarie simmetrie che emergono dentro il sistema che contiene circa 140.000 termini della lingua inglese. Si dimostra che gli elementi della rete non hanno la stessa natura, per cui il contributo innovativo della tesi è quello di mettere in evidenza come tali elementi si aggregano in motivi e dinamiche, fino a questo momento mai viste nella letteratura sull'argomento. Si precisano, inoltre, una serie funzioni che le grammatiche generative svolgono nella lingua naturale, con un livello di complessità maggiore in quanto le crescite, lungi dall'essere preordinate attraverso l'applicazione rigida delle regole, crescono con blocchi complessi, non trattabili dal punto di vista formale.

Le Conclusioni generali, infine, ritornano sui problemi del mondo reale affrontati in questa tesi, cercando di fornire risposte metodologiche nuove all'analisi dei linguaggi.

Capitolo 1. Aspetti teorici e formali dei linguaggi e delle lingue

1.1. Introduzione

Nella specie umana, la comparsa della capacità di parlare attraverso una lingua articolata ha permesso un'evoluzione senza precedenti dal punto di vista cognitivo-comportamentale. Si sono sviluppate la funzione propriamente comunicativa, per esprimere pensieri e stati mentali o emozioni; la funzione cognitiva, per rappresentare la realtà ed elaborarla nella propria mente, attraverso il processo di percezione-azione-simulazione del parlato (Buccino et al. 2005), la creazione di categorie concettuali (Rosch, 1978), che rappresentano la base di conoscenza incorporata in un sistema fisico; e la facoltà di interazione con gli altri membri della specie, per la costruzione del comportamento sociale (Gallese et al., 2004). Lo sviluppo della lingua nella specie umana è stato un processo ampio e articolato, che, espresso in un arco temporale molto vasto (Guasti, 2007), sottende il percorso filogenetico della specie umana. La lingua parlata contiene una serie di livelli, universali e condivisi in quasi tutte le lingue del mondo (fonetico-fonologico e morfo-sintattico-semantic) che permettono, nonostante la diversità di superficie (Chomsky, 1971), di attivare gli stessi meccanismi cognitivi di base (Chomsky, 1968). Nell'acquisire il pensiero maturo e interagire con il mondo esterno, il bambino perviene a diverse forme di rappresentazione (Bruner 1976), realizzando diversi stadi (Piaget, 1972), attraverso un coordinamento del sistema cognitivo e del corpo con i differenti processi funzionali della percezione-azione e della cognizione. Per cui, in un primo stadio evolutivo, la realtà è interpretata attraverso l'azione, in seguito attraverso le immagini, il linguaggio e altri sistemi simbolici. Se i sistemi di rappresentazione iconica possiedono ancora forti legami con la realtà, essi sono completamente abbandonati nei sistemi simbolici. Per cui, tra linguaggio e realtà esiste un rapporto di mediazione costituito dai significati delle parole. Ma la costruzione della realtà esterna avviene anche attraverso un potente mezzo di comprensione, le emozioni. Le emozioni sono sistemi funzionali che sviluppano l'apprendimento (Darwin, 1859). Un'ulteriore evoluzione dei linguaggi e delle lingue si è avuta con lo sviluppo dei linguaggi artificiali, che ha permesso la costruzione di sistemi simbolici di secondo livello (i modelli simulati del comportamento intelligente), che a loro volta hanno realizzato i computer come

potenti sistemi cognitivi e poi gli agenti intelligenti che incorporano programmi nel corpo fisico di robot (Bilotta, 1999). Particolare importanza rivestono i linguaggi formali e le grammatiche generative (Chomsky, 1957), che riflettono meccanismi di base o leggi generali che valgono per tutti i linguaggi e le lingue, di cui si darà discussione, per poi collegarsi ai processi di modellizzazione realizzati nei capitoli successivi. Il capitolo ha lo scopo di introdurre tale complesso e multidisciplinare insieme di concetti teorici e formali sui sistemi di comunicazione che fanno da cornice teorica alla tesi.

1.2. Evoluzione della facoltà del linguaggio e sviluppo della cognizione negli umani

Nell'uomo, il sistema linguistico si sviluppa attraverso una serie progressiva e incrementale di passi fino ai tre anni di età (Camaioni & Di Blasio, 2007). Lo sviluppo della lingua, inafferrabile e altamente imprevedibile, ha fatto sorgere una serie di ipotesi di ricerca e di accesi dibattiti sulla natura di tale incredibile facoltà, presente solo in parte in altre specie biologiche. Non sempre le lingue, i sistemi di comunicazione per eccellenza della specie umana, sono state così raffinate e evolute come lo sono attualmente. Una progressiva complessificazione della comunicazione e delle strategie comportamentali hanno portato il genere *Homo* a modificare profondamente l'ambiente, e a creare strumenti di supporto per migliorare le proprie capacità cognitive. L'architettura cognitiva degli esseri umani si è modificata nel corso del tempo. La capacità generale di comunicare, diffusa con gradi differenti fra tutte le specie (Darwin, 1859), i linguaggi, le lingue naturali, i sistemi di comunicazione mediatizzati, i sistemi robotici artificiali e incorporati (Bilotta, 1999; Nolfi & Mirolli, 2010) sono lo specchio di un costante adattamento cognitivo e fisico dell'uomo all'ambiente. Il cambiamento è avvenuto attraverso gradi che sono enucleati, teoricamente, nel pensiero di alcuni ricercatori che hanno studiato lo sviluppo comunicativo-cognitivo dei soggetti umani, anche con punti di vista contrastanti.

1.2.1. Il linguaggio è innato o appreso?

La prima più grande *querelle* per quanto riguarda l'origine del linguaggio, e, più in generale, l'origine delle capacità cognitive, ha come obiettivo lo studio della *dicotomia innato/appreso* o *natura/cultura*. Gruppi di ricercatori sostengono che la capacità comunicativa e la lingua siano facoltà apprese, altri, invece, che il sistema linguistico si sviluppi sulla base di meccanismi *innati e preformati*. Sin dai tempi di Pavlov (Eysenck, 1964), si capì che il condizionamento poteva diventare uno strumento di modellazione del comportamento umano. Per Watson (1985), anche la capacità linguistica, essendo essenzialmente un comportamento verbale, è acquisita attraverso processi di apprendimento, come qualsiasi altro tipo di comportamento: il possesso della lingua avviene attraverso il condizionamento ambientale. Noam Chomsky (1959) non è di questo parere e critica l'approccio del comportamentista Skinner (1957), ipotizzando l'esistenza di complessi dispositivi cognitivi in dotazione all'individuo (Paternoster, 2002), una grammatica universale (GU o facoltà di linguaggio). Su questa base si innesca il sistema di acquisizione di una lingua o L. A. D., *Language*

Acquisition Device, che elabora in input il linguaggio parlato, a disposizione del bambino. Dopo diverse elaborazioni che processano tali dati, estraendo il sistema fonologico e fonetico, il sistema morfologico-sintattico, si forma in uscita la grammatica di tale lingua. La grammatica lavora su simboli. E' proprio grazie a questa funzione simbolica che il bambino può rievocare realtà e oggetti assenti (Piaget, 1975). Dopo l'acquisizione di competenze nell'uso del corpo in interazione con l'ambiente, il bambino usa i simboli: è il passaggio dall'intelligenza sensoriale a quella simbolico-rappresentativa, che avviene intorno ai diciotto mesi di età. Sempre in questo periodo, i bambini acquisiscono altre capacità simboliche, come imitare e giocare a fare finta, che sono fondamentali per il ricordo e il gioco di simulazione. Bruner (1983) postula l'esistenza di un *Language Acquisition Support System* (L. A. S. S.), un sistema di supporto all'acquisizione del linguaggio, fornito dall'ambiente sociale: è il "*linguaggio delle mamme, motherese*" oppure il "*linguaggio diretto ai bambini*", quel modo di parlare con voce acuta e cantilenante, che sostiene lo sviluppo linguistico e rivela le emozioni del parlante, costituendo così il *canale* attraverso il quale il linguaggio esprime un significato. Le frasi sono brevi e sintatticamente semplici, l'intonazione è esagerata, il lessico è concreto e sono presenti numerose ripetizioni. Il bambino, interagendo con le altre figure del suo contesto sociale, comincia a realizzare la base su cui si fonda la comunicazione interpersonale. Bruner ritiene che le basi biologiche per lo sviluppo del linguaggio si strutturano nell'interazione sociale. Tomasello (2003) individua invece meccanismi generali che permettono al bambino l'acquisizione del linguaggio, con due funzioni principali. La prima quella di imitazione, la seconda l'uso dell'analogia per la generalizzazione dei concetti. Un importante contributo che integra i due punti di vista epistemologici dell'innatismo e del costruttivismo riguarda l'approccio neurocostruttivista di Karmiloff-Smith (1995). Questa ricercatrice afferma che l'acquisizione del linguaggio è un processo di progressiva specializzazione delle aree emisferiche delegate alle funzioni linguistiche, che risulta dall'interazione tra vincoli biologici (fattori innati) e esperienza (apprendimento). All'interno di tale prospettiva, lo sviluppo ontogenetico è inteso come il processo costruttivo, attraverso il quale i geni interagiscono con l'ambiente, a vari livelli, incluso quello esterno all'individuo, per produrre strutture biologiche caratterizzate da grande complessità, dove ogni stadio superiore permette al soggetto una maggiore adattività e organizzazione rispetto agli stati precedenti. Tale prospettiva implica una concezione di sviluppo come processo continuo, nel corso del quale ciò che cambia è il *formato delle rappresentazioni*, ossia il modo attraverso cui la conoscenza è rappresentata nella mente e nel cervello. Karmiloff-Smith parla di regole di *ri-rappresentazione da un formato ad un altro*. Secondo questo approccio, il linguaggio è costruito sulle basi cognitive necessarie ad acquisire altre conoscenze e abilità (Karmiloff-Smith, 1979), arrivando a conciliare e integrare fattori innati e apprendimento. La memoria semantica permette di stabilire un sistema di relazioni tra eventi del mondo dando quindi un "significato" ai vari termini all'interno di un dato schema. Essa consente la costruzione di un modello mentale del mondo (Tulving, 1972, 1983).

Per cui è chiaro che si sottende, in questo sistema di ri-scrittura un sistema simbolico. Infatti, il linguaggio si è sviluppato nella specie *Homo* passando da sistemi di rappresentazione iconici a quelli simbolici (Harnad, 1990). Ma come sono elaborate le rappresentazioni simboliche? Secondo il modello cognitivista (Neisser, 1967), la mente è considerata un elaboratore elettronico e le informazioni acquisite dall'ambiente sono analizzate tramite un processo di codifica. In seguito, esse sono tradotte in un formato linguistico, le cui caratteristiche principali sono l'arbitrarietà e la discretezza. I simboli sono arbitrari perché non c'è nessun collegamento tra un simbolo e ciò che essi denotano, e sono discreti perché o sono espressioni complesse che possono essere scomposte in parti più piccole, o sono simboli atomici, che non hanno nessuna struttura interna. Come fa notare Harnad (1990), un modello puramente simbolico non instaura connessioni tra i simboli ed i loro referenti, mentre un modello puramente connessionista difetta, invece, della proprietà compositiva. Secondo questo autore, affinché si abbia significato, occorre che cognizione, abilità senso-motorie ed il mondo nel quale si è immersi siano strettamente collegati (Brooks, 1990). Bisogna fondare i simboli in qualcosa che non siano altri simboli privi di significato. Questo approccio è noto come il *symbol grounding problem*. Esso rappresenta una svolta nelle Scienze Cognitive perché dà origine ad un filone di ricerca che, sostenuto dalle scoperte sui neuroni specchio e il sistema di simulazione su cui si basa il linguaggio (Rizzolatti & Arbib, 1998; Rizzolatti & Sinigaglia, 2006), afferma che l'origine dell'espressione comunicativa risiede nel corpo e nei processi percettivo-motori del corpo, in interazione con l'ambiente. La rappresentazione simbolica deve essere fondata in maniera "bottom-up" (dall'alto verso il basso) nelle rappresentazioni sensoriali non simboliche.

Gli studi sull'evoluzione dell'architettura del cervello, nel genere *Homo*, cercano di capire come si sia evoluto il nostro sistema di comunicazione. Esistono tre ipotesi. La prima ritiene che il parlato si sia evoluto da un primitivo *linguaggio gestuale in modo graduale* (Gentilucci & Corballis, 2006). Secondo questa ipotesi, nell'*Homo Habilis*, la progressiva specializzazione dell'area di Broca e dell'area di Wernicke (collocate nel lobo temporale sinistro), diede impulso alla realizzazione forme gestuali per la comunicazione, arricchite da suoni strutturati, fundamentalmente con funzione di segnalazione di pericolo, per gli altri componenti del gruppo. Il successivo sviluppo dell'apparato fonatorio, in *Homo sapiens*, diede la possibilità di articolare differenti suoni, a seconda dei punti di articolazione della lingua con la bocca e il palato. L'elasticità dello strumento vocale ha permesso il fiorire dell'enorme complessità e varietà delle lingue naturali. Questa ipotesi si basa sulla stretta interconnessione fra lo sviluppo di aree cerebrali deputate alla processazione linguistico-semantica e aree motorie. Si pensa che la cognizione sia strettamente legata al sistema motorio. *L'intelligenza è motricità, o cognizione in movimento* (Thelen & Smith, 1994). La seconda ipotesi sostiene le lingue naturali contemporanee siano un'evoluzione del linguaggio sonoro dei primati, che hanno una certa capacità di comunicare con grida e suoni, anche strutturati, soprattutto per i richiami di difesa e di attacco e i richiami relativi all'accoppiamento (Gentilucci & Corballis, 2006). La terza ipotesi basa la nascita del linguaggio sulle trasformazioni morfologiche

dell'apparato fonatorio umano. A partire dai comportamenti di imitazione e dalla funzione simulativa del comportamento dei conspecifici, per mezzo dei neuroni specchio, si è sviluppato un articolato insieme di funzioni cognitivo-motorie, legate alla comunicazione per mezzo di un linguaggio articolato (Rizzolatti & Sinigaglia, 2006; Molnar-Szakacs et al., 2006).

Per cui, nel corso dell'evoluzione, attraverso un progressivo passaggio dai gesti manuali ai gesti orali (Corballis, 2009; Zlatev, 2008), diverse specie di ominidi hanno introdotto innovazioni a livello della struttura e della funzionalità della produzione linguistica. Un possibile sviluppo potrebbe essere correlato alle seguenti manifestazioni (Corballis, 2009): l'*Homo habilis* plausibilmente inizia a comunicare attraverso una rudimentale forma di proto-linguaggio gestuale, che è rimasto a livello di comunicazione non verbale; *Homo erectus* produce comportamenti motori mimico-gestuali, che regolano l'espressione dell'emozioni, anche queste realizzate come un potente linguaggio di veicolazione di significati emozionali; mentre la specie *Homo sapiens* presenta una specializzazione delle strutture cerebrali (specie delle aree dell'emisfero sinistro e credibilmente ha anche sviluppato circuiti di neuroni specchio), che consentono di correlare la comunicazione gestuale-motoria, con le forme di articolazione motorio-vocale, *Homo sapiens-sapiens* usa in modo intenzionale e associa una funzione cognitiva alla comunicazione gestuale e vocale. Alla base di questo lento passaggio evolutivo pare ipotizzabile un *meccanismo istintivo* (Pinker, 1994) di atti comunicativi condivisi in un gruppo sociale, che avrebbero permesso di accostare e associare il significato del gesto a quello della parola (Arbib, 2005). Durante l'evoluzione filogenetica, quindi, il linguaggio verbale pare aver sfruttato le possibilità consentite dall'area di Broca e dai circuiti dei neuroni specchio, sia per quanto riguarda il riconoscimento degli schemi motori fonetici sia per l'articolazione vocale.

1.2.2. Né innato né appreso, ma tanti modelli cognitivi contribuiscono a determinarlo

Lieven (2010), analizzando campioni di forme morfologiche, sintattiche e lessicali negli input che ascoltano i bambini, ha scoperto che un fattore molto importante nell'acquisizione della lingua è la frequenza di esposizione dei soggetti al linguaggio. Tale fattore, diversamente dall'ipotesi associazionista, sarebbe determinato non tanto da strutture interiori che si modificano, ma dalla quantità d'apprendimento, in associazione alle condizioni ambientali. Una rappresentazione ripetuta è rinforzata, mentre non lo è un'altra rappresentazione che non viene ripetuta. In quest'analisi, esiste una difficoltà oggettiva nell'ottenimento di misurazioni empiriche adeguate della frequenza delle forme morfologiche cui i soggetti sono esposti e la specificità/modalità attraverso le quali si effettuano le misurazioni. Dal punto di vista pratico, si tratta di arrivare a determinare esattamente come misurare la quantità di input linguistici cui il bambino è esposto, considerando i mutamenti che sopravvengono durante lo sviluppo. Lieven non considera i bambini come dei "processori semplici di stimoli", ma analizza la frequenza delle ripetizioni, che potrebbe influenzare in modo notevole l'acquisizione. Secondo Roeper (2007), la frequenza non

causerebbe alcuna modifica all'informazione presente in una rappresentazione, perché quest'ultima è già o presente, o assente.

Un'altra ipotesi dell'acquisizione del linguaggio si basa su modelli probabilistici (Hsu et al., 2011). Secondo questi autori, le prospettive sull'acquisizione del linguaggio sono due: il modello di apprendimento discriminativo (Tolman, 1981) e quello generativo, elaborato da Chomsky (Hsu & Griffith, 2009). Nel modello discriminativo, il processo di apprendimento avviene stabilendo un confine tra le categorie, mappando gli input delle categorie da una serie di coppie/categorie. Nel linguaggio si hanno categorie di frasi grammaticali e non grammaticali. La principale difficoltà dell'approccio discriminativo risiede nel fatto che la capacità di apprendimento si fonda su esempi positivi, come le frasi grammaticali. Tuttavia chi apprende non ha sufficienti basi per stabilire i confini tra frasi che hanno senso, o che non hanno senso, a livello grammaticale. Per cui, le analisi sui processi di apprendimento di una lingua tendono ad essere poco efficaci (Nowak et al., 2002). Un approccio bayesiano per stabilire i processi di acquisizione della lingua, basato su inferenze di tipo probabilistico, con distribuzione probabilistica, cui si deve la produzione degli input del linguaggio, è proposto da Hsu e collaboratori (Hsu et al., 2011). Gli autori ritengono che l'acquisizione della lingua non è solo una questione di discriminazione tra forme linguistiche "*positive*", quindi da ritenere in successiva elaborazione o "*negative*", quindi da scartare perché non utili, ma piuttosto l'estrazione sistematica delle regolarità presenti nel linguaggio stesso. La lingua è acquisita a partire dalle regolarità del parlato presente nel contesto linguistico. Secondo l'ipotesi degli autori, molte costruzioni linguistiche, considerate come innate, possono essere acquisite per via probabilistica. Secondo Kuhl (2010), una complessa serie di modelli cognitivi interagenti, presenti nei soggetti di età evolutiva, contribuiscono allo sviluppo della capacità comunicativa attraverso una lingua. Tali sistemi, nella maggior parte dei casi, non sono analoghi a quelli degli adulti, ma permettono agli infanti umani di apprendere il linguaggio, sia tramite codici orali che vocali, sia tramite codici visivi e/o manuali. La plasticità neurale in età infantile decresce nel corso dello sviluppo. Ricerche sulla percezione fonetica nel primo anno di vita dei bambini hanno mostrato come l'infante sia dotato di un potente meccanismo per l'apprendimento del linguaggio, che comprende capacità di tipo computazionale, cognitivo e sociale. Sempre secondo Kuhl (2010), il meccanismo non presenterebbe alcuna similitudine con il modello di apprendimento basato sul rinforzo proposto da Skinner, né con le modalità descritte da Chomsky. I bambini apprendono in maniera molteplice, non riconducibile ad un'unica prospettiva. In generale, il periodo in cui è più facile l'apprendimento del linguaggio arriva fino all'età della pubertà (Lennenberg, 1982), per poi iniziare a decrescere. Inoltre, non tutti i livelli linguistici sono appresi contemporaneamente. L'acquisizione della fonetica è particolarmente intensa, prima della fine del primo anno di vita, mentre l'apprendimento della sintassi raggiunge il massimo intorno ai diciotto mesi di età; altri elementi fondanti del linguaggio, quali l'apprendimento dei significati (semantica) non presentano periodi rilevanti di sviluppo, ma tale capacità resta costante nel corso di tutta la vita dei soggetti umani. Come mostrano studi recenti (Heidar, 2012),

l'acquisizione/apprendimento della lingua restano argomenti controversi e continuano a permanere ipotesi diverse, persino fra loro contrastanti. Tutte le teorie, però, sono unanimi nel ritenere che la conoscenza linguistica si fonda sull'acquisizione di regole e principi. Più in generale, tra teorie d'ispirazione innatista o empirista perdura il dissenso: le due ipotesi devono ad ogni modo essere esaminate e verificate dal punto di vista sperimentale. Senza prove empiriche, l'acquisizione del linguaggio non può essere basata né su fattori innati né su fattori appresi o culturali. I due approcci teorici sono incompleti per procedere alla realizzazione di una teoria generale dell'acquisizione del linguaggio. L'approccio che vede l'acquisizione del linguaggio come un insieme di regole ribadisce la rilevanza della creazione di schemi linguistici e relativa memorizzazione in categorie da parte dei soggetti umani. Grazie a tali schemi è possibile realizzare un processo di generalizzazione, che permette la creazione di frasi sempre nuove, sullo stesso modello categoriale, attivando anche il riconoscimento di varianti degli stessi modelli. Nell'acquisizione del linguaggio, un ruolo primario è giocato dalla pragmatica della comunicazione umana. L'acquisizione del linguaggio, o meglio di ciascuno dei differenti linguaggi convenzionali, può essere compresa solo a partire dal processo comunicativo, inteso nel più ampio senso possibile. Infine, attualmente in merito all'acquisizione del linguaggio vi sono due diversi schieramenti, da un lato i ricercatori a favore dell'apprendimento "discriminatorio", e dall'altro lato quelli a favore dell'apprendimento "generativo" (Hsu & Griffiths, 2009, 2011). Secondo l'apprendimento discriminatorio il bambino fa delle discriminazioni tra frasi grammaticali e frasi non grammaticali, mentre per la teoria generativa l'acquisizione del linguaggio parte dagli input esterni che poi vengono generalizzati dal bambino (Hsu et al., 2011).

Tra le diverse impostazioni non è stato raggiunto un consenso unitario: attualmente nuove ipotesi esplicative del mistero della comunicazione umana coesistono tra loro, ma nessuna ha raggiunto una piena validazione sperimentale (Tomasello & Slobin, 2005).

1.3. La struttura organizzativa della lingua

Le basi dello studio scientifico del linguaggio sono state poste da De Saussure nel *Cours de linguistique générale*, pubblicato postumo nel 1916. In tale scritto si definisce linguaggio la capacità generale della specie umana di produrre segni e significati, interpretabili da soggetti che condividono lo stesso codice di comunicazione. Tale capacità è preesistente alla sua realizzazione concreta (il linguaggio parlato o scritto, il linguaggio gestuale). Si definisce invece lingua, uno dei circa 7000 sistemi di comunicazione, storicamente determinati nel tempo, che attualmente sono usati nel mondo. Ogni lingua è costituita da un gruppo di elementi minimali portatori di significato (i morfemi e i fonemi), organizzati fra loro in un complesso sistema di interrelazioni, di natura gerarchica. Rivedendo la teoria di De Saussure (1916) che pensava al linguaggio come un sistema di segni, alla luce della moderna Teoria della Complessità (Morin, 1983), ci si accorge che il linguaggio e le lingue naturali rientrano perfettamente nel novero dei sistemi complessi, in quanto possiedono una serie di elementi che interagiscono tra di loro e sono governati da un principio di

organizzazione che determina il comportamento locale e globale del sistema stesso. L'altra caratteristica propria delle lingue e dei sistemi complessi è che i diversi livelli (e ogni componente a ciascun livello), sono collegati da una grande varietà di legami, con una enorme quantità di interconnessioni. Le interazioni tra gli elementi di un sistema complesso non sono prevedibili apriori. Per cui, esse dipendono dalle condizioni iniziali. E' impossibile prevedere le scelte linguistiche che farà un parlante per esprimere un concetto in una lingua. Per questa imprevedibilità, e per le caratteristiche di emergenza che conferiscono agli elementi qualità non possedute singolarmente, ma globali e differenti rispetto a quelle di partenza, anche le lingue sono estremamente sensibili ai dati iniziali, e le dinamiche di interazione tra gli elementi possono essere considerate non lineari e caotiche (Bilotta & Pantano, 2008). E' notevole anche che linguaggi e lingue condividano processi che si strutturano a differenti livelli di complessità.

Il primo livello di organizzazione delle lingue naturali è quello fonetico e fonologico (De Mauro, 2008). La capacità di produrre suoni articolati negli umani è praticamente illimitata, anche se ogni lingua naturale ha un insieme relativamente piccolo di unità sonore, dette fonemi. Tali occorrenze fisiche sono studiate dalla Fonetica (Lyons, 1982) per quanto riguarda la loro produzione a livello del tratto articolatorio e del riconoscimento attraverso il canale sensoriale uditivo, mentre la Fonologia si interessa dei modelli mentali dei suoni (Simone, 2008), memorizzati e categorizzati dai neonati già fin dai primi mesi di vita. Il riconoscimento dei suoni e delle loro varianti si basa proprio sui modelli mentali fonologici. Un fonema è la più piccola unità di significato di una lingua naturale. I suoni associati ad ogni fonema di solito hanno una configurazione articolatoria in comune, che permette appunto il loro riconoscimento, pur nella variabilità di articolazione di ogni parlante. Ogni lingua ha in genere un inventario dei suoni. Un tale insieme di simboli costituisce un alfabeto fonetico, che si suddivide in fonemi *primari*, *semplici* e *composti* (Lyons, 1982). I fonemi primari semplici, nelle diverse lingue, variano circa da quindici a cinquanta (l'inglese standard ne possiede trentadue). I fonemi composti sono combinazioni di fonemi semplici, che si comportano come unità per quel che riguarda il significato e la struttura della parola (Valesio, 1969). Pertanto, l'individuazione dei fonemi di un sistema linguistico avviene attraverso la "prova di commutazione", il confronto cioè dei lessemi (o coppie minime), al fine di individuare l'elemento che, a parità di condizioni, provoca differenze di significato. Ciò significa che l'opposizione /t~d/ nel sistema linguistico dell'italiano rappresenta una coppia minima, e che i due elementi interessati sono fonemi del sistema fonologico italiano (Nespor, 1993; Nespor & Bafile, 2008). Considerando la molteplicità e la varietà dei suoni delle lingue e, all'interno della stessa lingua, la varietà individuale tra soggetti, è sorprendente che alcune differenze articolatorie siano tratti universali per tutte le lingue. Tali tratti riguardano: la distinzione tra i differenti tipi di consonanti, la differenza fra consonanti sorde e sonore (suoni per i quali esiste o meno una vibrazione delle corde vocali del tratto articolatorio), la distinzione fra vocali e consonanti.

Il secondo livello di organizzazione di una lingua è quello della morfologia. Gruppi di fonemi si uniscono, attraverso regole di combinazione specifiche per ogni lingua naturale, per formare unità significative di secondo livello, con funzioni grammaticali. La morfologia struttura la forma delle parole, come le parole si creano e la loro struttura interna (radice e desinenza, prefissi e suffissi). Per cui, possono essere considerati morfemi una singola parola come il verbo “vai” o le desinenze “o”, “i”, che individuano, nella lingua italiana, alcuni nomi maschili, rispettivamente singolari e plurali. Considerando termini come “ripensare”, possiamo spezzare la parola in tre morfemi differenti. Il morfema “ri”, il morfema “pens” e il morfema “are”. Il morfema “pens” è considerato il morfema radice, cui si attaccano gli altri pezzi. La radice può essere semplice, costituita solo da una parte, o complessa, costituita da più parti. La radice costituisce il cuore della parola cui si attaccano il prefisso “ri” e il suffisso “are”. L’approccio della linguistica strutturalista della prima metà del ventesimo secolo ha portato all’analisi particolareggiata di molte lingue naturali (Booij, 2007). Si pensava che i linguisti avessero bisogno di metodi dettagliati per l’analisi degli aspetti morfologici delle lingue, che li potessero rendere indipendenti dal loro oggetto di studio. Tale tipo di studi determina l’approccio *analitico*, estremamente concreto, anche quando si esaminano linguaggi lontani dalla tradizione indo-europea. Nello studio della morfologia delle lingue, l’approccio *sintetico* considera gli aspetti teorici, piuttosto che metodologici, ponendosi il problema di indagare come i morfemi si leghino tra loro, per scoprire leggi di combinazione morfologiche, in una lingua data. Presuppone una conoscenza funzionale-grammaticale dei morfemi, insieme con una ipotesi teorico-costruttiva sulle regole di accoppiamento fra loro. I problemi su cui indagano i due approcci sono comunque generali, e riguardano: come fa un parlante di una lingua a produrre una parola, quando ne ha bisogno; come si formano le parole; quali analisi si realizzano per il riconoscimento delle parole, nella lettura di testi scritti o nell’ascolto; quali sono i meccanismi che sovrintendono alla creazione di neologismi, a partire da morfemi pre-esistenti; come si sono modificati le strutture morfologiche e le parole nel corso del tempo.

Gli elementi dell’approccio analitico per lo studio della morfologia, ritrovati in quasi tutte le lingue sono:

1. morfemi con la stessa forma, organizzazione sonora e significato sono quasi sempre occorrenze concrete dello stesso morfema, in tutte le loro manifestazioni;
2. morfemi con lo stesso significato, ma con organizzazione sonora diversa, sono istanze dello stesso morfema ideale, se non si sovrappongono alla distribuzione dei morfemi cui somigliano;
3. morfemi non segmentabili, con forme non pronunciabili e dipendenti da altri morfemi per la loro realizzazione;
4. morfemi privi di allomorfi (morfemi che possono ricorrere negli stessi contesti, sovrapponendosi) specificano, per esempio, il singolare dal plurale o altre funzioni grammaticali.

Il terzo livello di organizzazione di una lingua è la sintassi, che si riferisce all’ordinamento delle parole all’interno di una frase. Le *regole sintattiche* variano da lingua a lingua e riguardano l’ordine delle parole e

le trasformazioni delle frasi (da attiva a passiva o interrogativa). Nel processo di comprensione delle frasi esistono due livelli fondamentali di analisi, che correlano strettamente la sintassi e la semantica:

- a. Analisi sintattica (grammaticale), o parsing;
- b. Analisi del significato della frase, o pragmatica.

In ogni lingua è possibile formulare un numero infinito di frasi, ma l'interpretazione delle frasi è sempre critica, anche per soggetti parlanti nativi di una specifica lingua. Chomsky (1957, 1959) e altri linguisti hanno elaborato un insieme di regole per spiegare la regolarità e la produttività del linguaggio. Con termine *grammatica* si definisce l'insieme finito di regole che fornisce tutte le combinazioni possibili tra le parole, per produrre le frasi di una lingua. Un modello computazionale di interpretazione di frasi, assegnando loro la struttura grammaticale, basato sull'interpretazione di soggetti umani, e incorporato in sistemi automatici è definito "parsing". È un processo tutt'altro che semplice perché in quasi tutte le lingue le frasi presentano problemi di ambiguità nella struttura grammaticale. Alcune frasi sono sintatticamente ambigue a livello globale, altre a livello locale. I parser rappresentano sistemi computazionali di analisi di testo. Sono basati o su teorie seriali o parallele. Nel processo interpretativo di un testo, parser basati su teorie seriali e/o parallele analizzano nel primo caso una frase alla volta, per poi passare all'altra, nel secondo caso le analisi sono realizzate in parallelo. Secondo il modello connessionista, proposto da MacDonald et al. (1994), per risolvere le ambiguità delle frasi bisogna tenere conto:

- della conoscenza acquisita, in dotazione al sistema esperto, che impone limiti all'interpretazione;
- del problema che le codifiche già sperimentate e conservate in memoria, legate ai termini da comprendere, sono tra loro collegate;
- del fenomeno dei termini ambigui, non sempre ambigui nello stesso modo, o con lo stessa valenza di significato;
- del fatto che le regole grammaticali sono variabili, e legate a interpretazioni già sperimentate in altri contesti di frase.

La *semantica* è quella disciplina della Linguistica che studia il significato delle parole e delle frasi, sia nella produzione scritta che nell'ascolto di lingua parlata. Lo studio del significato delle frasi e dei testi, è molto più complesso della disambiguazione delle singole parole, perché il significato combinato di diverse parole messe insieme dipende sia dal significato delle parole singole, che dal modo in cui queste sono strutturalmente combinate all'interno di una frase o di un testo. All'interno della semantica, importanti sono i processi pragmatici della lingua e i processi di comprensione e disambiguazione del testo.

La *pragmatica*, infatti, complica notevolmente l'interpretazione del parlato e dei testi. Tale settore della Linguistica, si occupa delle modalità con cui il linguaggio è usato a scopo comunicativo (Austin, 1976).

Austin ritiene che quando parliamo, si realizzano tre funzioni principali. La prima funzione riguarda l'ascoltatore, nell'atto di estrapolare il significato esatto di una frase, che sta ascoltando o leggendo; la seconda funzione riguarda i significati che il parlante, o il produttore di un testo, intende veicolare; la terza funzione riguarda l'effetto che la frase suscita su chi ascolta (compreso chi pronuncia la stessa frase). Inoltre, la comprensione/elaborazione di parlato o di testi sarebbe irrealizzabile senza accedere alle conoscenze conservate nella memoria a lungo termine. Il ruolo cruciale di tali conoscenze si manifesta nel processo di inferenza, definito da Schank (1976) *come il nucleo indispensabile per realizzare il processo di comprensione in una lingua naturale*. Esistono inferenze che collegano gruppi di significati e inferenze che sono ritenute processi di elaborazione. Le prime sono utili per connettere insieme parti fondamentali del parlato o di un testo scritto, nella dinamica temporale dell'ascolto o della lettura. Le inferenze di elaborazione servono invece in fase di progettazione di un testo o di un discorso, per aggiungere particolari, ampliare o fornire più spessore all'elaborato linguistico.

L'anafora è la forma più naturale di inferenza di collegamento. Essa permette di identificare un nome o un pronome con un altro nome o una frase nominale, in modo da far ricordare nel corso del discorso elementi significativi già menzionati. Secondo McKoon e Ratcliff (1992), le inferenze sono spontanee (realizzate inconsapevolmente dai soggetti) o intenzionali (realizzate con un preciso scopo di comprensione). Alcune inferenze automatiche stabiliscono una coerenza locale. Altre inferenze automatiche si basano su informazioni facilmente utilizzabili, perché fanno parte della base di conoscenza di ogni parlante di una specifica lingua, o perché riportate in modo chiaro nel contenuto del testo o del parlato. Le inferenze intenzionali (o strategiche secondo gli autori) sono espresse da un lettore/ascoltatore, al fine di pervenire a obiettivi di conoscenza specifici. Sono usate per creare connessione tra gli elementi fondamentali, all'interno di un testo. Gli autori sostengono che, in mancanza di processi intenzionali sviluppati dai soggetti, con specifico intento di comprensione, sono realizzati due tipi di processi di inferenza (ipotesi minimalista dell'interpretazione di un testo). Appartengono al primo tipo, le inferenze che costruiscono rappresentazioni locali ma compatte delle porzioni di un testo, al secondo, quelle che si fondano su indicazioni che sono usabili in modo naturale e rapido.

Il *lessico*, infine, è l'insieme dei morfemi di una data lingua o del repertorio linguistico di una data persona, strutturato in mappe semantiche e cognitive (Collins e Quillian, 1969; Collins e Loftus, 1975). Tali organizzazioni, stivate nella memoria a lungo termine (Tulving 1972, 1989) forniscono la base sulla quale si fondano tutti i meccanismi di produzione del parlato, in quanto contengono una descrizione approfondita delle categorie concettuali che sono state immagazzinate attraverso l'esperienza, durante le fasi di acquisizione del linguaggio

Questa breve rassegna di concetti teorici si chiude, analizzando le tappe fondamentali che portano i bambini all'acquisizione della lingua. A due anni di età cominciano a produrre le prime forme di parlato,

dapprima utilizzando parole, in seguito unendo tra loro almeno due parole e anche più di due. Questa capacità di produrre frasi si correla con lo sviluppo del lessico. E' la base della conoscenza linguistica che si stiva nella memoria a lungo termine, che è costituita da un minimo di 300 parole, quando i bambini hanno circa 3 anni di età. Il numero minimo è estremamente incerto e influenzato da numerosi fattori esterni, quali il contesto sociale di provenienza, l'abitudine dei genitori a parlare col bambino e la significatività dell'interazione, la qualità delle interazioni con la madre, la presenza di più persone nell'ambiente, e non da ultimo caratteristiche individuali. Secondo De Santis (2010), le prime locuzioni sono molto esplicative e pur essendo sprovviste di elementi grammaticali e/o funzionali, sono decisamente comprensibili dai soggetti adulti che vivono con il bambino. I rapporti sociali, infatti, permettono al bambino di affinare il processo di produzione linguistica.

L'acquisizione di una lingua procede attraverso tre tappe fondamentali:

- nella prima, i bambini producono frasi a struttura nucleare, con unioni di due termini. Le frasi sono generalmente formate dalle strutture: Nome/Nome, Verbo/Nome, Nome/Verbo;
- nella seconda tappa, le forme iniziali di produzione linguistica si allargano, con l'aggiunta di modificatori e avverbi. Per cui gli enunciati, sebbene formati da due o tre elementi soltanto, sono modificati sintatticamente, attraverso l'uso di parole che recano le informazioni aggiuntive, di solito veicolate dagli aggettivi e dagli avverbi.
- nel terzo periodo sono prodotte strutture complesse. Intorno alla fine del terzo anno di età sono utilizzate frasi complesse, formate dall'organizzazione Nome + Verbo + Articolo + Nome, oppure sono prodotte due frasi unite da rapporti di coordinazione e/o di subordinazione.

Dai tre/quattro anni il bambino emette enunciati più estesi e il suo linguaggio diventa sempre più vicino al modello di lingua di soggetti adulti. Numerosi morfemi grammaticali sono appresi a circa tre anni e mezzo. Verso i tre/quattro anni si riducono le omissioni di articoli, pronomi e preposizioni, mentre le flessioni verbali, nominali e gli accordi sostantivo/aggettivo, soggetto/verbo sono solitamente pronunciati in maniera corretta (Del Bo et al., 1988; Volterra et al., 2005). Un recente studio (Rowe, et al., 2012) si è concentrato su come gli elementi iniziali di acquisizione del linguaggio nei primi quattordici/quarantasei mesi influenzino la produzione linguistica nei successivi sviluppi (fino a cinquantaquattro mesi), per verificare l'effetto sullo sviluppo posteriore del vocabolario. Se nella prima infanzia l'apprendimento del dizionario mostra significative differenze (Fenson et al., 1994), l'acquisizione iniziale appare correlata positivamente a fattori di tipo socio-culturale, quali lo status socio-economico della famiglia di provenienza e la tendenza, da parte dei genitori, a parlare con i bambini (Hoff, 2006). Come si diceva precedentemente, l'acquisizione del linguaggio è influenzata dall'ambiente in cui vive il bambino e dalla qualità delle sue interazioni sociali (Snow, 1994).

1.4. Lingua e emozioni

Un altro fattore determinante che influenza il parlato (scritto e orale) è correlato con le emozioni. Dal punto di vista cognitivo, le emozioni sono difficili da studiare e da analizzare perché di solito rappresentano strutture complesse della cognizione, alle quali si uniscono fattori personali, ambientali, sociali neurofisiologici che determinano una frammentazione degli approcci di studio in tale settore. Di seguito sono presentati alcuni dei principali metodi di studio delle emozioni.

Secondo Plutchik (1997), sono state proposte nell'ultimo secolo almeno un centinaio di descrizioni del termine "emozione", tali da rendere grandemente inaccessibile una loro definizione operativa. In questa disorganizzazione definitoria, i ricercatori hanno posto l'accento sui processi cognitivi che sovrintendono alle emozioni (Mandler, 1975, 1984), sui processi fisiologici che le scatenano (Hebb, 1946; Tomkins, 1962), sulle funzioni evolutive (Scherer et al., 2001; Izard, 1971, 1977; Plutchik, 1980) che hanno determinato importanti cambiamenti nello sviluppo umano. Plutchik (1980) definisce le emozioni di base, paura, rabbia, gioia, tristezza, accettazione, disgusto, aspettativa, sorpresa. Le emozioni sono schemi biologici e comportamentali, evoluti nel corso del tempo, tipiche per ogni specie, geneticamente ereditate, ma modificabili attraverso l'esperienza e il processo di apprendimento. Nel volume che fonda lo studio delle Emozioni a livello etologico e neuro-fisiologico, *The Expression of the Emotion in Animals and Men*, Darwin illustra gli aspetti fisiologici del processo emozionale. Le emozioni sono legate alle dinamiche comportamentali - esperite attraverso una modalità sensoriale detta propriocezione - che modificano il nostro corpo (piangere, respirare affannosamente, sentire il cuore battere più forte, ecc.). Una improvvisa paura accresce la dose di cortisolo nel sangue. Questo ormone scatena una reazione a livello cerebrale, attivando il circuito dell'ippocampo, in collegamento con l'ipotalamo e della corteccia prefrontale con l'amigdala, che è avvertita con la sensazione di "essere agitati", pronti cioè a fuggire, come accadeva quando i nostri antenati dovevano scappare davanti ad un avversario temibile e sconosciuto. Per cui, secondo Darwin, tutte le emozioni (dal latino *e movere*) hanno un significato funzionale per la sopravvivenza, presente anche in altre specie animali. Gli stati emozionali interni sono molto mutevoli e poco classificabili in modo sistematico, a causa della compresenza nel vissuto esperienziale, di diverse e alle volte contrastanti, esperienze emozionali. Esse sono dispiagate in modo privilegiato dal volto (Ekman et al., 1972; Ekman, 2003), ma anche dalla postura del corpo e dai gesti, nella comunicazione non verbale (Argyle, 1988), così come dalle caratteristiche prosodiche e intonazionali delle lingue, sia nel linguaggio materno (Snow & Ferguson, 1977), che nel parlato (Magno Caldognetto et al., 2008). Molti studiosi ritengono che le emozioni siano dei mediatori tra il nostro corpo e l'ambiente, dei regolatori della nostra vita cognitiva e affettiva (Frijda, 1986; Schore, 1994). La vita emotiva può essere rappresentata come una serie di eventi. A partire da uno stimolo che occorre nel nostro ambiente, si sviluppa un processo cognitivo. Tale processo ci fornisce l'esperienza di una emozione, che a sua volta produce un comportamento. James definisce

l'emozione come "il sentire i cambiamenti vegetativi che hanno luogo a livello viscerale, a seguito dello stimolo che li determina". Questa teoria è detta periferica (James, 1884). Ogni emozione presenta una propria organizzazione di risposte fisiologiche (Canestrari,1984). Durante il processo emotivo, il soggetto percepisce l'oggetto, la percezione crea delle modificazioni organiche, che sono le risposte neurovegetative ed espressivo-motorie. L'attività cosciente e cognitiva, ci fa percepire le modifiche negli stati del nostro corpo. Cannon (1927), è il primo studioso che contrappone alla teoria periferica di James, una "teoria centrale" delle emozioni, spostando la sede dell'emozione a livello neurofisiologico, nella regione talamica del cervello. Secondo Cannon, alla base di tutte le emozioni, c'è una forte attivazione indifferenziata, che viene definita "reazione d'emergenza", come ad esempio fame, dolore o paura. Inoltre, al contrario della teoria periferica, Cannon afferma che tutte le emozioni presentano la stessa configurazione di risposte fisiologiche. Schachter & Singer (1962) individuano una dimensione psicologica nel processo emotivo. Secondo tale approccio, i fattori cognitivi sono determinanti nell'attivazione dell'esperienza emozionale attraverso un processo di percezione e attribuzione causale (Anolli, 2002). L'emozione è cioè il risultante di due componenti, una fisiologica di attivazione e una cognitiva di valutazione dello stimolo/etichettamento della propria esperienza emotiva. In questo processo, particolare attenzione è dedicata all'attribuzione causale che stabilisce una connessione fra queste due componenti. Inoltre, in base all'esperimento condotto da Schachter & Singer nel 1962, la natura dell'emozione, a parità di fenomeni somatici periferici, dipende dalla situazione ambientale e dall'esperienza del soggetto. Papez (1937) avanza l'ipotesi circa un circuito di strutture nervose coinvolte nell'elaborazione e regolazione delle emozioni (detto in seguito circuito di Papez), secondo il quale la parete mediale di ciascun emisfero cerebrale è collegata all'ipotalamo e la parete laterale al talamo. Inoltre, le vie sensoriali insieme ai circuiti che gestiscono il movimento, e anche il linguaggio-pensiero si suddividono a livello del talamo e trasmettono gli impulsi al cervello. Secondo la Teoria centrale, il cervello è dunque il substrato che coordina le complesse catene di comportamenti, chiamate emozioni (Cannon, 1927). Arnold (1960a e b) introduce il termine "appraisal" per indicare un atto cognitivo che integra la percezione di un dato stimolo esterno al soggetto, del quale si può diventare consapevoli solo a processo concluso. Questo approccio considera le emozioni come dipendenti dal modo in cui gli individui valutano e interpretano gli stimoli che percepiscono dall'ambiente. Esiste un forte legame tra aspetti emotivi e cognitivi. Tomkins (1962), valutando le emozioni come processi di adattamento evolutivo, ritiene che esse siano strettamente associate alla realizzazione di scopi universali, connessi con la sopravvivenza della specie e dell'individuo. Sulla base degli studi di Plutchik (1962), Ekman (2003), Izard (1971, 1977), Frijda (1986) si ritiene che, a partire dai bisogni fondamentali, si generano le emozioni primarie (gioia, collera, paura, disgusto, sorpresa, tristezza e disprezzo), che, a loro volta, producono emozioni secondarie o miste, che sono una miscela di due o più emozioni primarie secondo proporzioni diverse. Senza entrare nel merito dei modelli neurofisiologici della funzione

emozionale perché esula dal contesto di questa tesi, si cercherà di analizzare il ruolo delle emozioni nelle lingue naturali in vari contesti, attraverso un approccio semantico.

1.5. Linguaggi formali

Un modello completamente scevro da elementi emozionali è invece un linguaggio formale, un sistema matematico di natura simbolica, che è costituito da un insieme di parole di lunghezza finita organizzate in frasi che sono interpretabili dal punto di vista del significato (Harnad, 1990; Chomsky, 1965; Hofstadter, 1999). La formalizzazione mette in evidenza le caratteristiche essenziali di ogni sistema di comunicazione.

Gli elementi di base di un linguaggio formale sono: un alfabeto, che è un insieme di simboli e una grammatica composta da una lista di regole. I simboli, organizzandosi a differenti e gerarchicamente ordinati livelli di complessità, attraverso la lista di regole combinatorie creano frasi. Mentre l'alfabeto ha un insieme finito di simboli, il linguaggio al contrario può avere un numero infinito di stringhe o di frasi.

Da un punto di vista matematico, un linguaggio L è definito come un insieme (possibilmente infinito) di stringhe sopra un alfabeto finito.

Partendo da un alfabeto V , si possono definire stringhe o parole con sequenze finite di caratteri che appartengono a V^* , composte da tutte le stringhe di simboli di V . Denotiamo con λ una stringa non costituita da alcun simbolo (avente lunghezza zero). Indichiamo con V^+ tutte le stringhe finite contenute in V^* , tranne la stringa vuota. Per cui:

$$V^+ = V^* - \{\lambda\}.$$

Un linguaggio L costruito su un alfabeto V è un sottoinsieme di V^* .

Un linguaggio $L \subseteq V^*$ che non contiene λ è generalmente indicato con λ -free. La lunghezza di una stringa $l \in V^*$ è rappresentata da $||l||$.

Il numero di occorrenze di un dato simbolo $a \in V$ in una stringa $l \in V^*$ è indicato da $|l|_a$.

La sequenza 10001001 è una parola definita sull'alfabeto binario $\{0, 1\}$. La sua lunghezza $||l||$ è 8. Questa stringa appartiene a $\{0, 1\}^*$. $L = \{01, 0011, 000111, \dots\}$, ovvero potremmo avere infinite sequenze di stringhe.

Seguendo la teoria dei linguaggi formali (Hopcroft et al., 2007), si può dire che le grammatiche possono essere specificate in molti modi diversi. Le stringhe possono essere infatti prodotte da una grammatica formale (seguendo la gerarchia dei linguaggi formali di Chomsky, 1965), da espressioni regolari, da una macchina di Turing o da un automa a stati finiti (Bilotta & Pantano, 2010). Operazioni quali *concatenazione*, *intersezione*, *unione*, *complementarietà*, *scrittura inversa* possono essere usate per dare origine a nuovi linguaggi artificiali. Supponiamo che L_1 e L_2 siano Linguaggi su un Alfabeto comune. La concatenazione $L_1 L_2$ è realizzata da tutte le stringhe con forma νw where $\nu \in L_1$ e $w \in L_2$. L'intersezione L_1 e L_2 consiste di tutte le stringhe che sono contenute in L_1 e L_2 .

L'unione di L_1 e L_2 è fatta di tutte le stringhe che sono in L_1 o in L_2 . La complementarità del linguaggio L_1 coinvolge tutte le stringhe sull'alfabeto che non sono contenute in L_1 .

Il concetto di grammatica serve per caratterizzare un linguaggio L , ovvero definire quali stringhe o parole di V^* sono contenute in L e quali invece no. Se L avesse un numero finito di stringhe, noi potremmo semplicemente contarle. Ma tutti i linguaggi, hanno un numero infinito di stringhe. Questa è la capacità creativa di tutti i linguaggi e delle lingue naturali: realizzare infinite parole e frasi, partendo da un numero relativamente piccolo di simboli. La grammatica include un insieme di regole dalle quali è possibile generare stringhe. Introdotta dagli studi sulle lingue naturali dal linguista Noam Chomsky negli anni Cinquanta, una grammatica G , è una quadrupla:

$$G = (V_T, V_N, S, P)$$

dove V_T è l'insieme finito di simboli terminali, V_N è l'insieme finito di simboli non-terminali.

Le due tipologie di simboli danno origine all'alfabeto:

$$V = V_T \cup V_N.$$

Simboli terminali e simboli non terminali sono insiemi disgiunti:

$$V_T \cap V_N = \emptyset.$$

La funzione primaria del linguaggio è realizzata da $S \in V_N$ (dove S = scopo), mentre P è l'insieme finito di regole di Produzione:

$$P = \{\alpha \rightarrow \beta \mid \alpha \in V^* \cdot V_N \cdot V^*, e \beta \in V^*\}.$$

Le grammatiche formali possono essere divise in due principali categorie, generative e analitiche. Ai fini della studio, tratteremo soltanto le grammatiche generative. Una grammatica generativa è un insieme di regole attraverso le quali tutte le stringhe possibili in un linguaggio possono essere generate usando stringhe di riscrittura in modo sequenziale, partendo da un simbolo dato. Una grammatica generativa è descritta da un algoritmo che genera stringhe in un linguaggio. Per caratterizzare un Linguaggio artificiale, partendo da una Grammatica è necessario introdurre il concetto di Derivazione, che usa le regole di Produzione per generare una stringa partendo da un'altra stringa. Si può avere una funzione di:

(i) Derivazione Diretta (\Rightarrow).

Esempio: se $\alpha \rightarrow \beta \in P$ e $\gamma, \delta \in V^*$, allora, $\gamma\alpha\delta \Rightarrow \gamma\beta\delta$.

(ii) Derivazione (\Rightarrow^*).

Esempio: se $\alpha_1 \Rightarrow \alpha_2, \alpha_2 \Rightarrow \alpha_3, \dots, \alpha_{n-1} \Rightarrow \alpha_n$, allora $\alpha_1 \Rightarrow^* \alpha_n$.

E' possibile generare un Linguaggio partendo da una Grammatica. Un linguaggio $L(G)$ è generato da una Grammatica G , se:

$$L(G) = \{w \in V_T^* \mid S \Rightarrow^* w\}$$

Possiamo dire che due Linguaggi sono equivalenti se $L(G1) \equiv L(G2)$.

Seguendo le definizioni date precedentemente, si può dire che una stringa appartiene a un Linguaggio se e solo se:

- a. La stringa è costituita solo da Simboli Terminali;
- b. La stringa può essere Derivata dallo Scopo, S , del Linguaggio;

Stringhe che appartengono a un linguaggio L sono dette frasi.

Chomsky introduce una classificazione delle Grammatiche per descrivere la complessità strutturale delle frasi nelle lingue naturali. Vincoli differenti sulle regole di Produzione definiscono diverse classi di Grammatiche / Linguaggi.

Le Grammatiche di tipo più generale sono chiamate Grammatiche di Tipo 0. Sono grammatiche formali del tipo che abbiamo introdotto precedentemente:

$$G = (V_T, V_N, S, P)$$

tale che tutte le produzioni in P seguano la seguente condizione:

$$\text{Tipo 0. } \alpha \rightarrow \beta$$

con $\alpha \in V^* \cdot V_N \cdot V^*$ e $\beta \in V^*$.

Le Grammatiche Sensibili al Contesto (Context-Sensitive Grammars) sono invece di Tipo 1. Anche queste sono grammatiche formali del tipo $G = (V_T, V_N, S, P)$, tali che tutte le produzioni in P osservino le seguenti condizioni:

$$\alpha A \gamma \rightarrow \alpha \beta \gamma$$

con

$$\alpha, \gamma \in V^*, \beta \in V^+$$

e

$$A \in V_N.$$

Inoltre, è permessa una regola del tipo:

$$S \rightarrow \lambda$$

se S non compare sul lato destro di ogni regola.

Il significato del termine "Sensibile al Contesto" è dato dalla α e dalla γ che formano il contesto di A e determinano se A può essere rimpiazzata da β oppure no.

Le Grammatiche Indipendenti dal Contesto (Context-Free Grammars) sono invece di Tipo 2. Anche questo tipo di Linguaggi sono grammatiche formali del tipo: $G = (V_T, V_N, S, P)$, tali che tutte le produzioni in P rispettino le seguenti condizioni:

$$A \rightarrow \beta$$

con $A \in V_N$ e $\beta \in V^*$.

L'espressione "Grammatiche Indipendenti dal Contesto" deriva dal fatto che il termine non-terminale A può essere rimpiazzato da β , non avendo cura di quale sia il contesto nel quale si manifesta o occorre. Queste grammatiche sono importanti perché sono potenti abbastanza da descrivere la sintassi dei linguaggi di programmazione. Quasi tutti i linguaggi di programmazione sono definiti come Indipendenti dal Contesto. Queste grammatiche sono abbastanza semplici da permettere la costruzione di algoritmi di analisi che per stringhe date determinano se e come esse possono essere generate dalla Grammatica. La Grammatica dell'Esempio 1 è Indipendente dal Contesto.

Secondo Chomsky (1965), le Grammatiche Regolari sono di Tipo 3. Anche queste sono Grammatiche formali del tipo $G = (V_T, V_N, S, P)$, tali che tutte le produzioni in P rispettino la seguente condizione:

$$A \rightarrow aB, \text{ o } A \rightarrow a$$

con $A, B \in V_T$ e $a \in V_T$. Furthermore, a rule of the form:

$$S \rightarrow \lambda$$

se S non compare sul lato destro di ogni regola. Queste regole definiscono la Grammatica Regolare Destra.

Le seguenti Produzioni:

$$A \rightarrow Ba, \text{ o } A \rightarrow a$$

definiscono la Grammatica Regolare Sinistra.

Le Grammatiche Regolari Destra e Sinistra definiscono gli stessi insiemi di Linguaggi. Le Grammatiche Regolari sono comunemente usate per definire la struttura lessicale dei linguaggi di programmazione.

Questi tipi di Grammatiche definiscono una gerarchia di linguaggi, detta Gerarchia di Chomsky. Ogni Linguaggio Regolare è Indipendente dal Contesto, ogni Linguaggio Indipendente dal Contesto è Sensibile al Contesto e ogni Linguaggio Sensibile al Contesto è di Tipo 0. Esistono anche Linguaggi che non sono Sensibili al Contesto, che a loro volta non sono Indipendenti dal Contesto che non sono Regolari.

Teorema. Sia G una Grammatica Sensibile al Contesto, allora G è ricorsiva. Esiste un algoritmo che determina se ogni stringa $w \in L(G)$.

Un altro importante concetto da introdurre è l'Albero di Derivazione, uno strumento di visualizzazione per descrivere lo sviluppo delle regole di produzione. Di seguito si specifica l'Albero di Derivazione per le Grammatiche Indipendenti dal Contesto.

Si consideri $G = (V_T, V_N, S, P)$ una Grammatica Indipendente dal Contesto. Un albero di derivazione per G avrà le seguenti caratteristiche:

1. Ogni nodo ha un'etichetta, che è un simbolo di V ;
2. L'etichetta della radice è S ;
3. Se un nodo, n , etichettato con A ha almeno un discendente, allora A deve essere in V_N ;
4. Se i nodi n_1, n_2, \dots, n_k sono discendenti diretti del nodo n , con etichette A_1, A_2, \dots, A_k rispettivamente,

allora

$$A \rightarrow A_1, A_2, \dots, A_k$$

deve essere una produzione in P .

Un esempio di Albero di Derivazione è il seguente

Sia $G = (\{a, b\}, \{S, A\}, S, P)$, dove P è:

$$S \rightarrow aAS \quad S \rightarrow a$$

$$A \rightarrow SbA \quad A \rightarrow ba$$

$$A \rightarrow SS$$

L'albero che ne deriva è visualizzato in Figura 1.1.

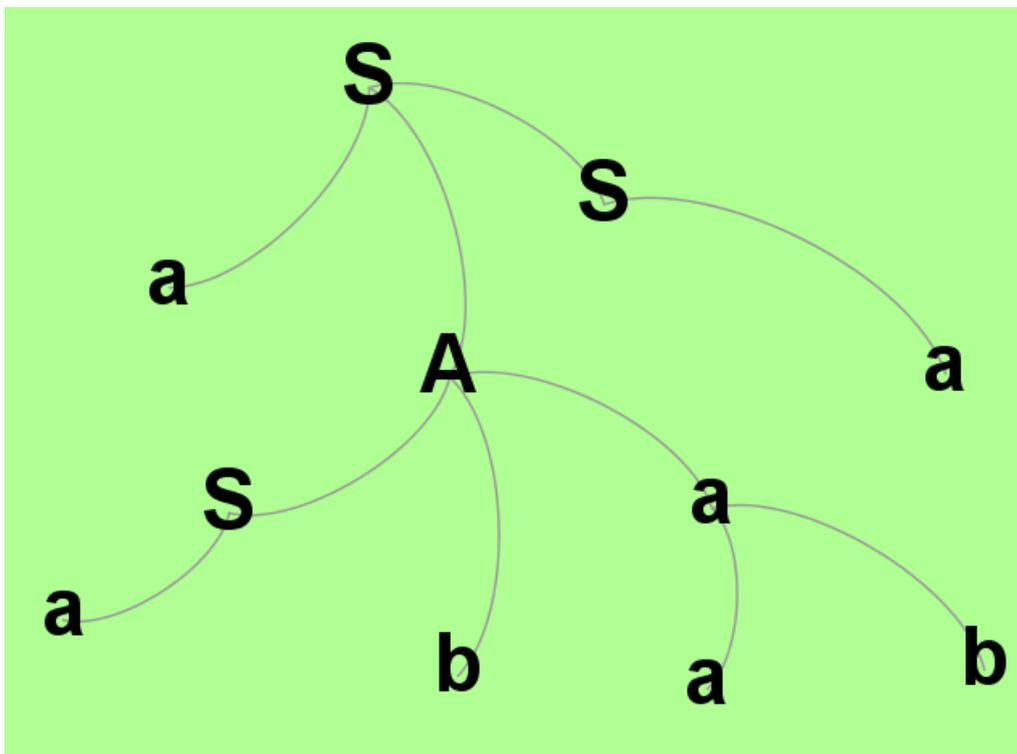


Figura 1. Albero di derivazione per una Grammatica Indipendente dal Contesto.

Si indicano come Foglie i nodi negli Alberi di Derivazione senza discendenti. Se si leggono le foglie da sinistra a destra, si ottiene una frase, detta anche risultato dell'albero di derivazione.

Capitolo 2. Aspetti formali delle reti complesse

2.1. Introduzione

I grafi sono sistemi matematici che catturano l'organizzazione strutturale e funzionale di numerosi fenomeni reali e artificiali, in diversi domini della conoscenza umana. L'idea di fondo dei grafi consiste nel collegare insieme elementi che sono in qualche modo distanti tra loro, attraverso legami di diversa natura. Sviluppato da Eulero nel 1736, il concetto di grafo è stato ripreso negli anni '60 dagli psicologi sociali (ogni persona è un nodo e gli archi sono le relazioni che gli individui instaurano tra loro) e ha poi trovato il suo sviluppo maggiore con l'avvento della rete Internet, che ha legato insieme tecnologie della comunicazione e reti sociali. Numerosi domini della conoscenza umana sono stati investigati tramite le reti. La teoria dei grafi descrive matematicamente le reti fisiche reali, realizzando metodi rigorosi e algoritmi computazionali atti a rappresentarle.

In questo capitolo si introducono i principali aspetti formali della teoria dei grafi e delle reti complesse. Gli esempi forniti sono riproducibili con diversi software. Per questo capitolo è stato utilizzato il software *Mathematica*.

2.2. Aspetti formali: Grafi e Reti

Le reti complesse sono rappresentate tramite grafi, dove una serie di *nodi* sono collegati tra loro tramite delle linee dette *archi*. Da un punto di vista formale un grafo è composto da due insiemi $G = \{X, L\}$ dove $X = \{x_1, x_2, \dots, x_N\}$ è l'insieme degli N nodi e $L = \{l_1, l_2, \dots, l_H\}$ è l'insieme degli H archi. Gli archi sono una coppia ordinata di elementi di X , cioè $l_i = (x_j, x_k)$ dove $x_j, x_k \in X$. In Figura 2.1 è riportato un grafo composto da 6 nodi e 7 archi, dove $X = \{1,2,3,4,5,6\}$ e $L = \{(1,2), (2,4), (2,3), (2,5), (3,5), (6,1), (6,5)\}$.

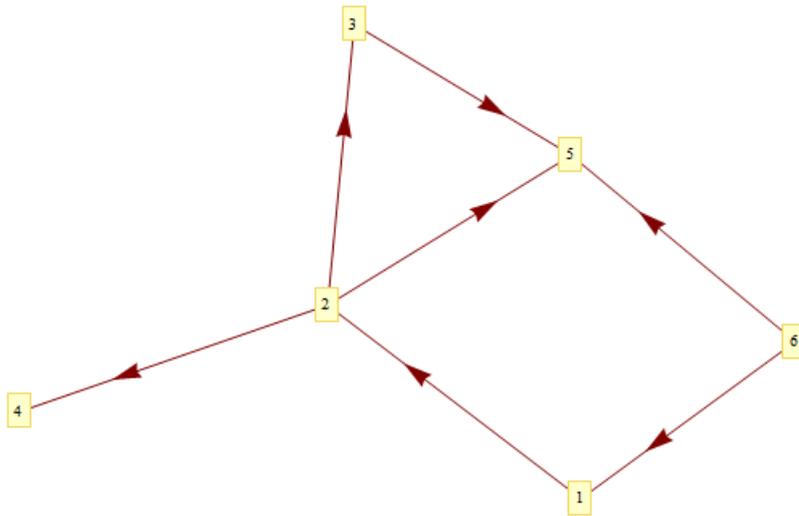


Figura 2.1. Un grafo composto da 6 nodi e 7 archi.

Se per ogni arco $l = (i, j) \in L$ esiste anche un arco $l' = (j, i) \in L$, allora il grafo è detto *indiretto*, altrimenti il grafo è detto *diretto*. Nel caso di un grafo indiretto gli archi si rappresentano tramite segmenti invece che tramite frecce (Figura 2.2).

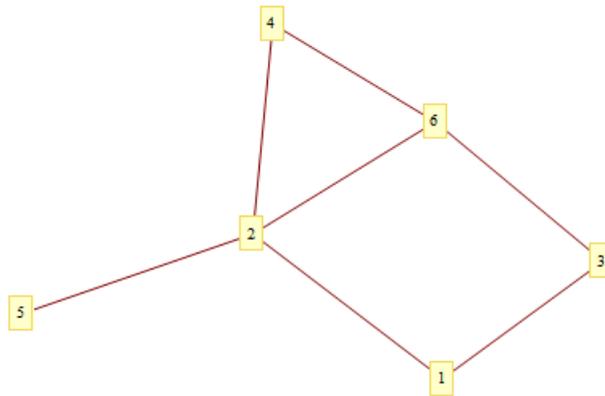


Figura 2.2. Rappresentazione di un grafo indiretto.

Un modo alternativo di rappresentare un grafo è la forma matriciale. In questo caso si costruisce una matrice A , di N righe e N colonne, dove N è il numero di nodi, detta *matrice di adiacenza*. Gli elementi a_{ij} di A , saranno uguali a 1, se esiste una coppia $(i, j) \in L$, cioè esiste un arco che unisce i e j , altrimenti $a_{ij} = 0$. Nel caso dell'esempio di Figura 2.1, la matrice di adiacenza è la seguente

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Un modo alternativo di rappresentare la matrice adiacenza è disegnare un quadrato di N^2 caselle e colorare le caselle in corrispondenza dei nodi connessi (vedi Figura 2.3).

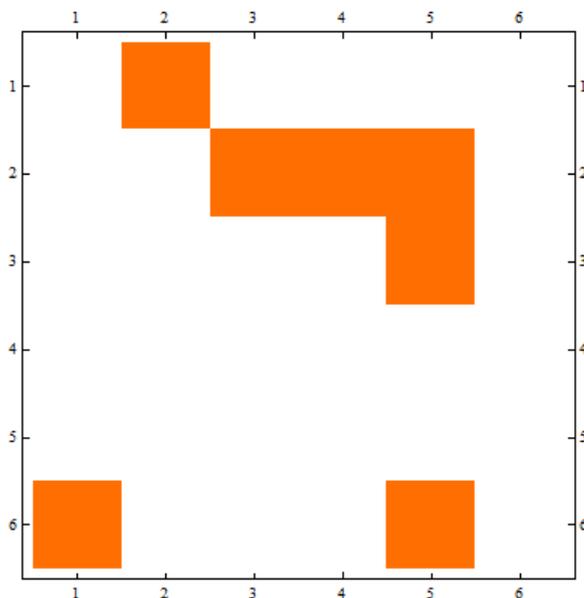


Figura 2.3. Rappresentazione grafica della matrice di adiacenza del grafo di Figura 2.1.

Un cammino è un modo per passare da un nodo origine ad un nodo termine, utilizzando gli archi del grafo G . Da un punto di vista formale, un *cammino* c è un insieme di nodi tra i quali sussiste un arco, cioè: $c = \{n_1, n_2, \dots, n_n\}$ dove gli n_i sono nodi del grafo; per ogni nodo del cammino esiste un arco uscente dal nodo che lo collega al successivo. Ad esempio, nella Figura 2.4a, esiste un cammino tra il nodo 2 e il nodo 6: $c=(2,3,5,6)$, che può anche essere scritto come $2 \rightarrow 3 \rightarrow 5 \rightarrow 6$.

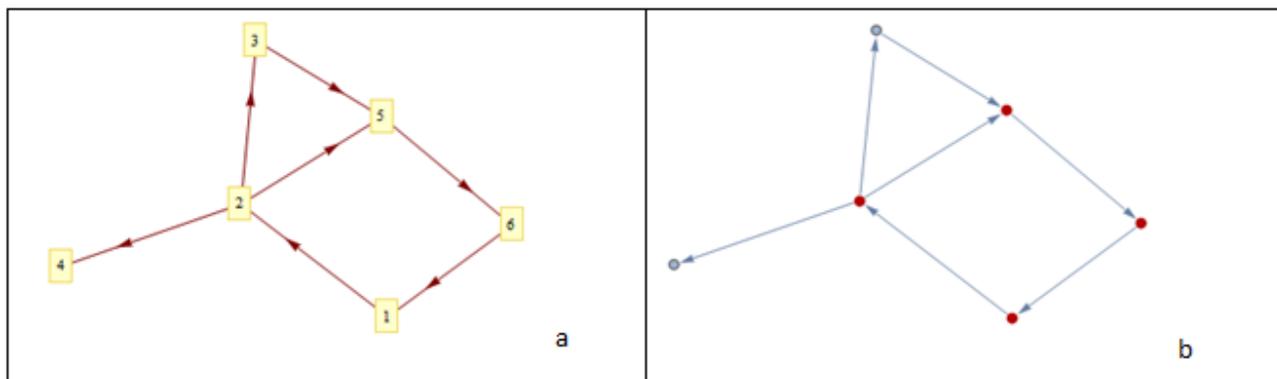


Figura 2.4. Il grafo (a) possiede il ciclo $1 \rightarrow 2 \rightarrow 5 \rightarrow 6 \rightarrow 1$. I nodi del ciclo sono indicati in rosso nella Figura (b).

La *lunghezza* di un cammino è il numero di archi che compaiono nel cammino stesso. Nell'esempio precedente questa lunghezza è 4.

Due nodi si dicono *connessi* se esiste un cammino (o percorso) che li collega. Due nodi che non sono connessi, sono detti *sconnessi*. La *lunghezza* di un percorso è il numero di archi necessari per connettere i nodi che definiscono il percorso stesso. La *distanza* d_{ij} tra due nodi i e j è la lunghezza del percorso più breve che li collega. Se non esiste alcun collegamento tra due nodi, la loro distanza è considerata convenzionalmente infinita, ed è indicata con il simbolo ∞ . Se il grafo è diretto, la distanza d_{ij} tra il nodo i e il nodo j , può essere differente dalla distanza d_{ji} che rappresenta la distanza tra il nodo j e il nodo i . Ad esempio, nel grafo di Figura 2.1, la distanza tra il nodo 1 e il nodo 2 è pari a 1, mentre la distanza tra il nodo 2 e il nodo 1 è pari ad ∞ , perché non esiste nessun cammino possibile che porti dal nodo 2 al nodo 1. Anche le distanze tra i nodi possono essere rappresentati da una matrice D i cui coefficienti sono le d_{ij} . Ad esempio nel caso del grafo di Figura 2.1, la matrice delle distanze sarà

$$D = \begin{pmatrix} 0, 1, 2, 2, 2, \infty \\ \infty, 0, 1, 1, 1, \infty \\ \infty, \infty, 0, \infty, 1, \infty \\ \infty, \infty, \infty, 0, \infty, \infty \\ \infty, \infty, \infty, \infty, 0, \infty \\ 1, 2, 3, 3, 1, 0 \end{pmatrix}$$

che può essere rappresentata tramite un diagramma analogo a quello di Figura 2.3, dove ora si sceglie una scala di colore per individuare il valore dei coefficienti della matrice. In Figura 2.5, è visualizzata una possibile rappresentazione della matrice delle distanze per il grafo di Figura 2.1. Il colore più scuro rappresenta una distanza infinita.

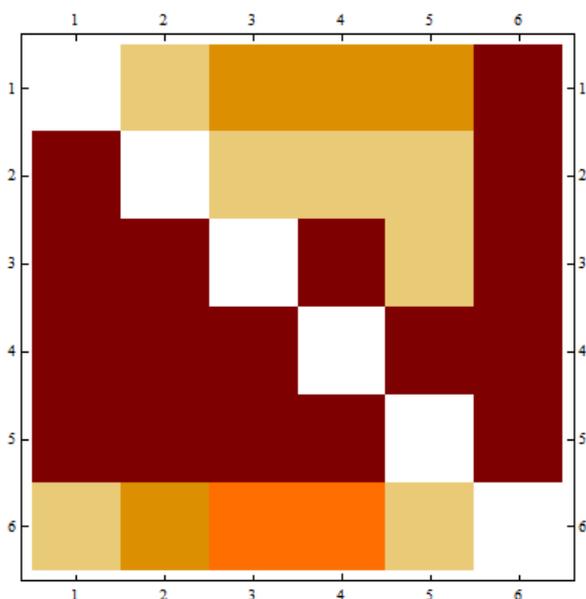


Figura 2.5. Rappresentazione grafica della matrice delle distanze del grafo di Figura 2.1.

Un *ciclo* è un cammino che inizia e termina nello stesso nodo. A differenza del grafo di Figura 2.1 che non possiede cicli, in Figura 2.3 possiamo osservare un nuovo grafo che possiede il ciclo $1 \rightarrow 2 \rightarrow 5 \rightarrow 6 \rightarrow 1$.

Per ogni nodo possiamo individuare il numero di archi entranti k_{in} (*in-degree*), il numero di archi uscenti k_{out} (*out-degree*), ed il totale dei due, detto *grado* (*degree*) del nodo, indicato usualmente con k : $k = k_{in} + k_{out}$. Per i 6 nodi del grafo di Figura 2.1, gli in-degree valgono rispettivamente 1, 1, 1, 1, 3, 0, gli out-degree valgono 1, 3, 1, 0, 0, 2; i gradi sono 2, 4, 2, 1, 3, 2. In termini dei coefficienti della matrice di adiacenza, per un nodo i , il valore dell'indegree è $k_{in}^i = a_{i1} + a_{i2} + \dots + a_{iN} = \sum_{j=1}^N a_{ij}$, il valore dell'out-degree è $k_{out}^i = \sum_{j=1}^N a_{ij}$, mentre il grado del nodo è $k_i = \sum_{j=1}^N (a_{ij} + a_{ji})$.

2.3. Grandezze Statistiche

Una serie di grandezze statistiche ci forniscono le caratteristiche strutturali delle reti complesse. Di seguito, sono riportate le principali grandezze presenti nella letteratura (Barabási,1999), che saranno utili ai fini della modellazione dei linguaggi.

a) Il Grado Medio Totale (*Average total degree*) k , rappresenta il grado medio dei nodi, ovvero il numero medio di collegamenti posseduti da un nodo. Il suo valore è definito da:

$$(2.1) \quad k = \frac{\sum_{i=1}^N k_i}{N}$$

Alcuni autori, nella formula (2.1), usano un 2 al denominatore per non contare lo stesso arco due volte, come entrante in un nodo e come uscente da un altro nodo. Nel caso del grafo di Figura 2.1, il suo valore è 2,333.

b) Si definisce la Densità ρ di un grafo, il rapporto tra gli archi esistenti nel grafo e quelli possibili. Se abbiamo un grafo composto da N nodi e H archi, è possibile connettere ciascun nodo a tutti gli altri $(N-1)$ nodi. Si escludono da questo computo i *loop*, che, come detto precedentemente, è un nodo che può essere connesso anche a se stesso, con un arco uscente dal nodo. Quindi tutti gli archi possibili saranno $N(N - 1)$. Pertanto, in un *grafo direzionato* la densità è:

$$(2.2) \quad \rho = \frac{H}{N(N-1)}$$

Nel caso dell'esempio di Figura 2.1, $H = 7$ ed $N = 6$, per cui applicando la (2.2) $\rho = 0,233$.

Nei grafi non direzionati, un arco tra due nodi può comparire un'unica volta in quanto non si distingue tra gli archi entranti e uscenti. Il numero di archi possibili è pari a $\frac{N(N-1)}{2}$. Si deduce che nei *grafi non direzionati* la densità è:

$$(2.3) \quad \rho = \frac{2H}{N(N-1)}$$

c) Si indica il Cammino medio L come la media tra tutte le distanze d_{ij} tra due nodi, non contando quelle pari a 0 e pari a ∞ :

$$(2.4) \quad L = \frac{\sum_{i=1}^N \sum_{j=1}^N d_{ij}}{M}$$

dove M è il numero di distanze diverse da 0 e da ∞ . Nel caso dell'esempio di Figura 2.1, $M = 13$ ed il valore al numeratore della formula 2.4 è pari a 21. Per cui, $L = 1,615$.

d) Il diametro d della rete è dato dalla distanza massima tra tutti i nodi. Nel caso del grafo di Figura 2.1, questa distanza è pari a 3, che rappresenta la distanza del nodo 6 dal nodo 3 e dal nodo 5.

e) Se dobbiamo inviare un segnale dal nodo i al nodo j all'interno di una rete, questo dovrà utilizzare vari archi. Una rete sarà tanto più efficiente, quanto minore è il numero di archi da attraversare e quindi è legata all'inverso dei cammini. L' Efficienza globale E_g misura in un certo senso quanto efficiente è il flusso di informazioni all'interno di una rete. Essa è definita come la media degli inversi delle distanze tra i nodi:

$$(2.5) \quad E_g = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{d_{ij}}$$

E' ovvio che in questa formula non consideriamo le distanze nulle.

f) In un grafo non tutti i nodi hanno la stessa importanza. I concetti di centralità cercano di misurare l'importanza di un nodo rispetto ad un altro. In questo senso, una grandezza significativa è la Betweenness Centrality (che potremmo tradurre liberamente con al centro del centro, nel cuore del centro). Consideriamo in tal senso un nodo i e chiediamoci quanti sono i cammini che congiungono due nodi j e k passano per i . E' ovvio che più cammini minimi passano per i più il nodo è importante. Possiamo definire allora la Betweenness Centrality di un nodo i come la somma dei rapporti tra il numero di cammini minimi tra due nodi che passano per il nodo i ed il numero di cammini minimi tra due nodi. In un certo senso, se due nodi j e k hanno un unico cammino e passa per i questo viene considerato, altrimenti non viene considerato. Nel caso in cui tra i nodi j e k esiste più di un cammino minimo, consideriamo come contributo il rapporto tra quelli che passano per i rispetto al numero totale. Da un punto di vista formale, la Betweenness Centrality b_i di un nodo i è data da

$$(2.6) \quad b_i = \sum_{\substack{j=1 \\ j \neq i}}^N \sum_{\substack{k=1 \\ k \neq i}}^N \frac{n'_{jk}}{n_{jk}}$$

dove n'_{jk} è il numero di cammini minimi che tra i nodi j e k che passano per i , ed n_{jk} è il numero di cammini minimi tra i nodi j e k . Se tra due nodi non esistono cammini possibili, allora il rapporto $\frac{n'_{jk}}{n_{jk}}$ viene posto convenzionalmente uguale a 0.

A titolo di esempio consideriamo il grafo di Figura 2.1. Dal nodo 1 passano solo i cammini che congiungono il nodo 6 agli altri nodi. In particolare da esso passano i cammini minimi $6 \rightarrow 1 \rightarrow 2$, $6 \rightarrow 1 \rightarrow 2 \rightarrow 4$ e $6 \rightarrow 1 \rightarrow 2 \rightarrow 3$ e questi cammini tra il nodo 6 e i nodi 2, 4 e 3 sono unici. Quindi la $b_1 = 3$. Dal nodo 2, oltre ai

già citati cammini, passano altri due cammini minimi $1 \rightarrow 2 \rightarrow 4$ e $1 \rightarrow 2 \rightarrow 3$ e sono gli unici cammini più brevi che uniscono 1 ai nodi 3 e 4 e quindi la $b_2 = 5$. Dai nodi 3, 4, 5 e 6 non passa alcun cammino più breve e quindi la loro Betweenness Centrality è uguale a 0.

Un'altra misura dell'importanza (centralità) di un nodo i è la Centralità di grado, che è pari a k_i . In un certo senso un nodo è tanto più centrale se più grande è il numero di archi entranti e uscenti.

L'eccentricità di un vertice i , è la lunghezza massima dei cammini possibili che uniscono un vertice a tutti gli altri vertici. Se due vertici non sono collegati, questo cammino non viene considerato. Nel caso dell'esempio del grafo di Figura 2.1, il valore dell'eccentricità di ogni vertice è pari a 2, 1, 1, 0, 0, 3 rispettivamente.

g) Il Coefficiente di Clustering misura le aggregazioni presenti nella rete. Dato un vertice i , è possibile definire l'intorno I del nodo i come l'insieme di tutti gli s nodi j che hanno distanza 1 dal nodo i . Quindi $I = \{n_1, n_2, \dots, n_j, \dots, n_s / a_{ij} \neq 0 \vee a_{ji} \neq 0\}$. Questo intorno ha dei cluster se esistono dei collegamenti tra i nodi dell'intorno differenti da quelli con i . Per i grafi diretti, il coefficiente di clustering locale c_i può allora essere definito come:

$$(2.7) \quad c_i = \frac{\sum_{j=1}^s \sum_{m=1}^s a_{ij} a_{jm} a_{mi}}{k_i(k_i - 1)}$$

Per un grafo non direzionato questo valore va moltiplicato per 2.

Il coefficiente di clustering dell'intera rete può allora essere definito come la media aritmetica dei coefficienti di clustering:

$$(2.8) \quad C = \frac{\sum_{i=1}^N c_i}{N}$$

Se consideriamo la rete di Figura 2.1, si osserva che $c_1 = c_4 = c_6 = 0$, infatti i nodi e i loro intorni non producono triangoli. L'intorno del nodo 2 è composto da 4 nodi e sviluppano un solo e quindi $c_2 = \frac{1}{12}$, essendo $k_2 = 4$. $c_3 = \frac{1}{2}$ e $c_5 = \frac{1}{6}$ e quindi

$$C = \left(\frac{1}{12} + \frac{1}{2} + \frac{1}{6}\right) \frac{1}{6} = \frac{1+6+2}{72} = \frac{1}{8} = 0,125$$

2.4. Modelli di rete

L'attuale teoria delle reti descrive una varietà di sistemi fisici, chimici, biologici, sociali e tecnologici. Le comunità scientifiche stanno utilizzando le reti complesse per modellare questi sistemi studiando la loro topologia, l'evoluzione e la struttura. Tradizionalmente, lo studio delle reti complesse utilizza la teoria dei grafi (Clark & Holton, 1995). I principali modelli di rete attualmente disponibili sono quelli di Erdős & Rényi

(1959), Watts & Strogatz (1998) e Barabási & Albert (2000), e sono noti come modelli di *grafi random*, *piccolo mondo* e *scale-free*.

A Partire dagli anni Cinquanta, due matematici ungheresi Erdős & Rényi (1959) hanno identificato grafici casuali, con nessuna organizzazione apparente. Questo modello, dominante per anni nella letteratura di settore, considera un grafo di N nodi ed una probabilità p che due coppie di nodi siano connessi fra loro. Questo porta ad una rete i cui archi H sono circa $\frac{pN(N-1)}{2}$ e questi collegamenti sono distribuiti in modo casuale. In Figura 2.6, sono riportati due grafi di 50 nodi e 100 archi, costruiti in modo casuale.

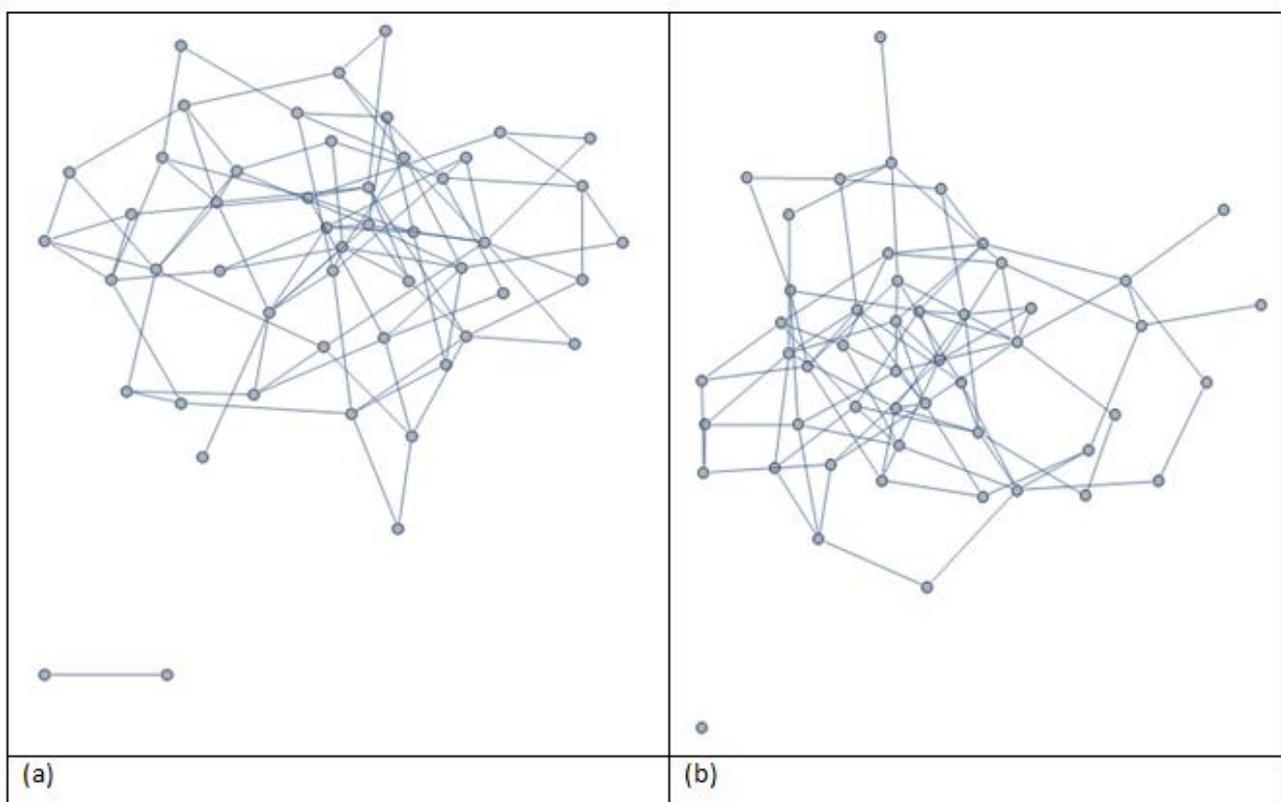


Figura 2.6. Due grafi random di 50 vertici e 100 archi, realizzati tramite il comando `RandomGraph[{50, 100}]` di Mathematica. Sia la rete (a) che la rete (b) non sono connesse. La prima ha due punti isolati, connessi fra loro, mentre la seconda ha un solo punto isolato.

Ma l'interesse nella modellazione dei fenomeni biologici, tecnologici e fisici ha portato all'idea che i sistemi complessi abbiano principi organizzativi, incorporati nelle loro strutture. Se i modelli non sono casuali, forse altri strumenti quantitativi sono necessari per analizzare l'organizzazione strutturale di questi sistemi. Uno di questi concetti è il concetto di small-world (piccolo mondo), che sottolinea l'emergenza di sistemi organizzati e aggregati (cluster) all'interno delle reti. Il concetto di clustering o di aggregazione affonda le sue radici in sociologia (Wassermann & Faust, 1994). Milgram (1967) ha dimostrato che, nelle comunità

sociali, ogni individuo è separato da un altro in media da sei gradi di separazione. Oggi diremmo che, se ad ogni individuo corrisponde un nodo, un arco e una qualche forma di relazione tra individui, il diametro di questa rete è 6. La natura di questa rete non è casuale: nella sua struttura sono presenti molti cluster (ad esempio, gli amici di un amico sono quasi sempre anche amici tra loro, creando le *clique*, ovvero sotto-grafi con ogni nodo collegato all'altro). Insieme di cluster formano le *comunità* o *reti sociali*, ovvero aggregazioni più vaste di gruppi di soggetti altamente connessi tra loro, mentre il diametro di questa rete non cambia in modo sostanziale, rispetto alle reti casuali. La presenza di cluster non va a discapito dell'efficienza della rete. Questo tipo di rete sembra caratterizzare diversi sistemi, in domini diversi, dalla rete di attori di Hollywood alle reti telefoniche, alle reti biologiche, etc. Il coefficiente di clustering fu introdotto da Watts & Strogatz nel 1998 per misurare l'aggregazione di un gruppo di nodi che compongono una rete. Un coefficiente di clustering vicino a 1 indica che il gruppo è molto compatto, ma se tende a 0, il gruppo è molto frammentato. Watts & Strogatz hanno quindi proposto un nuovo modello di reti che prevede l'esistenza di raggruppamenti, i cui nodi hanno molte strutture in comune, mentre i legami tra nodi appartenenti a differenti gruppi sono molto ridotti. In queste reti, i nodi non sono collegati né in maniera ordinata né casuale; queste reti stanno tra due estremi (ordinati- casuali) e sono dette di piccolo mondo. La Figura 2.7 presenta una rete di piccolo mondo, costruita sul modello di Watts & Strogatz.

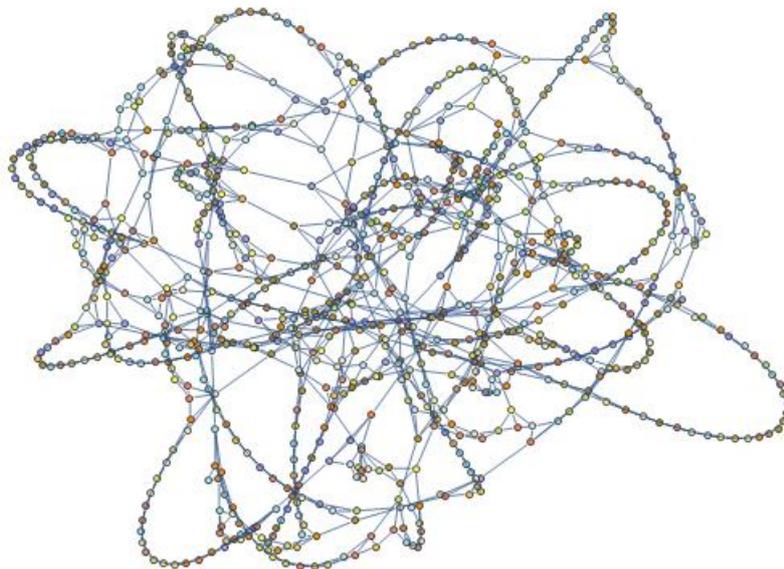


Figura 2.7. Rete di piccolo mondo costruita sul modello di Watts & Strogatz costruita con 900 nodi e 1800 archi. Questa rete è stata generata con il software Mathematica.

Le reti di piccolo mondo possono essere create in modo semplice. Si parte da una rete ordinata composta da N nodi nella quale un nodo i è collegato ai suoi due vicini di destra $i+1$, $i+2$ e di sinistra $i-1$, $i-2$; a loro volta, solo i vicini prossimi sono collegati fra loro. Per questi reti si considera una condizione al contorno

periodica, nella quale il vicino di sinistra di 1 è il nodo N e il vicino di destra del nodo N è 1. In Figura 2.8, si riporta una rete di questo tipo, composta da 5, 10 e 20 nodi rispettivamente.

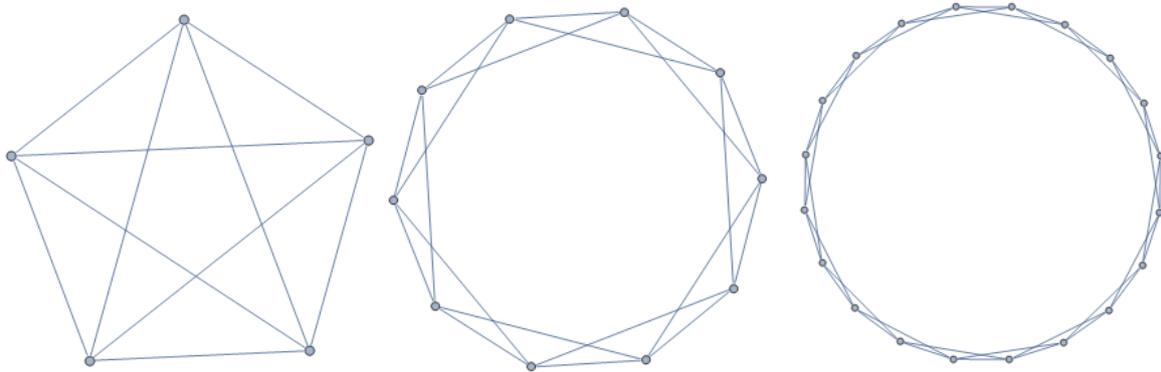


Figura 2.8. Una rete di 5, 10 e 20 nodi nella quale ogni nodo è collegato con i suoi vicini prossimi, che a loro volta sono collegati fra loro. Il numero di archi è 10, 20 e 40 rispettivamente.

I coefficienti di clustering delle tre reti di Figura 2.8 sono molto alti (1 per la prima rete e 0.5 per le altre due). In realtà questo valore si stabilizza su 0.5 perché nel vicinato solo il 50% dei triangoli possibili sono presenti. Il problema di queste reti è la loro scarsa efficienza; infatti, il diametro, così come la lunghezza media, crescono molto velocemente. Per la prima rete il diametro e la lunghezza media sono pari a 1; nella seconda rete valgono 3 e 1,667. Nella terza rete diventano 5 e 2,895. In una rete di 100 nodi i valori diventano addirittura 25 e 12,879. In una rete di 900 nodi questi valori sono diventati 225 e 112,875. Quindi, più grande è il numero di nodi, più i cluster intaccano l'efficienza della rete, con cammini troppo lunghi. Il problema si risolve prendendo alcuni degli H archi della rete e, invece di usarli per collegare nodi prossimi, si collegano a caso nodi anche distanti dal nodo dato. Ad esempio, la rete a 900 nodi di Figura 2.8 è ottenuta tramite una riscrittura con probabilità $p = 0,05$. In questo caso, il diametro della rete si è ridotto a 24, e la lunghezza media è pari a 11,664. Anche il coefficiente di cluster è diminuito a 0,438, perché alcuni archi sono stati utilizzati non per formare triangoli, ma per collegare nodi distanti. La Figura 2.9 presenta una rete di 50 nodi, i cui archi sono riscritti con probabilità 0,03, 0,07, 0,2 e 0,3 rispettivamente.

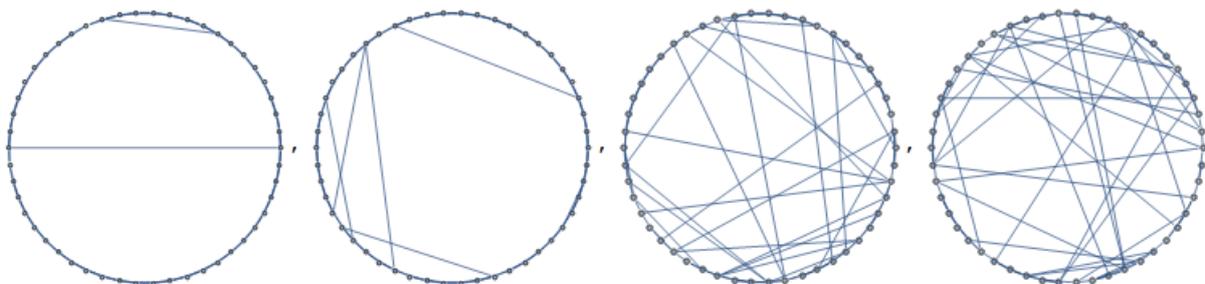


Figura 2.9. Rete di 50 nodi i cui archi sono riscritti con probabilità 0,03, 0,07, 0,2 e 0,3 rispettivamente.

Se per la rete di 900 nodi si aumenta la probabilità di riscrittura a 0,2, si ottiene una ulteriore diminuzione del diametro, pari a 12 e una lunghezza media pari a 6,75. Il coefficiente di clustering diminuisce a 0,259.

Riassumendo, una rete ordinata ha un alto coefficiente di clustering, ma una grande lunghezza media; una rete random ha un basso coefficiente di clustering, ma una bassa lunghezza media; ci si aspetta che una rete di piccolo mondo abbia un alto coefficiente di clustering, ma una lunghezza media non molto più bassa di una rete random corrispondente, cioè che consenta aggregazioni come nelle reti ordinate, ma mantenga un'efficienza paragonabile a quella di una rete random.

Come possiamo sapere se una rete può essere classificata di piccolo mondo? Possiamo rispondere a questa domanda calcolando tre nuovi parametri per la rete

$$(2.9) \quad \gamma = \frac{C}{C_{rand}}$$

$$(2.10) \quad \lambda = \frac{L}{L_{rand}}$$

$$(2.11) \quad \sigma = \frac{\gamma}{\lambda}$$

dove C_{rand}, L_{rand} sono i valori medi di una rete random che abbia lo stesso numero di nodi e lo stesso numero di archi della rete sotto esame. Affinché una rete sia di piccolo mondo allora, in accordo con la definizione data da Watts & Strogatz, dovrà risultare: $\gamma \gg 1$ (cioè un numero di cluster più grandi rispetto a quello di una rete random), $\lambda \approx 1$ (cioè una lunghezza media paragonabile a quella di una rete random) e $\sigma > 1$. σ rappresenta, in un certo senso, una misura di quanto la rete è di piccolo mondo.

Un anno dopo il lavoro di Watts & Strogatz, Barabási & Albert (1999) si accorsero che un grande numero di reti reali presentavano pochi nodi molto connessi e molti nodi poco connessi, cioè il valore di k_i non seguiva una distribuzione gaussiana, come ci si sarebbe aspettato nelle reti random o nelle reti di piccolo mondo, ma la distribuzione di grado seguiva una legge di potenza, completamente diversa dalla distribuzione normale, che può essere espressa come:

$$(2.12) \quad P(k) = ak^{-\beta}$$

Il significato di una legge di questo tipo individua un particolare modello di crescita della rete stessa: cioè un nodo i che possiede un grado k_i tenderà ad aumentare il suo grado man mano che la rete cresce in misura tanto maggiore quanto maggiore è il suo grado. Quindi i nodi più connessi diventeranno via via più connessi. Al contrario, i nodi con un minor numero di connessioni aumenteranno i legami assai lentamente. Questo vale ad esempio per il World Wide Web, (Albert et al., 1999), per Internet (Faloutsos et al., 1999), e per le reti metaboliche (Jeong et al., 2000). Barabási e collaboratori (Barabási et al., 1999) definirono queste reti *scale free*, in quanto presentano le tipiche invarianze di scala delle strutture frattali.

La (2.12) può essere scritta in forma logaritmica come

$$(2.13) \quad \log P(k) = \log a - \beta \log k$$

e quindi su scala logaritmica otterremo una retta che lega il logaritmo della probabilità della sua distribuzione al logaritmo del grado.

In questo tipo di reti emergono naturalmente dei nodi con grado altissimo, che sono detti *hub*. Lo studio e la distribuzione degli *hub* è di grande interesse pratico perché è strettamente legato all'efficienza della rete. La rimozione di un *hub* da una rete non avrà lo stesso effetto della rimozione di un qualsiasi altro nodo dalla rete.

Queste reti possono essere create facilmente. Ad esempio, la rete di Figura 2.10 è stata creata a partire da un grafo di tre nodi connessi tra loro, aggiungendo un nodo con due archi ad ogni passo. I due archi sono attaccati a caso agli altri nodi, seguendo una distribuzione proporzionale al grado del nodo. Gli *hub* sono chiaramente visibili perché hanno le dimensioni più grandi rispetto agli altri nodi.

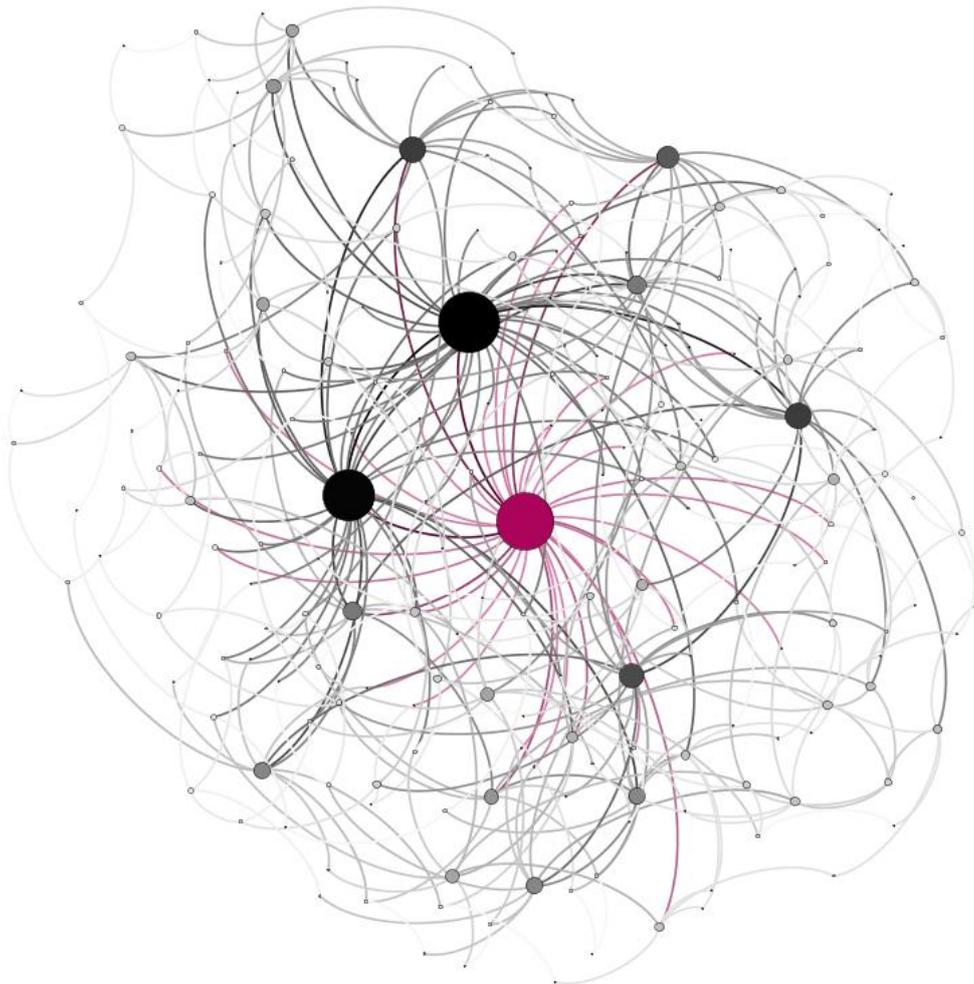


Figura 2.10. Una rete di 200 nodi generata tramite il modello di Barabási con un attaccamento preferenziale.

Nella Figura 2.11, si presenta come si distribuiscono i nodi in funzione del grado.

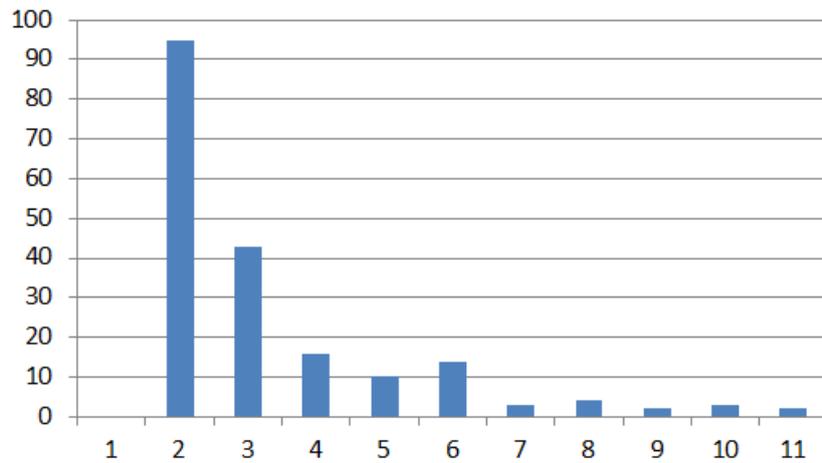
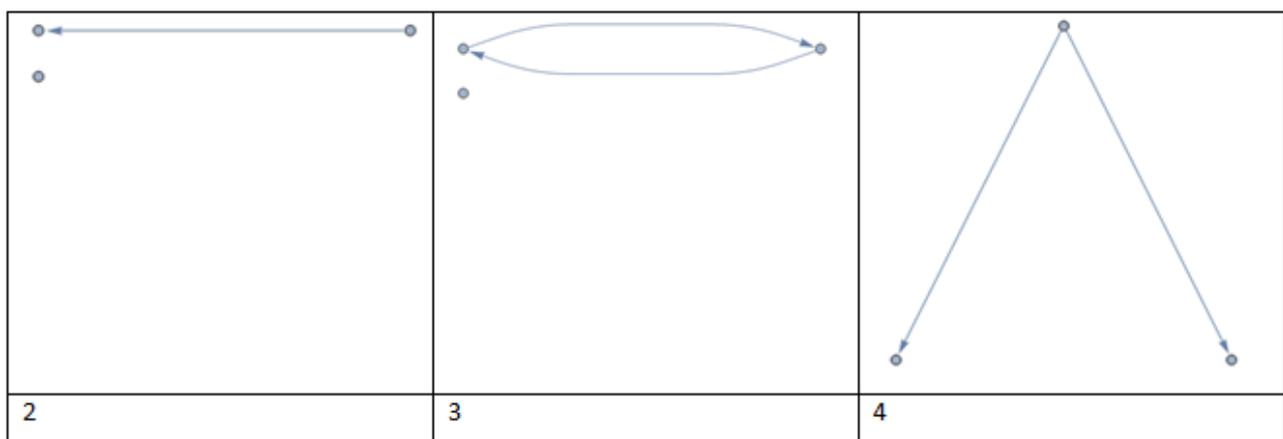
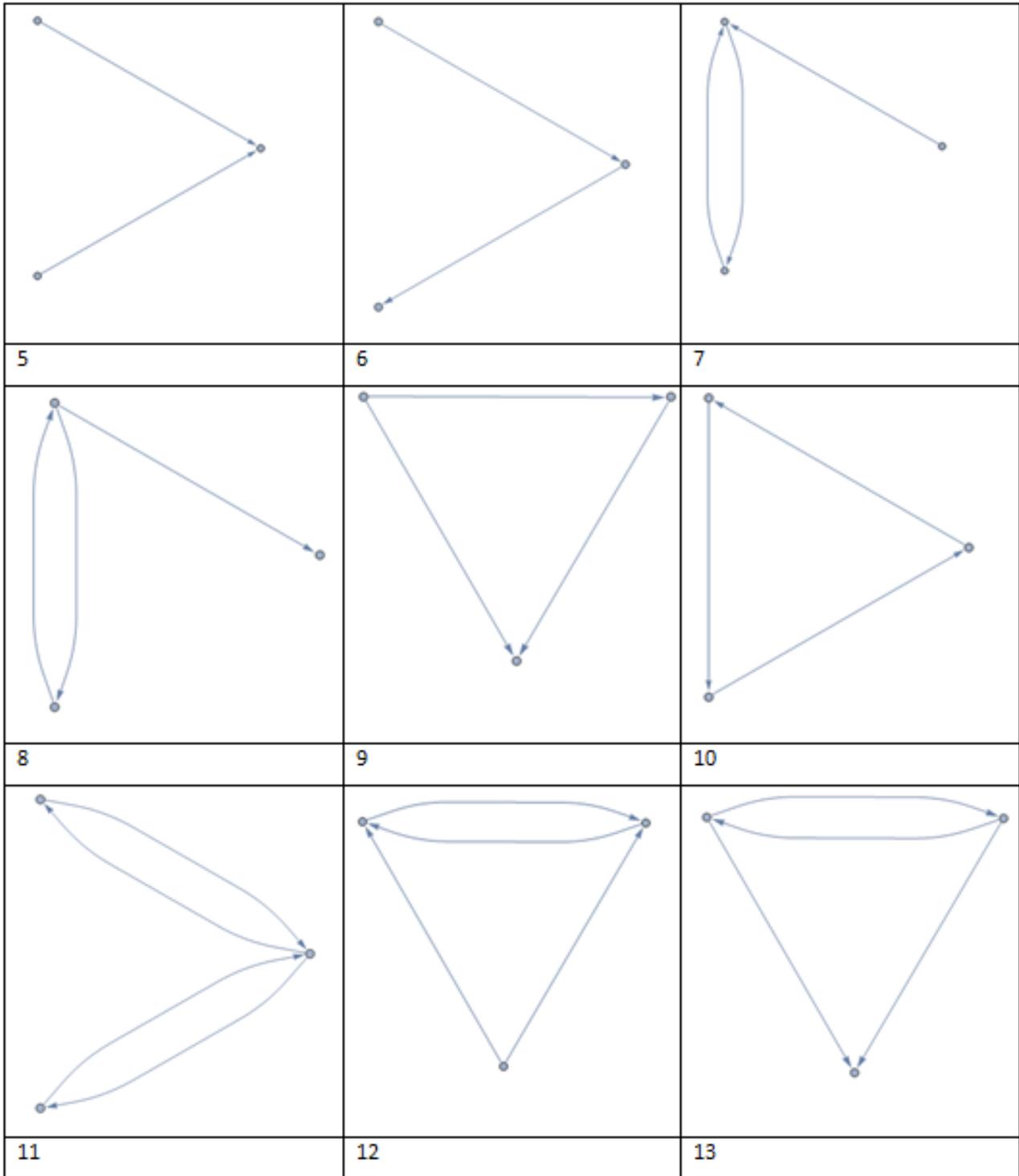


Figura 2.11. Distribuzione dei nodi in funzione del grado nella rete della Figura 2.10 seguendo il modello di distribuzione di Barabási.

2.5. Motivi

All'interno di una rete, come si rapportano tra loro più nodi attraverso gli archi che intercorrono tra essi? Come si trasferisce l'informazione da un nodo all'altro? Si è già evidenziato come la presenza di triangoli individua i cluster e le comunità. Ma in contesti biologici queste relazioni possono avere anche particolari significati e individuare cooperazione genica, processi di feed-back, connessioni cerebrali, ecc. (Milo et al., 2002; Sporns & Kotter, 2004), oltre ad indurre nella rete direzioni preferenziali di organizzazione e di crescita (Itzkovitz & Alon, 2005). E' quindi opportuno individuare queste strutture di base note come *motivi* e determinarne le caratteristiche.





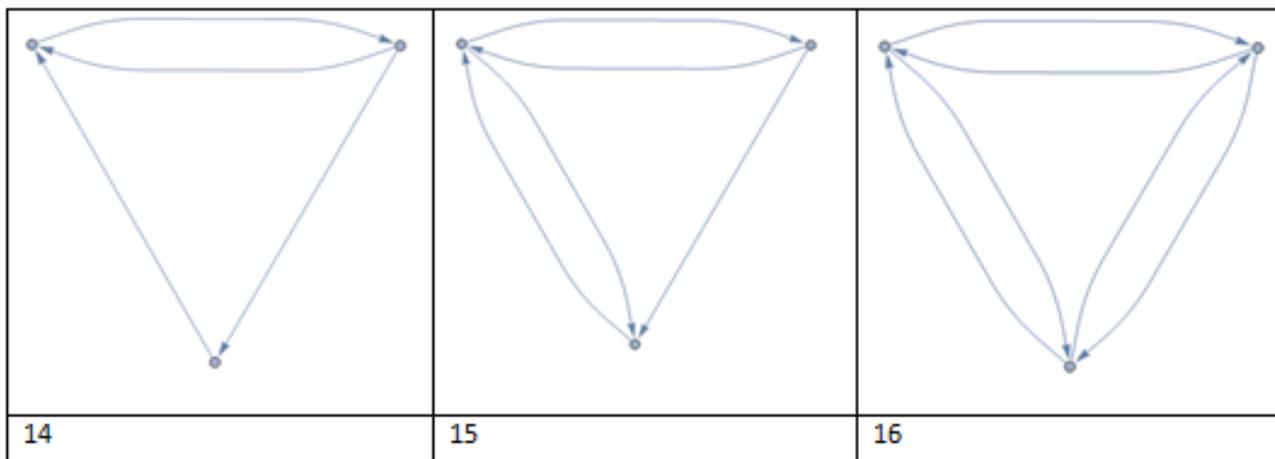
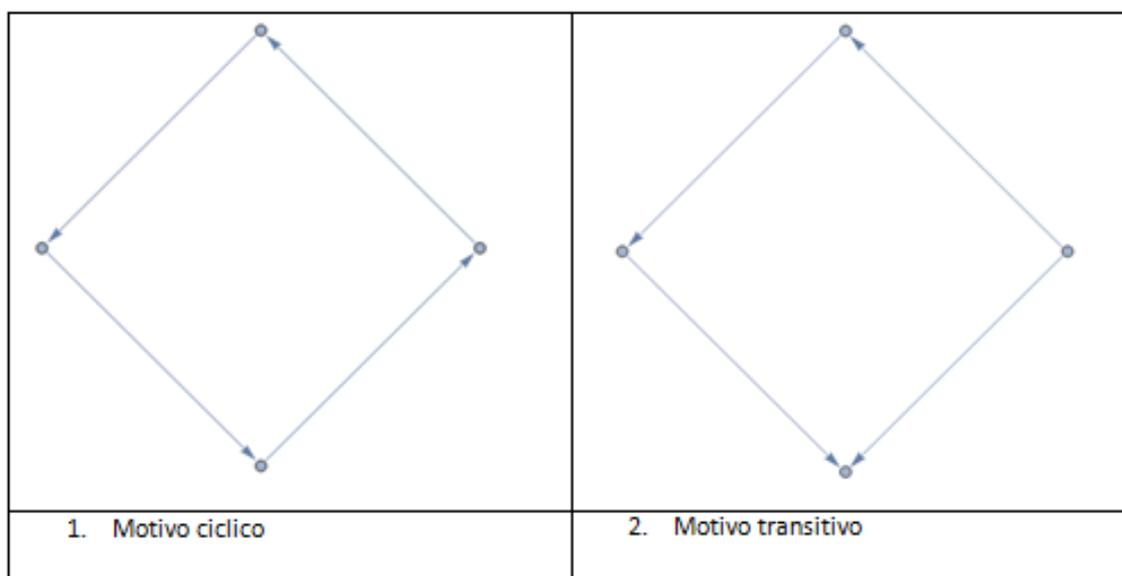


Tabella 2.1. In questa Tabella sono riportati tutti i possibili motivi costituiti da 3 nodi.

Tre punti possono organizzarsi in 16 modi diversi. In Tabella 2.1. sono evidenziati 15 modalità, numerate da 2 a 16. La modalità 1 è relativa a tre punti isolati, motivo per il quale non viene riportata. L'esempio è stato realizzato con Pajek (Wouter de Nooy et al., 2005), per cui è stata usata la stessa terminologia. Le 5 triadi, da 12 a 16, sono responsabili dei clustering all'interno della rete. Queste triadi sono rappresentazioni *ideali* di come effettivamente i vari nodi *reali* si dispongono e interagiscono fra loro. Ad esempio, la triade 10 rappresenta un *ciclo*, mentre la 9 una *transitività*.

Accanto a motivi di 3 nodi, si possono considerare motivi che coinvolgano insieme maggiori di elementi di una rete. In Tabella 2.2 sono riportati alcuni motivi realizzati da 4 nodi, che rivestono particolare interesse, in quanto le dinamiche che realizzano sono completamente diverse.



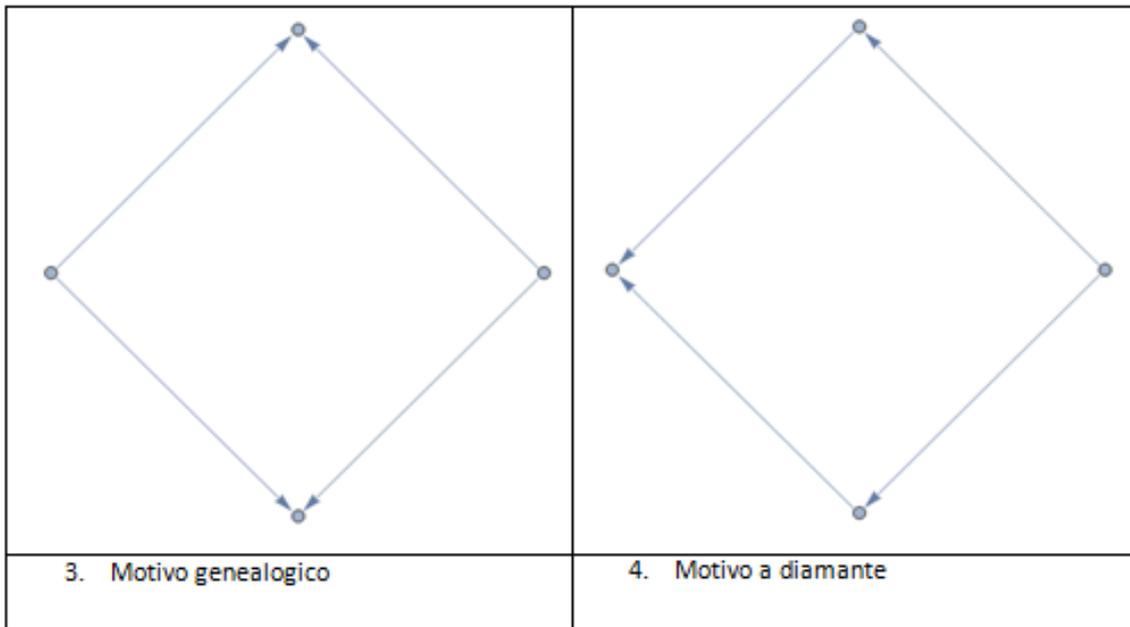


Tabella 2.2. Alcuni motivi interessanti costruiti con 4 nodi.

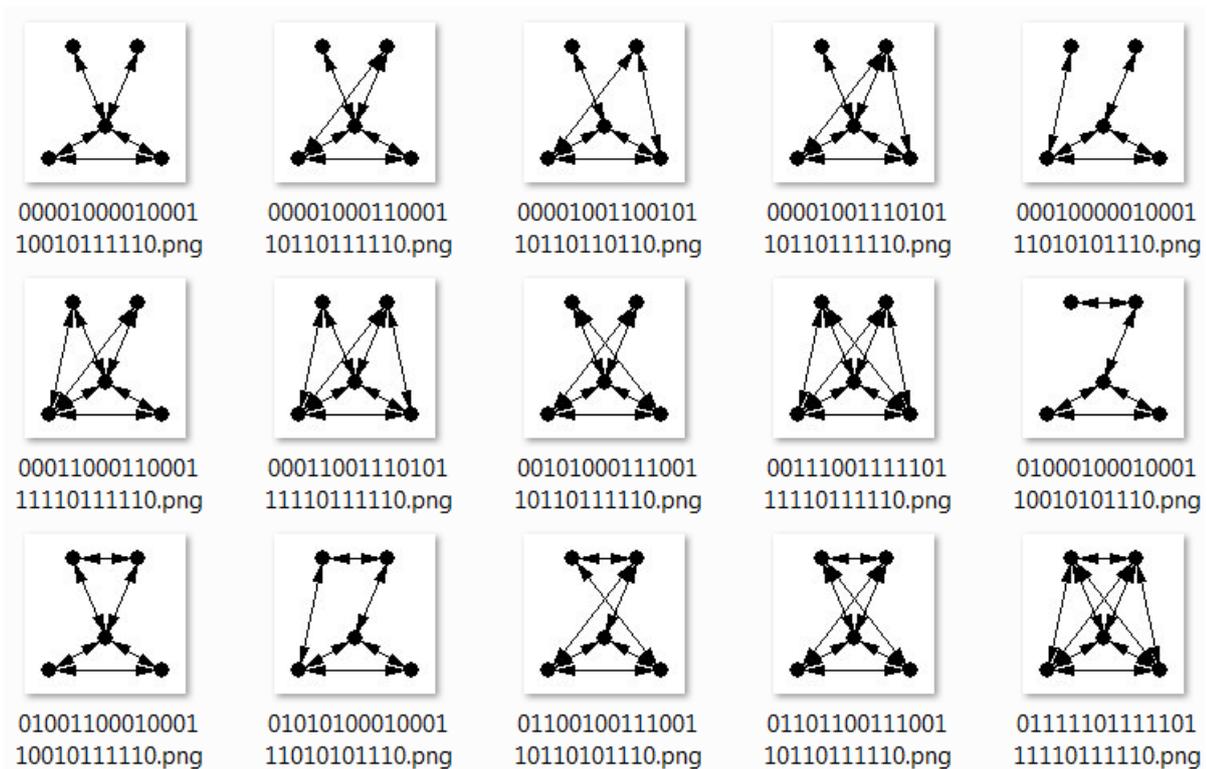


Figura 2.12. Motivi presenti una rete musicale ottenuti tramite il software FanMod. I codici numerici fanno riferimento ad opportune codifiche tipiche di quel software.

Aumentando il numero di punti, è possibile ottenere una grande quantità di motivi interessanti, la cui interpretazione nel contesto può fornire molti e variegati spunti di riflessione. I motivi saranno infatti uno strumento di analisi delle organizzazioni delle reti nei tre contesti analizzati. Molto rilevanti ai fini della loro

organizzazione complessa sono i motivi rilevati nella lingua naturale, oltre ovviamente ai motivi ritrovati in musica e nel cervello biologico. In Figura 2.12, sono riportati alcuni motivi di 5 nodi trovati all'interno di una rete musicale, utilizzando il software FanMod (Wernicke, 2005; Wernicke & Rasche, 2006). E' importante osservare che la ricerca di motivi è un problema NP-compleso e la richiesta di calcolo aumenta all'aumentare della dimensione della rete.

E' del tutto evidente l'interesse dei motivi ciclici come il numero 10 della Tabella 2.1, o il numero 1 della Tabella 2.2.

Completiamo questa sezione introducendo le *clique*, cioè grafi completi di N vertici, dove ogni vertice è legato a tutti gli altri vertici. Noi abbiamo già incontrato delle clique com'è ad esempio il motivo 16 della Tabella 1 o il primo grafo della Figura 2.8, composti rispettivamente da 3 e 5 punti. Nella Figura 2.13 osserviamo clique composte da grafi indiretti di 4, 8 e 16 punti.

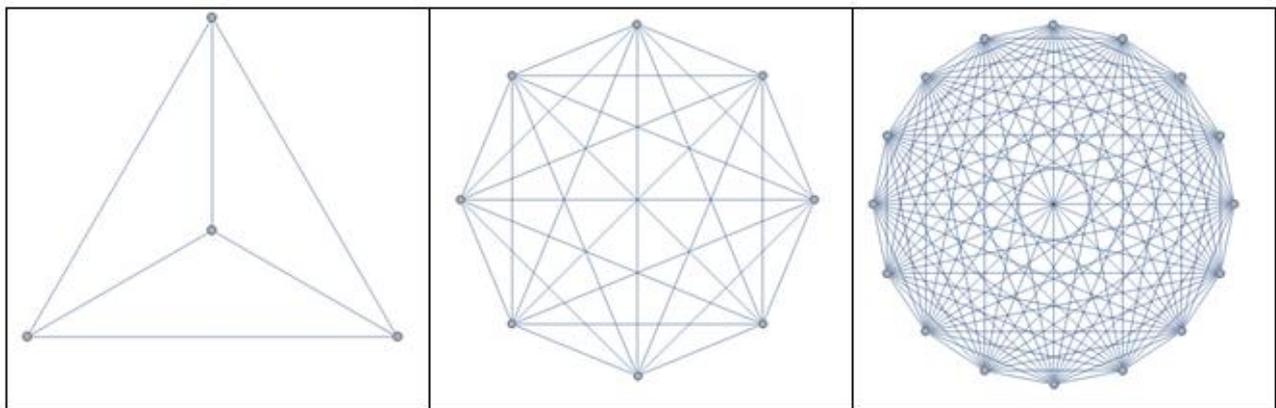


Figura 2.13. Clique composte da 4, 8 e 16 nodi

Motivi, clci e grafi giocheranno un ruolo centrale nella trattazione successiva, perché sono le strutture portanti delle reti dei linguaggi che sono stati analizzati in questa tesi.

Capitolo 3. Modellazione della Musica

3.1. Introduzione

Considerata fra “le più misteriose abilità dell’uomo” (Darwin, 1871), la musica, come il linguaggio, ha caratteristiche generali, tra le quali le più significative sono:

- a. è straordinariamente antica, i primi strumenti risalgono al Paleolitico (Wallin et al., 2000; Mithen S., 2006);
- b. è universale per tutte le culture, organizzata come una grammatica, a diversi livelli gerarchici, e con unità che si relazionano fra loro per formare strutture sempre più articolate (Baroni & Jacobini, 1978; Lerdhal & Jackendoff, 1983);
- c. è un tipico tratto umano nella comunicazione biologica (Peretz, 2006);
- d. è acquisita nella prima infanzia in modo spontaneo, attraverso uno stretto rapporto di comunicazione, su base non verbale, con la madre o il care-taker (Sundberg & Lindblom, 1991; Trevarthen & Aitken, 2001; Thaut, 2005);
- e. è un sistema adattivo, su base innata, che coinvolge tutti i processi cognitivi, dalla percezione all’apprendimento, dall’emozione all’azione (Meyer, 1956; Krumhansl, 1997; Pinker, 1997; Levitin, 2010).
- f. si sviluppa sugli scambi sociali (Malloch, & Trevarthen, 2010), in quanto ha una profonda base armonica, collegata ai ritmi biologici del cervello e del corpo (per cui contiene in sé elementi di dinamica non lineare (Patel & Balabam, 2000; Garcia et al., 2006), collegati ai modelli matematici sottesi a tali dinamiche.

La musica è un fenomeno biologico e culturale complesso, il cui sviluppo, nei soggetti umani, può essere paragonato a quello dell’acquisizione di una lingua naturale. Alcuni ricercatori sostengono che esista una base neurofisiologica, comune per il linguaggio e la musica, e che le strutture cerebrali dedicate alla cognizione musicale siano funzionalmente simili alle architetture cognitive che gestiscono i processi linguistici. La connessione essenziale che lega il linguaggio e la musica è stata dimostrata da alcune ricerche

che utilizzano strumenti di *brain imaging*, e che hanno analizzato l'influenza della musica sul cervello e hanno rilevato che musicisti esperti usano aree cerebrali largamente disperse ma interconnesse, quando si concentrano su differenti aspetti di un pezzo musicale, quali il ritmo, la melodia e l'armonia (Patel, 2003; Patel et al., 1998). Altri ricercatori (Parsons, 2003; Levitin & Menon, 2003) hanno messo in evidenza che esiste un'area nell'emisfero cerebrale destro, che funzionalmente serve per leggere le note musicali scritte e passaggi di note, che è analoga all'area collaterale sinistra, nota per essere responsabile per l'interpretazione delle lettere e delle parole scritte (vedi Capitolo 4, sulle funzioni cerebrali). Le ricerche hanno messo in evidenza che le strutture cerebrali deputate alla cognizione musicale, e l'uso che di esse fanno i soggetti umani, sono simili alle strutture cerebrali deputate alle funzioni linguistiche. La regione destra del cervello processa le note e i passaggi musicali e corrisponde alla regione sinistra, che processa lettere e parole. Il funzionamento di tali meccanismi neurali, presenti in ognuno dei due emisferi cerebrali, è adatto per simili scopi cognitivi, ma con differenti informazioni o contesti. Tutti questi lavori dimostrano che le risposte musicali sono gestite da aree largamente distribuite in tutto il cervello, piuttosto che localizzate in una singola regione, come altre funzioni quali la visione o il movimento. La conclusione è che esiste una similarità fra musica e linguaggio (Maess et al., 2001; Panksepp, 2003; Levitin, 2006) e l'evoluzione della cognizione musicale è connessa con lo sviluppo del linguaggio.

Così come esistono milioni di lingue parlate, gli umani hanno sviluppato differenti sistemi musicali, in culture diverse. I soggetti umani amano la musica, la cercano attivamente, hanno comportamenti d'ascolto tipici, la possono ascoltare come sottofondo piacevole, anche quando sono impegnati in altre attività. Dallo stadio prenatale in poi siamo sottoposti all'influenza che il parlato, i suoni e la musica, presenti nell'ambiente, esercitano su di noi. L'acquisizione del linguaggio musicale si sviluppa nel tempo. Tale processo implica, come nel caso delle lingue naturali, la competenza nel riconoscimento e nella produzione di un'organizzazione di distinzioni riccamente strutturate di cui il linguaggio musicale, ma anche il sottostante livello sonoro, sono dotati. I soggetti imparano semplicemente ascoltando e riproducendo motivi e ciò implica che il cervello umano abbia una specializzazione, geneticamente determinata, che è finalizzata al linguaggio musicale. Per cui, si può dire che le nostre capacità musicali siano una complessa combinazione di natura e cultura, poiché come per il linguaggio, esse si strutturano proprio dall'interazione dinamica tra il soggetto e l'ambiente in cui vive. L'interazione con l'ambiente è complementare alla dotazione genetica. Un bambino acquisisce la capacità di gestire un fenomeno complesso come il linguaggio e ad assimilarlo in un tempo relativamente breve.

Il dominio della cognizione musicale (composizione, percezione e esecuzione) è estremamente complesso e influenzato dalle idee di Chomsky (1965). Come altre lingue naturali, la musica è un sistema di regole (una grammatica generativa, la teoria generativa della musica tonale o Generative Theory of Tonal Music, GTTM) con una struttura superficiale e una struttura profonda (Lerdahl & Jackendoff, 1983a, b). La presenza di

strutture universali nella musica permette di “sintonizzarsi” anche su brani musicali lontani dalla tradizione culturale cui un soggetto appartiene. Sfortunatamente, non esiste alcuna testimonianza dell’evoluzione filogenetica del linguaggio e della musica, se non i resti di strumenti dei nostri antichi progenitori (Kunej e Turk, 2000). Numerosi studi simulano l’evoluzione del linguaggio, da un primitivo linguaggio sonoro, attraverso reti neurali (Bharucha & Todd, 1989), e attraverso agenti che interagiscono con l’ambiente (Steels, 2003). La maggior parte delle regole che controllano l’evoluzione della cognizione musicale sono ancora sconosciute (Fitch, 2006), anche se il ruolo centrale di molti degli universali sono stati rilevati (Dowling & Harwood, 1986). Ma una teoria completa della produzione musicale è ancora carente.

La natura interna della musica è riferibile ad una grammatica, condividendo molti aspetti formali con modelli matematici. La ricerca di proprietà matematiche, statistiche e cognitive, (Wiggins, 2012; Colton & Wiggins, 2012) della musica è migliorata di recente nei sistemi di Intelligenza Artificiale (Backer & Van Kranenburg, 2005; Bellaachia & Jimenez, 2009; Barutcuoglu & De Coro, 2006), che si possono anche incorporare in nuove applicazioni musicali, basate su un uso non convenzionale dei sistemi dinamici (Bertacchini et al., 2007; Cupellini et al., 2008), da usare anche in classe con studenti di scuola secondaria, per sviluppare la motivazione e la creatività (Bertacchini et al., 2012a, b, c; Bertacchini et al., 2013). Resta comunque molto difficile comprendere l’organizzazione sfuggente e complessa delle strutture profonde della musica (Bilotta et al., 2005).

E’ possibile svelare le caratteristiche principali della musica di Bach? Può la teoria delle reti essere utile a questo scopo? In questi ultimi decenni, l’approccio metodologico delle reti è stato usato per descrivere una grande varietà di sistemi, che vanno dal World Wide Web (Barabási & Albert, 1990) al linguaggio umano (Ferrer i Chanco, 2001), alla musica (Liu et al., 2010), con l’idea di cogliere i principi organizzativi che sistemi diversi incarnano nelle loro strutture. In questo capitolo, attraverso una formalizzazione e trasformazione dei file MIDI (noto sistema di interfacciamento musicale) in file .MET, le melodie musicali di Bach sono rappresentate come reti e le caratteristiche strutturali di queste reti sono esplorate. Come vedremo più dettagliatamente nel capitolo 6, l’emergere di rapporti ha rivelato organizzazioni di piccolo mondo (Small-World Networks-SWN) (Watts & Strogatz, 1998) strutture Scale-Free Troncate (Barabási, 2004). Si dimostrerà che alcune di tali reti musicali hanno comunità di nodi e clique, che identificano organizzazioni strutturate delle composizioni di Bach, i rapporti musicali nello spazio musicale, e altre organizzazioni, tipiche della teoria musicale. Oltre a considerazioni statistiche, le reti musicali, possono contribuire a spiegare l’organizzazione strutturale e funzionale della musica, e in ultima analisi, il pensiero cognitivo nella produzione musicale.

In questo capitolo, si forniscono i principali elementi di modellizzazione di questo complesso sistema di comunicazione, i codici che sono stati utilizzati, e tutti i passi seguiti nel processo di modellizzazione, per la riproduzione del metodo.

3.2. Modellazione Matematica delle Melodie Musicali

E' ben noto che matematica e musica sono tra loro strettamente connesse (Ashton, 2003). Questa relazione ha una duplice valenza: da una parte possiamo utilizzare modelli e strutture matematiche per creare melodie, dall'altra può essere interessante investigare le proprietà matematiche di una melodia per mettere in evidenza relazioni numeriche e rapporti geometrici (Hofstadter, 1979), da utilizzare nello sviluppo di strumenti di sintesi. Sistemi Dinamici, sia continui (Bilotta & Pantano, 2005; Bilotta, Pantano & Gervasi, 2005; Bertacchini et al., 2007; Bilotta et al., 2009) che discreti (Bilotta & Pantano, 2001; Bilotta et al., 2000a, b) possono utilmente essere utilizzati per sintetizzare musica. Per quanto riguarda i sistemi dinamici continui, il circuito di Chua, con il suo gran numero di attrattori caotici (Bilotta & Pantano, 2008; Bertacchini et al., 2009), permette di creare melodie con caratteristiche sempre diverse. Strutture frattali sono anche ampiamente utilizzate per la creazione di musica (Bilotta et al., 2002a, 2002b), perché è noto che la musica proveniente da frattali è percepita come consonante o piacevole, dal momento che ha una legge tipicamente di potenza. Tale legge, individuata per la prima volta da Voss e Clark (1975), è stata poi ritrovata anche in altri prodotti della musica occidentale (Manaris, 2005; Bigerelle & Iost, 2000).

Le leggi dell'armonia musicale sono state dedotte dai discepoli di Pitagora da una serie di esperimenti, rilevanti per la sistematicità e per il tempo in cui sono stati realizzati. La comprensione dei rapporti armonici sonori è avvenuta con il monocordo, uno strumento a singola stringa, che si estendeva su una cassa armonica, con due punti estremi fissi. Esso aveva un ponte mobile, per cambiare il grado di tensione e la lunghezza della corda. Usando una diversa lunghezza e larghezza della stringa, e variando la tensione, ruotando le viti cui si allacciano, si potevano creare suoni diversi. Pitagora osservò che la stringa più corta produce sempre la nota più alta e, la più lunga, la nota più bassa. Egli scoprì in seguito i tre accordi perfetti: l'ottava (corrispondente ad un rapporto di lunghezza di 1:2), la quinta (corrispondente al rapporto che si ottiene da 2:3) e la quarta (corrispondente al rapporto che si ottiene da 3:4). I rapporti tra numeri piccoli (1:2, 1:3, 2:3, ...) producono suoni piacevoli. L'armonia, la teoria di base della consonanza, è quindi strettamente legato ai rapporti tra piccoli numeri. Esiste pertanto una specifica correlazione tra matematica e musica. La musica occidentale è un perfetto sistema matematico. Partendo da una nota campione (convenzionalmente A4, La nella tradizione italiana), è possibile generare l'ottava, raddoppiando la frequenza precedente, e ottenere le altre note, opportunamente dividendo questo intervallo in 12 parti. Questo processo permette di costruire la scala musicale ben temperata. Le altre note, appartenenti a ottave diverse, possono essere ottenute nello stesso modo. Durante il tempo di Bach, per esempio, diversi musicisti avevano rifiutato le scale temperate, disapprovate per le caratteristiche di eccessivo meccanicismo e per il non rispetto delle regole dell'armonia. Johann Sebastian Bach, invece, era del parere opposto: credeva fermamente che la possibilità di passare da una tonalità all'altra era una caratteristica essenziale della vitalità della musica e che la simmetria e la regolarità nelle sequenze di note non erano

deplorevoli, ma una garanzia di ordine e di trasparenza. Per affermare questo sistema, Bach compose il primo e il secondo libro del Clavicembalo ben temperato, una rinomata collezione di 48 preludi e fughe, in cui ogni pezzo viene eseguito in tutte le tonalità della scala temperata, sia nei modi maggiori che minori. La scala ben temperata permetteva:

1. consonanza – rapporti di terza, quarta, quinta, unisono, ecc.
2. Completa libertà di modulazione, si può passare da un tasto qualsiasi ad un altro.
3. Convenienza nell'utilizzo pratico di strumenti a tastiera, che possono in questo modo suonare insieme senza risintonizzarsi.

Quando ascoltiamo una melodia musicale, in realtà noi ascoltiamo una sequenza di suoni sovrapposti generati da una serie di strumenti musicali. Ciascuno strumento rappresenta una *voce*. In realtà questi suoni non sono casuali, ma sono opportunamente codificati come *note* in quanto hanno una definita altezza ed una precisa durata temporale. Le armoniche superiori generate dallo strumento forniscono il timbro dello strumento stesso. Il pitch p è la frequenza percepita di queste note (Machlis & Forney, 1990) e nella notazione standard (Cavanagh, 2009) è correlato alla frequenza fondamentale del suono dalla seguente relazione:

$$(3.1) \quad p = 69 + 12 \log_2 \frac{f}{440\text{Hz}}$$

In questa convenzione p è un numero naturale.

Per poter analizzare una melodia, è necessario creare una rappresentazione della melodia stessa, sulla quale poter operare in modo formale con tecniche matematiche. In generale questo processo non è banale e sono utilizzate diverse statistiche (Pearce et al., 2004; Clauet et al., 2009), e numerosi metodi (Bengio, 1996; Basili et al., 2004). Anche se molto spesso si introducono errori ed ambiguità, vista la complessità del fenomeno musicale. Il metodo più usato è l'individuazione di pitch ai quali associare una simbolo: la successione di simboli rappresenterà utilmente anche se non esaustivamente la melodia (Manaris et al., 2003, 2005, 2007; Machado et al., 2007).

Nella musica elettronica, spesso al posto dei pitch si utilizza il codice MIDI. Il MIDI è un sistema formale utilizzato nella sintesi musicale a partire da uno strumento elettronico. Le note ed altre informazioni come gli strumenti sono codificati in un codice binario, che in genere è riconosciuto da altri strumenti elettronici, in grado di riprodurre la stessa melodia. È importante notare che il codice MIDI è un registratore di eventi, e quindi l'attacco di una nota avviene ad un determinato tempo, indicato dall'evento *Note On*. La registrazione cessa ad un determinato tempo (*Note Off*), anche se non sempre questo evento viene esplicitamente rappresentato, visto che può corrispondere ad un concomitante evento di *Note On*,

introducendo in tal senso delle ambiguità. Questo è tanto più evidente quando si tenta di rappresentare un file MIDI con l'usuale notazione visiva della rappresentazione delle note, sul pentagramma. Il sistema automatico, in questo caso, tenta di associare ad una serie di eventi, una nota, di una particolare durata e, questa rappresentazione non è ovviamente univoca. Molto spesso, due rappresentazioni fatte da software differenti producono articolazioni diverse per esempio nell'organizzazione del valore della nota, anche se il risultato percettivo, pur leggermente modificato, è largamente coincidente nella fase di sintesi. Nella Tabella 3.1, si riporta una corrispondenza delle note musicali con il relativo codice MIDI.

C	C#	D	D#	E	F	F#	G	G#	A	A#	B	Ottava
0	1	2	3	4	5	6	7	8	9	10	11	-1
12	13	14	15	16	17	18	19	20	21	22	23	0
24	25	26	27	28	29	30	31	32	33	34	35	1
36	37	38	39	40	41	42	43	44	45	46	47	2
48	49	50	51	52	53	54	55	56	57	58	59	3
60	61	62	63	64	65	66	67	68	69	70	71	4
72	73	74	75	76	77	78	79	80	81	82	83	5
84	85	86	87	88	89	90	91	92	93	94	95	6
96	97	98	99	100	101	102	103	104	105	106	107	7
108	109	110	111	112	113	114	115	116	117	118	119	8
120	121	122	123	124	125	126	127					9

Tabella 3.1. Corrispondenze tra la nota musicale e il codice midi. Le note sono espresse con la notazione inglese, ormai di largo uso in tutti i software musicali.

Un metodo alternativo per costruire una successione di simboli per rappresentare la melodia è partire da un file che è già in formato MIDI (Campolongo & Vena, 2006; Liu et al., 2010), rispetto alla segmentazione diretta dei suoni (Manaris et al., 2003, 2005, 2007; Machado et al., 2007).

Da un punto di vista formale, possiamo pensare la nota m come una terna, composta da un numero intero h che rappresenta l'altezza della nota in un opportuno codice, da un valore discreto T che rappresenta la durata del suono e da un numero intero I che rappresenta lo strumento. In tal modo, la nota m può essere espressa come:

$$(3.2) \quad m = (h, T, I)$$

Man mano che il tempo scorre (a intervalli discreti), la singola traccia di una melodia può essere rappresentata come un'applicazione f del tempo discreto nello spazio delle note M , i cui elementi sono composti da terne del tipo (3.2): $M \subset \mathbb{N}^3$. Allora

$$(3.3) \quad f : \mathbb{N} \rightarrow \mathbb{N}^3$$

Per cui, a ogni tempo (discreto) associamo una terna di numeri interi del tipo (3.2). Dunque la singola voce di una melodia, in formato MIDI può essere rappresentata da una sequenza di elementi di M :

$$(3.4) \quad Me = \{m_1, m_2, \dots, m_i, \dots, m_n / m_i \in M \}$$

Nel resto della tesi, rappresentiamo la nota solo con una coppia, in quanto, per semplicità di analisi, la parte di notazione relativa allo strumento non è considerata. Per cui, la (3.2) diventa:

$$(3.5) \quad m = (h, T)$$

dove h fa riferimento ad una nota nella specifica ottava, come riportato nella Tabella 3.1. T rappresenta il valore della nota, e convenzionalmente assume, al posto di un intero, uno dei seguenti valori convenzionali (4/4, 2/4, 1/4, 1/8, 1/16, 1/32, 1/64, 1/128). Con questa codifica, lo spazio musicale diventa uno spazio bidimensionale discreto, la cui cardinalità è $128 \cdot 8 = 1024$. A titolo di esempio consideriamo la prima traccia del Preludio n.1 di Johann Sebastian Bach.



Figura 3.1. La prima parte della partitura del Preludio No.1 di Johann Sebastian Bach.

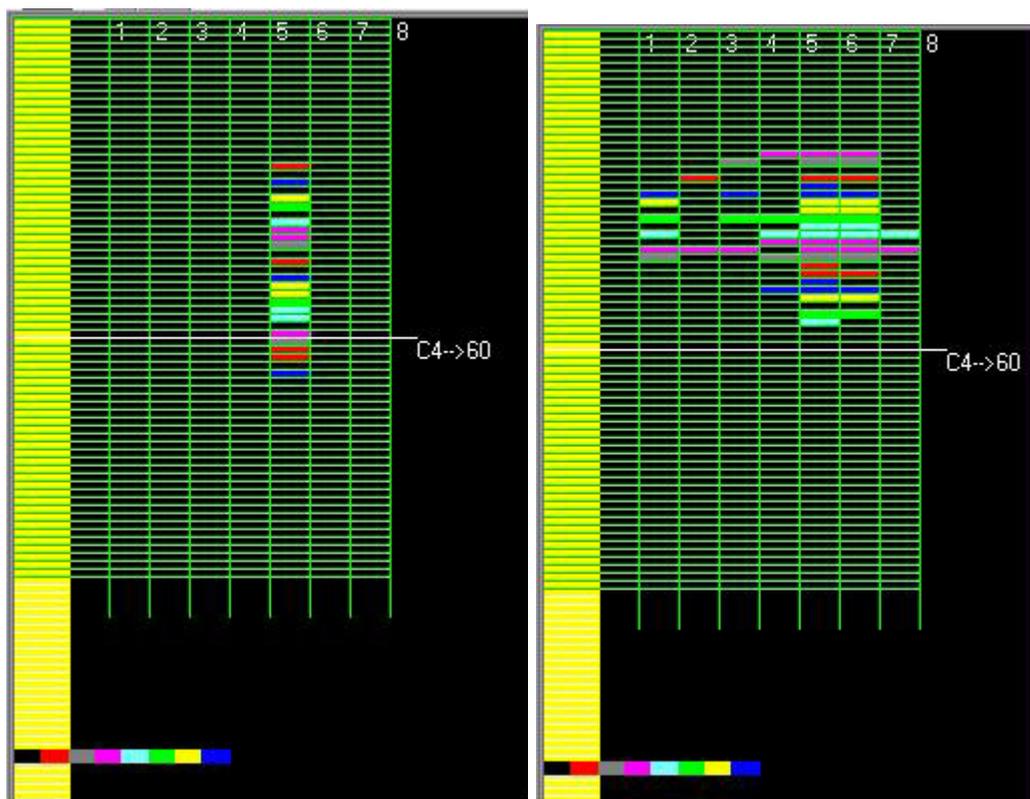


Figura 3.2. Rappresentazione grafica dello spazio musicale $M \subset \mathbb{N}^2$. Le singole note della melodia sono rappresentate come rettangoli colorati.

In questo caso le note del tipo (3.2) utilizzate sono 25: $M = \{X_1, X_2, \dots, X_{25}\}$. Possiamo rappresentare queste differenti coppie nello spazio discreto bidimensionale (Figura 3.2a) dove abbiamo posto sull'asse x la durata della nota, convenzionalmente posta da 1 a 8, corrispondente al tempo 4/4, 2/4, 1/4, ecc., e sull'asse verticale l'altezza della nota corrispondente, rappresentata da un numero tra 1 e 127, seguendo la Tabella 3.1. Il Do (che corrisponde a C nella notazione inglese) sulla IV ottava, è individuato dal numero 60. La Figura 3.2a mostra le 25 differenti note che compongono la melodia, che hanno tutte la stessa ascissa, in quanto hanno la medesima durata temporale pari a 1/16. Nella Figura 3.2b, è riportata la disposizione delle note nello spazio musicale per un'altra melodia. In basso un codice colore identifica le note La, Si, Do, Re, Mi, Fa Sol, avendo utilizzato lo stesso colore per la nota e le deformazioni della scala tonale (diesis e bemolle).

3.3. Simmetrie e complessità

La modellazione (3.2) delle note ci consente una rappresentazione visiva della melodia come in Figura 3.3. Questa può essere ottenuta nel modo seguente:

1. consideriamo un piano bidimensionale xy dove sull'asse x rappresentiamo il tempo t e sull'asse delle y l'altezza delle note;
2. ad ogni nota di una melodia, associamo un segmento orizzontale, posto in corrispondenza dell'altezza della nota;
3. mettiamo in successione le varie note della melodia, avendo cura di lasciare uno spazio vuoto quando incontriamo una pausa, pari alla durata della nota stessa.

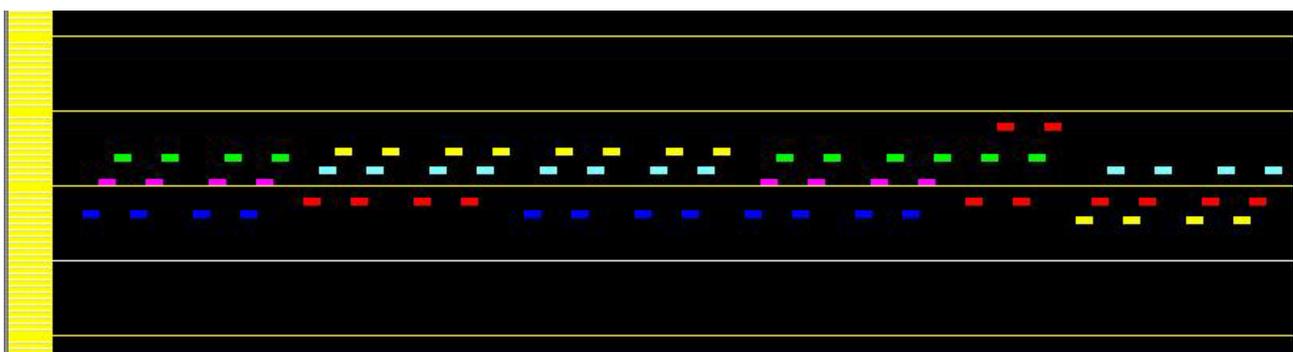


Figura 3.3. Rappresentazione visiva del Preludio N. 1 di Bach. Si noti lo sviluppo della melodia e la successione delle note al progredire del tempo.

Se uniamo le note consecutive tramite un segmento, avendo cura di lasciare uno spazio in presenza di una pausa, risultano del tutto evidenti le simmetrie presenti nella melodia, come strutture ripetute (e/o variate), a differenti tempi o su differenti altezze (Figura 3.4).

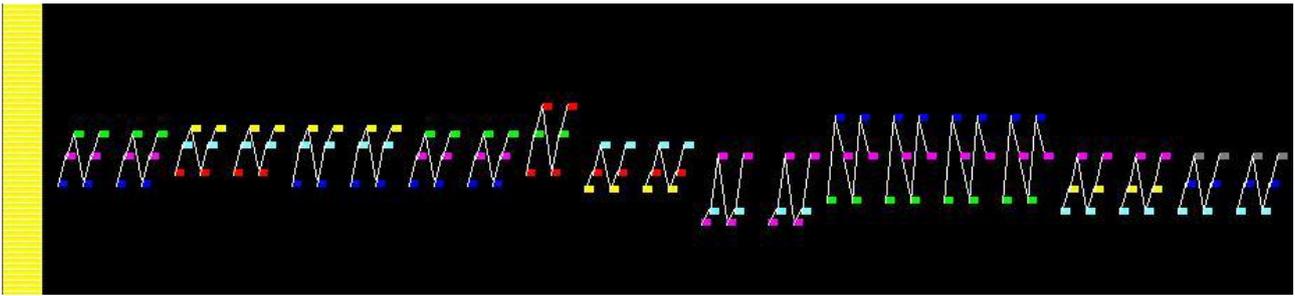


Figura 3.4. La presenza di simmetrie all'interno delle melodie è messa in risalto dalla ripetizione dei pattern.

Le simmetrie possono essere di tipo differente come traslazioni (orizzontali e verticali), riflessioni e/o contrazioni (Frova,1999; Pierce, 1992). Le *variazioni* possono essere considerate come *rotture di simmetria*. Un esempio di traslazione orizzontale occorre quando un blocco di note viene ripetuto *inalterato* nel tempo all'interno di una melodia. Una traslazione verticale interviene quando un blocco di note viene ripetuto ad una differente frequenza, mantenendo la stessa durata (valore) delle note e gli stessi rapporti armonici. Una contrazione e/o dilatazione interviene quando un gruppo di note viene ripetuto inalterato modificandone la durata/valore. Se un blocco di note appare in ordine inverso allora si parla di *riflessione*.

Lo strumento che ci permette di misurare le simmetrie è l'*entropia* (Shannon, 1948; Adami, 1998). La misura dell'entropia è stata usata da Meyer (1956) per suggerire un legame diretto tra il significato musicale e la tensione, secondo questo autore interpretabile musicalmente sia come una misura di informazioni sia di incertezza sul contenuto musicale, espresso attraverso gli andamenti dei contorni melodici, solitamente realizzati da tensioni/ rilassamento. In ambito musicale, altri ricercatori hanno usato misure di entropia in molti modi diversi, tra cui l'analisi della struttura nella musica atonale (Hiller e Fuller, 1967), la variazione stilistica nella musica tonale (Knopoff & Hutchinson, 1981), e le differenze tra gli stili musicali (Margulis & Beatty, 2008).

L'entropia è una grandezza statistica che misura la distribuzione dell'informazione in un sistema. Se si ha un sistema di M eventi, e questi eventi occorrono in m modalità diverse x_i , ed ogni evento viene ripetuto $rip(x_i)$, allora l'entropia per definizione è pari a

$$(3.6) \quad E = \sum_{i=1}^m -\frac{rip(x_i)}{M} \ln\left(\frac{rip(x_i)}{M}\right)$$

Esistono due valori estremi:

a) il primo quando il numero di eventi M ed il numero di modalità m coincidono ($M = m$); questo significa che, se un evento occorre un'unica volta, e tutti gli eventi sono differenti, allora $rip(x_i) = 1$. Quindi la formula (3.6) diventa:

$$(3.7) \quad E = \sum_{i=1}^M -\frac{1}{M} \ln\left(\frac{1}{M}\right) = \ln(M)$$

Questo fenomeno accade quando una persona digita in modo casuale i tasti del pianoforte. La misura dell'entropia, in questo caso, coincide col massimo del disordine/caos.

b) il secondo caso si ha quando abbiamo un'unica modalità x ($m = 1$) e, quindi, gli M eventi sono identici. Allora $rip(x) = M$. La formula diventa in questa occasione:

$$(3.8) \quad E = \sum_{i=1}^m -\frac{rip(x_i)}{M} \ln\left(\frac{rip(x_i)}{M}\right) = -\ln(1) = 0$$

Questo accade ad esempio quando uno strumento meccanico preme lo stesso tasto del pianoforte per lo stesso tempo M volte. Gli altri casi stanno nel mezzo. Quindi l'entropia di un sistema di M eventi che occorrono in m modalità diverse x_i ha un valore compreso tra 0 e $\ln(M)$.

Se si considera l'entropia relativa come il rapporto tra l'entropia e il massimo, $E = \frac{E}{\ln(M)}$, questo valore si colloca in un intervallo compreso tra 0 e 1. E' quindi possibile fare una comparazione tra sistemi diversi. Se abbiamo una melodia composta da M note, gli eventi saranno le singole note, ovvero differenti m note, che comporranno la melodia. E' possibile in tal modo calcolare l'entropia delle note di una melodia.

Consideriamo ad esempio il Preludio N. 2 di Bach, seconda traccia, estratto dalla raccolta *The Well Tempered Clavier, libro I*. In tale sonata, si individuano 34 note differenti, mentre la melodia è composta da 493 note, ovvero sono suonati 493 eventi, e esistono 34 possibilità di realizzazione di tali eventi. Un risultato di tale organizzazione è riportato nella Figura 3.5.

Per questa melodia l'entropia è 0,443936367573034, che rappresenta un valore intermedio tra l'ordine (rappresentato da 0) e il caos (rappresentato da 1).

Una coppia di note che si ripetono (traslazione orizzontale senza dilatazione/compressione) costituiranno un blocco di due note; una terna di note un blocco di 3, e così via. Nella Figura 3.6, è possibile osservare l'entropia dei vari blocchi, nelle prime 10 melodie della collezione *Art of Fugues di Bach (AOF)*. Coppie di note uguali (uguale frequenza e durata) non sono infrequenti; le terne ripetute, meno frequenti, non sono comunque trascurabili.

Distribuzione di eventi

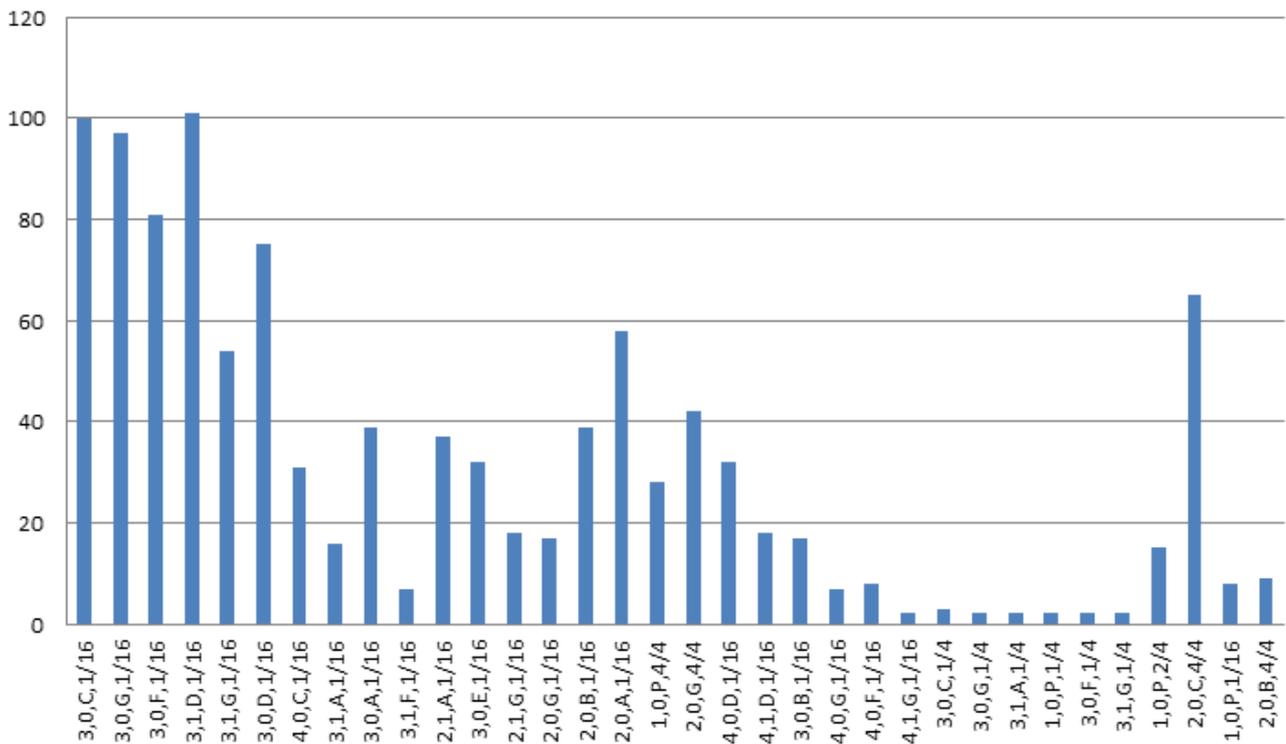


Figura 3.5. Esempio di distribuzione delle note in tutta la melodia. Come si può notare, la nota più frequente è il Mi# 3, seguita da Do, Sol e Re, sempre nella stessa ottava. Le 34 note differenti sono indicate nel codice Met, la cui specifica si trova alla fine del capitolo.

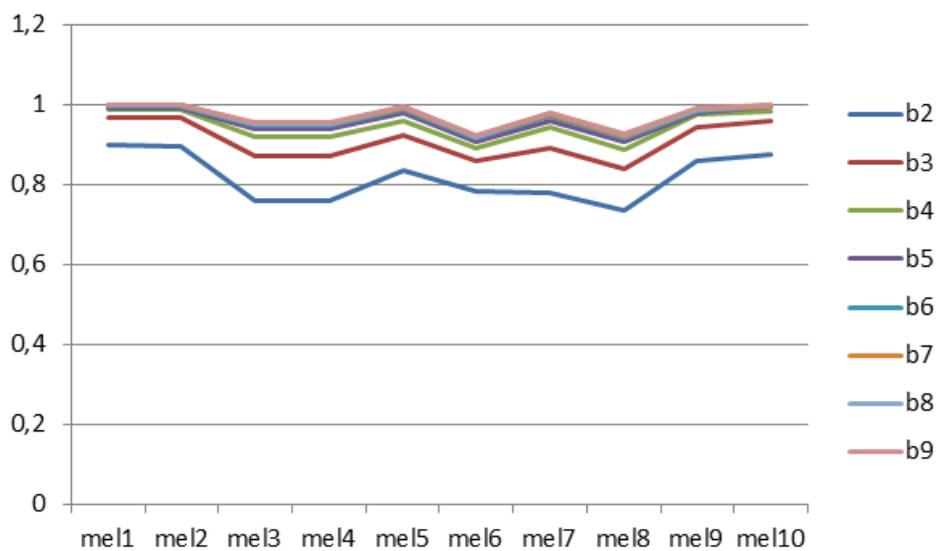


Figura 3.6. Entropia dei blocchi individuati nelle prime 10 melodie della collezione The Art of Fugues di Bach. Il numero che segue la lettera b, individua la dimensione del blocco.

Questo diagramma descrive solo una delle possibili simmetrie presenti in una melodia. Nella Figura 3.7 osserviamo il rapporto tra le simmetrie precedenti (indicate con S1), nel caso della melodia m1, considerando blocchi di note ripetute senza distorsioni e in presenza anche di traslazioni verticali di ottava.

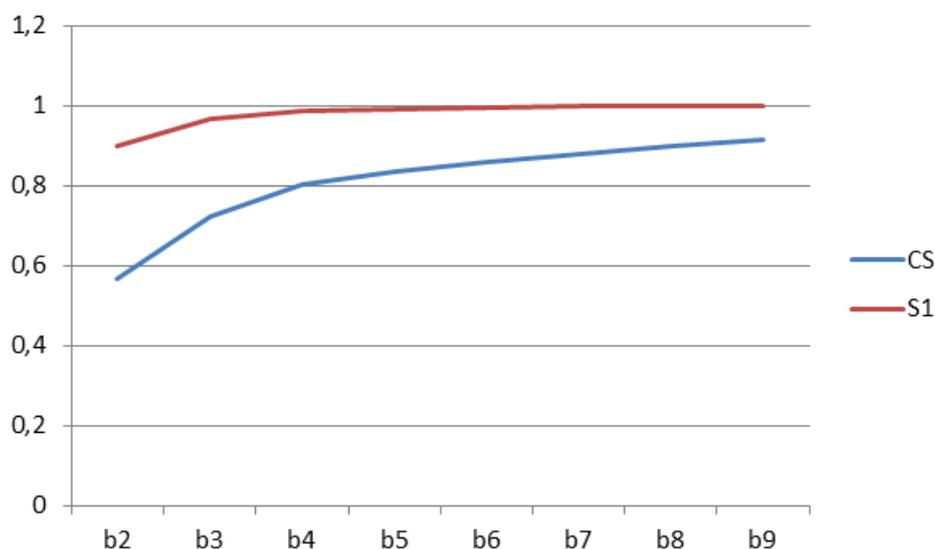


Figura 3.7. Confronto tra le entropie nel caso di simmetrie di traslazione (S1) ed altre simmetrie (dilatazione/compressione, traslazione verticale di ottava e traslazione orizzontale) nella prima melodia della collezione *The Art of Fugues di Bach*.

3.4. Le reti melodiche

Come avviene per altri sistemi biologici e artificiali dove si predilige una associazione lineare di elementi (Albert et al., 1999; Barabási & Albert, 1999; Albert & Barabási, 2002), man mano che una melodia si sviluppa, le varie note intrecciano relazioni tra di loro. Noi possiamo rappresentare queste relazioni per mezzo di un grafo (Campolongo & Vena, 2006; Liu et al., 2010).

Possiamo associare un grafo ad una melodia, operando nel seguente modo:

1. associamo un nodo v_i ad ogni nota X_i .
2. Percorrendo la stringa che rappresenta la melodia, consideriamo un arco c_{ij} tra due note X_i e X_j se X_j è un successore di X_i . In questo modo, possiamo avere una rappresentazione grafica della melodia, ponendo in modo opportuno i nodi nello spazio e disegnando gli archi tra questi nodi.

Nella Figura 3.8, possiamo osservare una rappresentazione del grafo generato dalla traccia 1, della collezione *The Goldberg Variations di Bach*. La rete risulta ovviamente molto intrecciata per la presenza di molti nodi (193), pari al numero di note differenti e ha anche un numero considerevole di archi (1513).

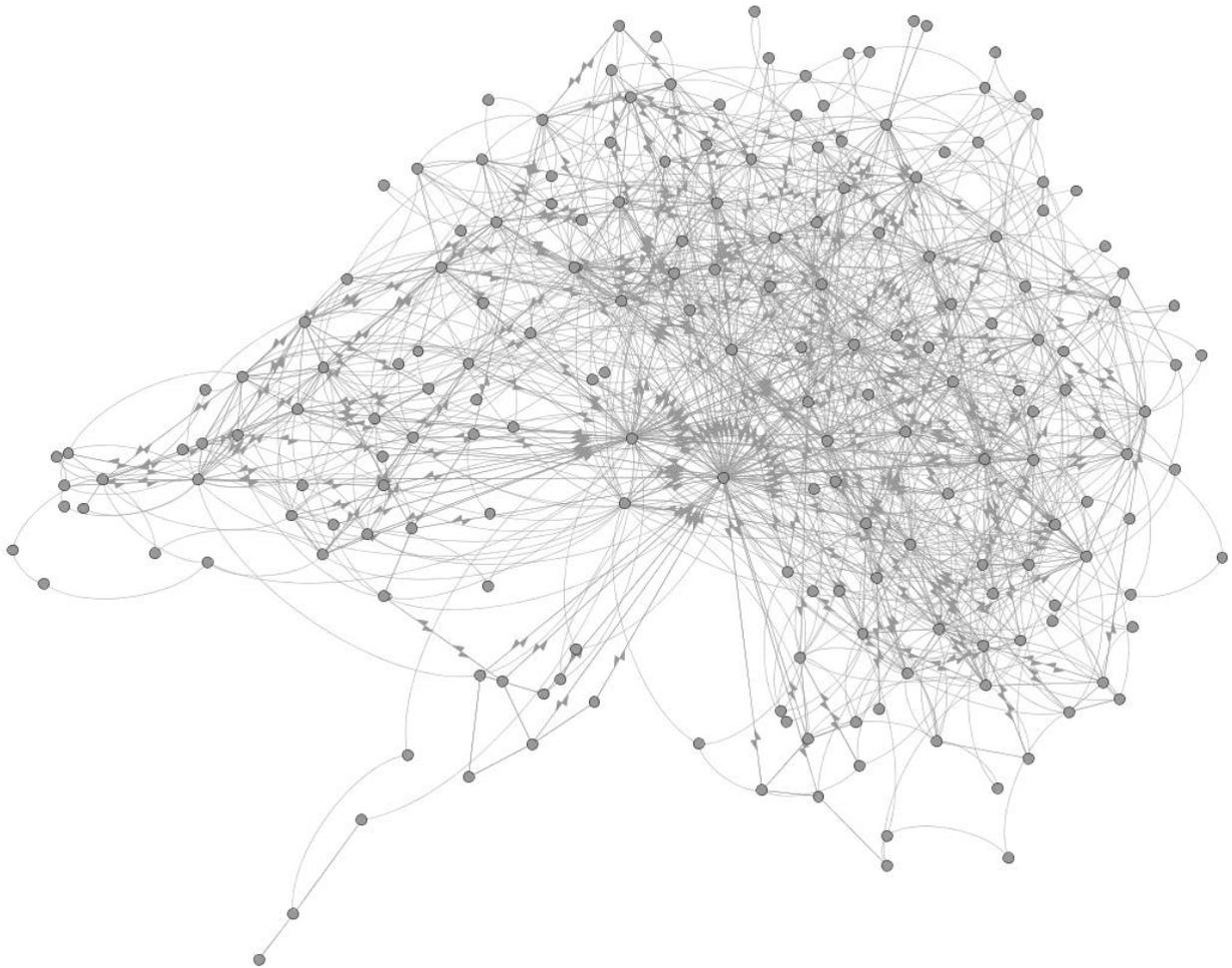


Figura 3.8: Rappresentazione del grafo corrispondente alla traccia 1, melodia Goldberg, categoria Gold di Bach.

Il problema di disporre gli archi e i nodi in modo adeguato per poterli visualizzare è un problema molto complesso e vari metodi possono essere utilizzati. In generale la scelta è quella di creare il numero minimo di sovrapposizioni tra gli archi e il massimo numero di spaziature tra i vertici. I metodi più popolari utilizzano 2 algoritmi: il modello a molla e il modello elettrico. Si suppone che tra i nodi si vadano ad esercitare delle forze e l'algoritmo tende a minimizzare l'energia del sistema. Ovviamente questi modelli fisici possono essere utilizzati anche nel piano, dove, in assenza della terza dimensione, gli oggetti sono meglio rappresentati.

Queste rappresentazioni si basano sulla relazione antecedente-successore presente nella melodia. Questi grafi possono contenere degli attributi come ad esempio quante volte un arco è ripetuto, che è equivalente a quante volte una coppia di note compare nella melodia. Un modo per rappresentare quante volte un arco è ripetuto è quello di dimensionare lo spessore dell'arco in funzione delle sue ripetizioni come in Figura 3.9.

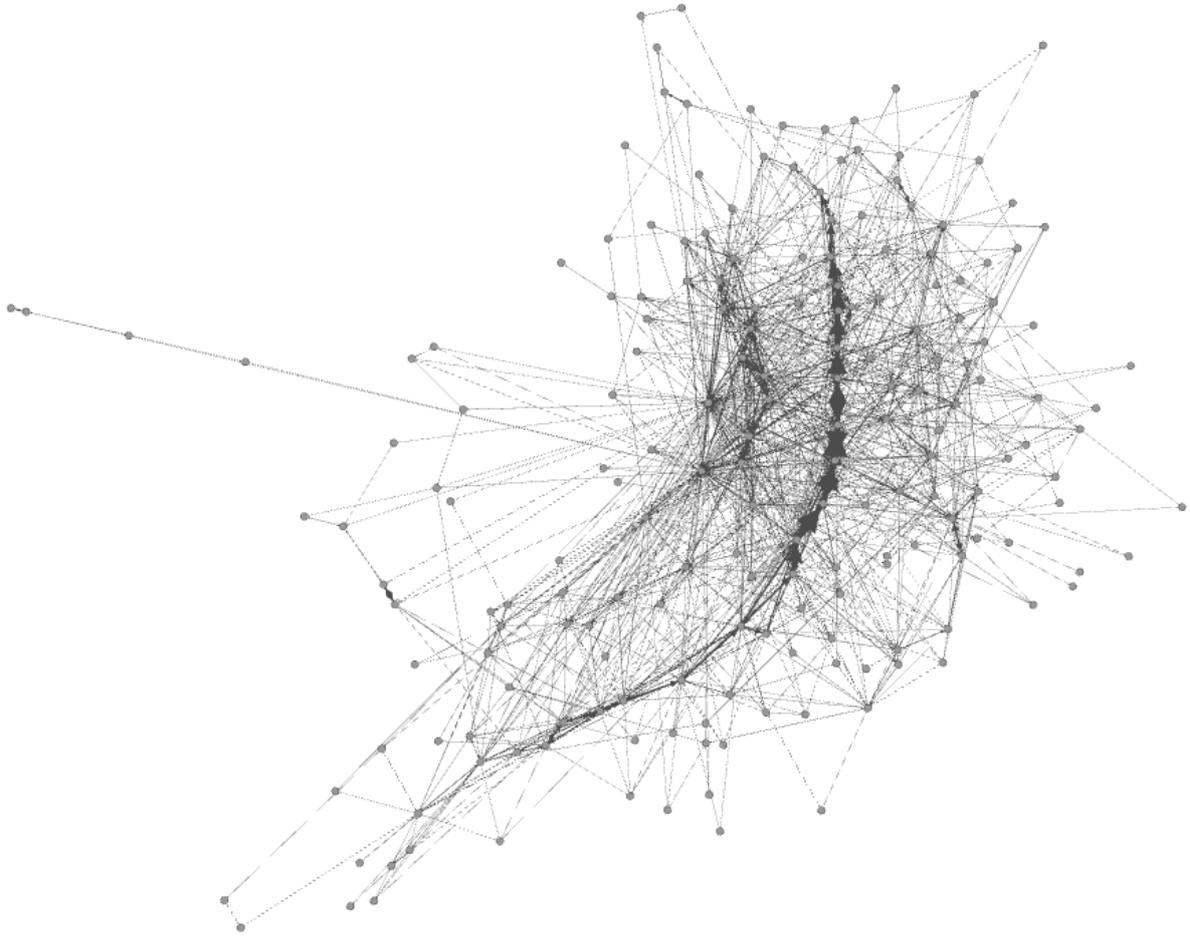


Figura 3.9 .La dimensione degli archi è proporzionale al numero di volte in cui lo stesso blocco di 2 note che compare nella melodia.

In questa Figura possiamo osservare come si manifesti un'organizzazione centrale della traccia, con molti passaggi ripetuti tra gruppi di note. E' importante notare che esiste una relazione tra l'entropia di coppie di note e il peso degli archi. In un certo senso l'entropia misura la distribuzione dei pesi sugli archi. In questo caso l'entropia è pari a 0,78 da cui si evince che queste ripetizioni, che esprimono simmetrie all'interno della rete, non sono trascurabili.

E' possibile inserire le etichette di ogni nota, per vedere a quali note corrispondono i vertici. In questo caso, per favorire la leggibilità delle note, abbiamo utilizzato il codice .MET Il primo numero rappresenta l'ottava, il secondo (0 o 1) rappresenta l'assenza o la presenza del diesis, la lettera rappresenta la nota in notazione inglese e il quarto numero il valore (la durata temporale) della nota.

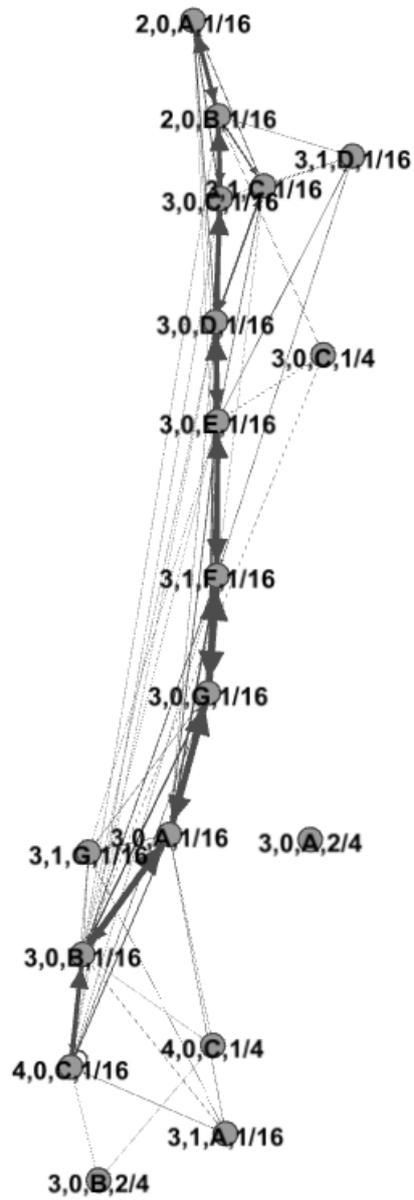


Figura 3.10. Un dettaglio di alcune relazioni particolarmente importanti relative alla traccia 1, melodia Goldberg, categoria Gold di Bach.

Se le reti non sono molto estese o particolarmente connesse, i pattern ci possono fornire delle interessanti informazioni. In Figura 3.11, le singole note sono rappresentate attraverso un codice colore.

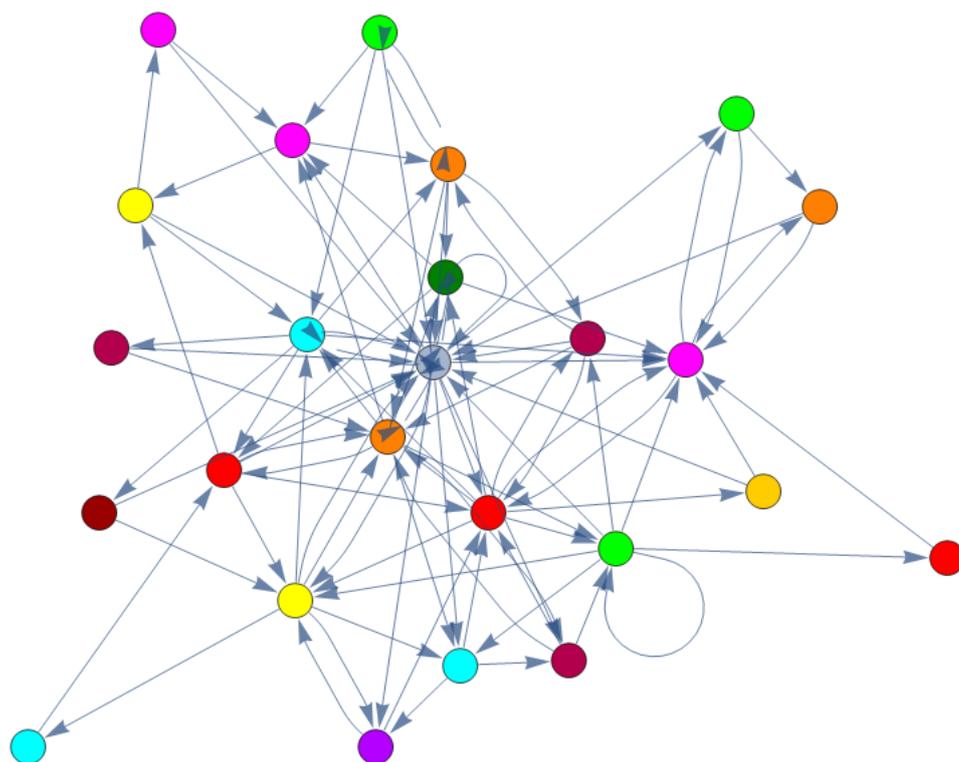


Figura 3.11. Rappresentazione di una traccia con un codice colore per indicare le note.

Per concludere il capitolo riportiamo le caratteristiche del file .MET prodotto dal sistema computazionale partendo da un file MIDI:

- 1) il file inizia con un numero che indica il numero dei canali seguito da un “;”
- 2) ciascun canale deve essere indicato con la seguente stringa: "#canale#", seguito dal numero dello strumento utilizzato in formato MIDI;
- 3) segue una nota per ogni riga, separando le informazioni su di loro da virgole. Le informazioni che devono essere fornite per ogni nota sono:
 - a. Numero di ottava (che può variare da 1 a 7).
 - b. Deformazioni (che può essere 1, se si vuole indicare la presenza di forte, e 0 altrimenti).
 - c. I nomi delle note sono riportati nella notazione musicale inglese (C = Do. D = Re. E = Mi. F=Fa. G = Sol. A = La. B = Si).
 - d. La durata della nota è espressa in quarti. Ogni nota può assumere i seguenti valori: 1/4. 2/4. 3/4. 4/4. 1/8. 1/16. 1/32. 1/64.
- 4) tutte le righe del file .MET devono terminare con il carattere ";"
- 5) Il file deve essere chiuso con il carattere ":".

Capitolo 4. Modellazione di reti cerebrali

4.1. Introduzione

Quali sono i circuiti che regolano le competenze linguistiche a livello cerebrale? Come fa un sistema fisico come il cervello a produrre il linguaggio-pensiero? La dicotomia fra mente e cervello (LeDoux, 2002; Solms & Turnbull, 2002; Oliverio, 2004) agita ancora la scienza contemporanea. La scienza delle reti ci può aiutare nella comprensione delle interconnessioni cerebrali per individuare il funzionamento di vaste aree cerebrali deputate alla processazione ed elaborazione linguistica, studiando come si organizzano i circuiti neurali del cervello umano per acquisire e elaborare informazioni che danno, in uscita, questa capacità globale che è il possesso di una lingua. Ci può anche spiegare come aree cerebrali così interconnesse siano ad un tempo un fattore di organizzazione e di interpretazione dell'ambiente che ci circonda, e specchio della pragmatica sociale. In questo capitolo, dopo una breve esposizione sull'organizzazione strutturale e funzionale del cervello umano, e sui circuiti cerebrali inerenti il linguaggio, passibili di modellazione, saranno forniti i concetti principali della *connettistica* (Sporns, 2011), la scienza per modellare reti cerebrali. Tecniche recenti di brain imaging adottate nel settore delle Neuroscienze, la trattografia e la connectomica permettono di realizzare reti cerebrali tridimensionali. Sono descritti in dettaglio i metodi per modellare differenti reti cerebrali, a diverse scale di grandezza, ottenute attraverso l'uso di trattografia diffusiva da Immagini di Risonanza Magnetica (Magnetic Resonance Imaging) in soggetti sani e malati. Le motivazioni allo studio delle reti cerebrali sono le seguenti:

- a. Molti metodi per lo studio delle reti sono stati ampiamente applicati già in diversi settori scientifici (Strogatz, 2001), incluso lo studio delle reti inerenti l'interazione delle proteine, ecosistemi, comunicazione e trasporto, il world wide web, i sistemi sociali, e finalmente, il cervello umano (Bullmore & Sporns, 2009).
- b. Diversi studi hanno dimostrato che le reti anatomiche della corteccia cerebrale dei mammiferi esibiscono degli attributi tipici delle small world networks (Bassett et al., 2008).

- c. Strutture di piccolo mondo sono state riscontrate nella corteccia di primati non-umani, servendosi di dati ottenuti tramite “tract tracing” (Sporns et al., 2000; Sporns & Zwi, 2004), e tramite Risonanza Magnetica diffusiva del cervello umano (Gong et al., 2009; Hagmann et al., 2008, 2007; Iturria-Medina et al., 2008).

4.2. Strutture cerebrali

La nostra conoscenza del cervello e di come si sia evoluto il linguaggio dai primati non umani al genere *Homo* è ancora esigua, anche se si fanno avanti numerose ipotesi genetiche sulla loro evoluzione (Prauss, 2012). Ma come si è evoluto il linguaggio, perché c'è stata questa specializzazione del cervello umano (Geschwind, 1972; Hockett, 1960)? Secondo Mac Lean (1984), il cervello dell'uomo e dei primati ha tre livelli di organizzazione. Il mesencefalo, una parte importantissima del cervello umano che praticamente funge da connettore per tutte le vie nervose motorie e per i canali sensoriali, insieme con i gangli di base costituiscono il primo settore. Questi moduli, che sono la sede delle attività riflesse, e risalgono a 200 milioni di anni fa, sono posseduti anche dai rettili e dagli anfibi. Il sistema paleoencefalico-limbico costituisce il secondo livello, che è sede dei processi emotivi, risalenti a 100/200 milioni di anni fa, e ritrovati nei mammiferi primitivi. La neocorteccia, il sistema che presiede ai processi di apprendimento e di adattamento, presente nei mammiferi più evoluti, di circa 20 milioni di anni fa, in concomitanza dell'uso di strumenti (Beck, 1980; Chiarelli, 1985). La corteccia, configurazione di neoformazione, è la struttura che permette l'attività motoria e le funzioni adattive, tipiche della specie umana, per cui rappresenta il terzo livello di sviluppo del cervello umano. Costituita da circa 8 miliardi di neuroni, è un tessuto denso di neuroni la cui funzione principale è dedicata a distinguere, comparare, ordinare e processare dati in ingresso, per la loro categorizzazione e la loro stiva nelle memorie specializzate (Tulving, 1989). La corteccia si divide in quattro parti, distinte e simmetriche, dette lobi (frontali, parietali, temporali, occipitali,) che costituiscono l'emisfero destro e sinistro del cervello umano. I due emisferi sono collegati da una struttura di interconnessione detta corpo calloso. I lobi più sviluppati (frontali e parietali), sono collegati alla produzione linguistica.

In particolare, i lobi **frontali**, presiedono alle funzioni primarie del cervello; sono coinvolti nel controllo muscolare, nel controllo emotivo e degli impulsi primari, nell'attività di produzione e ideazione del parlato, nelle scelte decisionali, nella risoluzione di problemi, nella pianificazione delle azioni e nei processi creativi. I neuroni specchio, che rivestono grande importanza nella comprensione del linguaggio, per via delle loro capacità simulative del comportamento motorio, legato ai verbi di moto si trovano vicini alla corteccia motoria, nei lobi frontali.

I lobi **temporali**, situati dietro le tempie, sono implicati nella processazione degli stimoli uditivi, e coinvolti in processi di memorizzazione e elaborazione di contenuti emozionali. Infatti, i lobi temporali

sovrintendono alla funzione di memorizzazione, alla comprensione/produzione del linguaggio (con le aree di Broca e di Wernicke poste nell'emisfero sinistro), alle capacità musicali e alla visione immaginativa (tipiche dell'emisfero destro), alla regolazione del comportamento.

I lobi **parietali** governano l'acquisizione e processazione di informazioni sensoriali provenienti non solo dai canali sensoriali, ma anche dal corpo. In questa sede avvengono numerose importanti funzioni legate al linguaggio, si crea la base di conoscenza dell'ambiente esterno e si creano connessioni tra fonemi, morfemi, parole, frasi, pensieri. In questa sede si strutturano le associazioni fra immagini verbali e visive.

I lobi **occipitali**, posti nella parte posteriore di ciascun emisfero cerebrale, sono coinvolti nella visione. I quattro lobi si incontrano nel mesencefalo, l'area del cervello in cui avviene la fusione dell'informazione sensoriale. Il mesencefalo è coinvolto in numerose malattie neurodegenerative, come si vedrà nel successivo capitolo 7.

Le funzioni dei due emisferi cerebrali sono le seguenti. L'emisfero sinistro del cervello umano è la sede delle strutture che sovrintendono al linguaggio nella specie umana. Nelle 11 componenti modulari di tipo corticale, e quattro di tipo subcorticale, sono localizzate le funzioni fonologiche, sintattico-semantiche e ideative del linguaggio, per le abilità linguistiche di produzione/comprendimento e scrittura/lettura. L'area cerebrale di Broca, situata nel lobo frontale della corteccia cerebrale, e l'area di Wernicke, connessa all'area di Broca, tramite il fascicolo arcuato, sono deputate alla produzione/comprendimento del parlato e a tutte le funzioni cognitive legate al pensiero-linguaggio. Che il linguaggio sia connesso all'area di Broca e di Wernicke, situate nell'emisfero sinistro è stato scoperto attraverso studi di soggetti affetti da afasia (mancanza di produzione linguistica ai differenti livelli della struttura linguistica) in correlazione a diversi tipi di danni cerebrali o traumi. Nell'area di Broca sono iscritte tutte le memorie motorie, utili per la produzione linguistica, mentre l'articolazione fonica, che genera la produzione di parole, necessita della cooperazione di altre aree motorie della corteccia e di strutture sotto-corticali, costituite dai gangli della base e dal cervelletto. Lesioni all'area di Broca producono difficoltà nell'organizzazione grammaticale-sintattica delle frasi e difficoltà a nominare gli oggetti, nell'articolazione fono-articulatoria. Danni al fascicolo arcuato producono afasia di conduzione, che si presenta con una difficoltà a ripetere il linguaggio udito, senza danni per la capacità di comprensione e produzione di significato. Alla struttura che circonda la parte posteriore della scissura temporale (indicata con il nome di scissura di Silvio) è affidata la comprensione del linguaggio. Tale regione cerebrale è un'area di interconnessione tra la rappresentazione uditiva della parola, processata nell'area di Wernicke, e il suo significato, stivato sotto forma di rappresentazioni nella corteccia associativa sensoriale. Lesioni in quest'area causano l'afasia transcorticale sensoriale, che si manifestano con difficoltà nella comprensione del linguaggio e alta produzione di parole senza senso. La capacità di ripetere il parlato in fase di ascolto rimane inalterata. Causata da lesioni temporali e parietali, l'anomia è un tipo di afasia che si manifesta con una difficoltà a menzionare gli oggetti

con i nomi. Lesioni nell'area frontale di Broca producono invece una difficoltà ad articolare le azioni espresse dai verbi. L'anomia può essere causata anche da lesioni specifiche nell'area associativa, generando impossibilità nella produzione di nominare parole, associate a significati. Situata nel lobo temporale, nella regione posteriore-superiore, l'area di Wernicke contiene le memorie uditive, utili per la comprensione del linguaggio parlato. Lesioni a quest'area producono incapacità di comprensione del parlato. Rimangono invece intatte le capacità di comprendere testi scritti e di produzione linguistica, sia scritta sia orale. Al contrario, l'emisfero destro non è coinvolto nella processazione del linguaggio, sebbene sia coinvolto nella gestione di enunciati automatizzati, come contare, elencare liste. Rientrano in tali funzioni automatiche anche il ritmo dell'eloquio e la consapevolezza del "prendere il turno" per parlare, nell'interazione sociale. Presiede a processi connotativi (estremamente soggettivi) della processazione semantica, mentre al contrario l'emisfero sinistro sovrintende i processi di analisi dei significati di tipo denotativo (socialmente condivisi). La parte destra del cervello umano processa i parametri extralinguistici del parlato, che comprendono gli andamenti intonazionali, l'intensità, il timbro o *colore* della voce, le espressioni tonali delle emozioni e le qualità articolatorie dei fonemi di una lingua. Attraverso tali variabili extralinguistiche, l'emisfero destro processa gli aspetti pragmatici del linguaggio, che si riferiscono agli elementi contestuali e sociali che permettono di disambiguare il significato delle parole. Importanti dal punto di vista dell'estrazione di significato sono gli elementi emozionali del linguaggio. L'emisfero destro ha la funzione di analizzare dati testuali e visivi, interpretare discorsi complessi, estrarre contenuti metaforici, riconoscere le intenzioni comunicative espresse nel parlato dal soggetto, seguire le relazioni che si dispiegano nella sintassi del periodo, attraverso la concatenazione di enunciati correlati, estrarre elementi di coerenza tra le frasi in un discorso, organizzare l'ordine degli eventi, discriminare gli elementi di primo piano rispetto a quelli di sfondo, riconoscere anche le funzioni linguistiche. I processi realizzati dall'emisfero sinistro avvengono in parallelo con quelli dell'emisfero destro. Ben lontani dall'essere dicotomici, i due emisferi realizzano funzioni cognitive complesse e complementari per produrre e interpretare i significati della lingua, anche se ancora non conosciamo le modalità di organizzazione.

4.3. Circuiti cerebrali cognitivi

Circuiti cerebrali connettono vaste aree del cervello realizzando processi di informazione non solo a livello del singolo neurone, ma fra gruppi di neuroni, livelli più vasti di agglomerati neurali, reti cerebrali specializzate, come quelle deputate alla produzione linguistica. Infatti, le reti cerebrali si dispiegano su più scale spaziali, dalla micro scala di singole cellule e sinapsi, alla macro scala dei sistemi neurali, incorporati in sistemi cognitivi. Lo studio di come si connettono le diverse aree cerebrali rappresenta un settore emergente nel campo delle Neuroscienze. La connettività svolge un ruolo rilevante per chiarire la struttura neuroanatomica del cervello umano, lo sviluppo neurologico, la struttura elettrofisiologica dei circuiti neurali, le funzioni del cervello rilevate attraverso il *Brain Imaging*, le strutture neurali deputate alla

cognizione in generale e al linguaggio in particolare. Il poter analizzare la struttura delle connessioni fra diverse aree cerebrali fornisce importanti chiarimenti sulla funzione di integrazione che il cervello svolge. Inoltre la modellazione di dati provenienti da Risonanza magnetica funzionale può farci capire come emerge la complessità organizzativa delle strutture del cervello, anche se costituite da elementi neurali sparsi ma interconnessi da sinapsi, attraverso gli assoni. La rete ci permette di analizzare l'architettura di tali connessioni per individuare i circuiti specializzati del linguaggio e di altre funzioni cognitive. Infatti, la modellazione matematica delle reti, e i metodi quantitativi che essa sottende, ovvero la traduzione in rete di immagini provenienti da Risonanza Magnetica permette la verifica di principi di organizzazione dell'architettura cerebrale che definisce l'anatomia del cervello, per definire come sia singoli neuroni, sia strutture via via più complesse, organizzano risposte fisiologiche a catene di stimoli provenienti dalle aree sensoriali. Queste risposte, inoltre risultano dalle interazioni tra un gran numero di singoli neuroni, collegati in circuiti locali, nonché in circuiti sempre più vasti, che individuano le aree dei lobi (frontali, parietali, temporali, occipitali,) che costituiscono, come detto precedentemente, l'emisfero destro e sinistro del cervello umano. L'integrazione strutturale e funzionale di questi circuiti cerebrali, che lavorano in parallelo, determina i differenti stati cognitivi che gli umani sperimentano, all'interno degli stati cerebrali associati al sonno e/o alla veglia. Differenti stati cognitivi corrispondono a differenti pattern di attivazione cerebrale, vale a dire a differenti circuiti neurali o percorsi nell'architettura di rete. Questi modelli non sono statici ma dinamici e realizzano la base delle dinamiche cerebrali. Dinamiche differenti possono realizzare processi cognitivi diversi e esistono anche notevoli variazioni non solo tra un individuo e l'altro, ma anche all'interno dello stesso individuo: variando le connessioni circuitali e le relative dinamiche, cambia la connessione o l'architettura di rete. Tali dinamiche sono molto più importanti e in continua evoluzione durante lo sviluppo del cervello in soggetti di età evolutiva. Crescono e si sviluppano nuove connessioni, mentre altre possono essere abbandonate. Inoltre, nella corteccia motoria ha sede una rappresentazione fisica del corpo. Cervello e corpo sono accoppiati in modo dinamico attraverso cicli continui di percezione-azione. Quali sono i principali circuiti cognitivi? Cosa avviene al cervello quando è colpito da malattie neurodegenerative?

Ricerche di neuroscienze hanno rilevato che le aree corticali e subcorticali sono in comunicazione attraverso complessi circuiti neurali che regolano i meccanismi d'attivazione e d'inibizione dei sistemi neurali, e che sottendono l'organizzazione cognitiva (Leonardi, 2008). I circuiti cortico/subcorticali comprendono i gangli basali, il talamo, l'amigdala, e l'ippocampo. Le aree della corteccia sono invece connesse dai circuiti cortico/corticali. Questi due tipi di circuiti attivano la corteccia frontale per cui sono anche chiamati circuiti fronto/corticali e fronto/sottocorticali. Appartengono ai circuiti fronto/corticali le seguenti aree cerebrali:

a) il fascicolo arcuato, che come abbiamo già menzionato, connette l'area di Wernicke con l'area di Broca;

- b) il fascicolo fronto/occipitale superiore, che collega l'insula e la corteccia frontale con i lobi temporali ed occipitali;
- c) Il fascicolo uncinato, che lega la corteccia frontale orbitale e polare con i lobi temporali anteriori;
- d) l'area del cingolo che unisce il lobo frontale con l'ippocampo e le aree parietali posteriori ed occipito/temporali.

I circuiti fronto/sottocorticali individuati sono le seguenti connessioni cerebrali:

- e) circuito motorio, comprensivo delle specifiche di tale area costituita dall'area motoria aggiuntiva, dalle aree premotorie, dall'area motoria primaria e dall'area somatosensoriale;
- f) circuito oculomotore, che si diparte dalle aree oculari frontali;
- g) circuito prefrontale/dorsolaterale, che collega la superficie dorsolaterale del lobo frontale con altre aree frontali e quelle parietali;
- h) circuito orbitofrontale/laterale, che comprende i giri orbitali laterali con connessioni alla corteccia temporale dell'insula; questa struttura è collegata con il circuito dorsolaterale;
- i) circuito orbitofrontale/mediale, che ha origine nel giro retto e nel giro orbitario mediale, in collegamento con la corteccia cingolata anteriore e riceve afferenze da strutture subcorticali e mesencefaliche coinvolte con la gratificazione ed il piacere;
- l) circuito del cingolo anteriore, ripartito nelle sottoregioni rostrali, con funzione affettiva; dorsale, con funzione cognitiva; caudale con funzioni motorie;
- m) circuito corteccia laterale/cervelletto con funzioni esecutive. Questi circuiti sono reciprocamente interconnessi.

Come fa ad emergere una funzione cognitiva da un semplice circuito con funzione associativa e trasmettendo informazioni attraverso le varie connessioni da un'area all'altra? Le reti cerebrali modellate attraverso dinamiche di rete possono aiutare a comprendere come emergono le connessioni da semplici scambi di informazioni.

La modellizzazione attraverso la Connectomica, studia il cervello umano come una rete complessa con l'obiettivo di avere una descrizione globale dell'architettura di tale rete, costruendo una mappa completa dei suoi elementi e delle sue connessioni. Lo studio di una rete cerebrale prevede inoltre il calcolo di parametri tipici della teoria dei grafi, volti a individuare le aree più centrali, ad esaminare la tolleranza della rete in questo caso alla neurodegenerazione strutturale e funzionale, per calcolare quanto la rete elabori processi di connettività globali e locali.

4.4. Modellazione matematica: dalla scansione di Risonanza Magnetica alle Reti Cerebrali

La Connettomica (Sporns 2011) è un nuovo campo emergente delle neuroscienze che combina la Risonanza Magnetica Diffusiva e le metodologie della trattografia con gli strumenti di analisi delle reti. Negli ultimi

anni, le tecniche di brain imaging hanno permesso di disvelare la comprensione della struttura e del funzionamento del cervello umano. In particolare, tramite Immagini di Risonanza Magnetica diffusiva (Pierpaoli et al., 1996), si è riusciti ad evidenziare le fibre e i collegamenti che esistono tra le diverse strutture corticali e sub-corticali. Scoperta dai fisici Felix Bloch ed Edward Purcell, premi Nobel nel 1952, la Risonanza Magnetica analizza il moto dei protoni degli atomi di idrogeno all'interno della materia, che sono sottoposti all'azione di un campo magnetico statico e di onde elettromagnetiche dello spettro delle onde radio (RF), che forniscono loro energia (Carr & Purcell, 1954). La Risonanza Magnetica Diffusiva ha ricevuto grande interesse per la prima volta nei primi anni novanta quando venne utilizzata come biomarker per l'individuazione di ictus (Moseley et al., 1990). In seguito venne messo in luce come, attraverso tale tecnica, fosse possibile giungere al tracciamento delle fibre che caratterizzano la materia bianca (Conturo et al., 1999; Mori et al., 1999; Wedeen, 1996), sfruttando le proprietà di anisotropia di tale regione (Basser et al., 1994, 2000; Douek et al., 1991). La Risonanza Magnetica Diffusiva ha dato vita allo studio non invasivo della *mappatura strutturale della connettività del cervello*, partendo dalla conoscenza dei percorsi delle fibre ottenuto dalla studio della diffusione locale dell'acqua nei tessuti cerebrali. Le molecole d'acqua all'interno dei tessuti, a temperatura corporea o ambiente, mostrano un moto casuale dovuto all'energia termica; questo fenomeno è comunemente chiamato moto Browniano (Le Bihan, 1995). All'interno di un volume limitato, ad esempio un voxel di alcuni mm^3 , lo spostamento di una popolazione di molecole d'acqua può essere descritto attraverso una funzione di densità di probabilità (PDF) (Callaghan, 1991). Ovviamente, la forma della PDF è fortemente influenzata dall'ambiente circostante, ad esempio, l'orientamento delle barriere locali create dalle membrane cellulari, guaine di mielina, macromolecole, influenzano direttamente questa funzione, che è anche differente in ogni posizione del cervello, a causa della diversa architettura dei tessuti (Beaulieu, 2002). Il sistema diffusivo ha rivelato che il moto delle molecole d'acqua all'interno dei tessuti può essere definito attraverso l'applicazione di un forte gradiente bipolare (Le Bihan, 2003).

In presenza di un campo magnetico variabile, spazialmente indotto attraverso un gradiente magnetico, le molecole in moto emettono segnali alle radiofrequenze, con delle leggere differenze di fase. In un piccolo volume (voxel), contenente un gran numero di molecole d'acqua, queste fasi sono distribuite in maniera casuale, riflettendo direttamente la traiettoria delle molecole (processo di diffusione). Questa diffusione, correlata alla distribuzione di fase, si traduce in un'attenuazione del segnale di risonanza Magnetica, attenuazione che dipende quantitativamente dalle caratteristiche del gradiente e dal coefficiente di diffusione (D).

I dati di Risonanza magnetica sono acquisiti tramite uno scanner. Per cui, per ottenere reti cerebrali è necessario fare una scansione di un soggetto umano con un sistema di risonanza magnetica, ottenere i relativi file DICOM (lo standard per lo scambio di immagini mediche tra diversi sistemi operativi e anche tra diversi sistemi di comunicazione, Digital Imaging and COmmunications in Medicine), e procedere con i passi

della trattografia e della connettomica. Di seguito si riportano i principali concetti utilizzati. In Figura (Figura 4.1), è visualizzata un'immagine di risonanza ottenuta con una macchina a 3-Tesla della General Electrics.

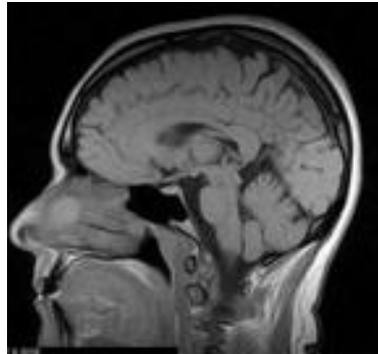


Figura 4.1. Immagine di Risonanza Magnetica, prodotta dallo standard DICOM.

Nel capitolo 7, si presenterà un esempio specifico di come le immagini di risonanza e le reti cerebrali connesse possono essere utilizzate per fare una serie di inferenze su malattie neuro-degenerative.

I file di interesse per la Connettomica e la Trattografia sono realizzati nella prima fase della scansione di Risonanza Magnetica. Infatti, i file DICOM contengono anche altri tipi di file, detti T1 e TDI, di solito ottenuti per Risonanza Magnetica Diffusiva, in una singola sequenza.

Per arrivare dalle scansioni Immagini di Risonanza Magnetica alle reti è necessario:

- a. Determinare il tensore di diffusione, che caratterizza la diffusione delle molecole d'acqua all'interno del cervello e attraverso cui si risale alla presenza e all'orientazione delle fibre. (Figura 4.2).
- b. Tracciare le fibre a partire dalla fibre stesse (Figura 4.3). La Diffusion Spectrum Imaging (Wedeen, 1996; Wedeen et al., 2005, 2008) è una tecnica caratterizzata dall'acquisizione di un gran numero di immagini, applicando un forte gradiente magnetico in direzioni multiple. Questo processo, non solo richiede lunghi tempi di acquisizione, ma anche un hardware ad alte prestazioni. Per queste ragioni, solo negli ultimi anni questo metodo ha iniziato a giocare un ruolo fondamentale nel Brain Imaging, spiegando perché la DTI, tecnica molto più semplice e che non necessita di alte prestazioni, è largamente usata, nonostante i suoi limiti (Krueger, 2008). In Figura 4.3, sono riportate differenze di dettaglio nei due metodi di tracciamento.

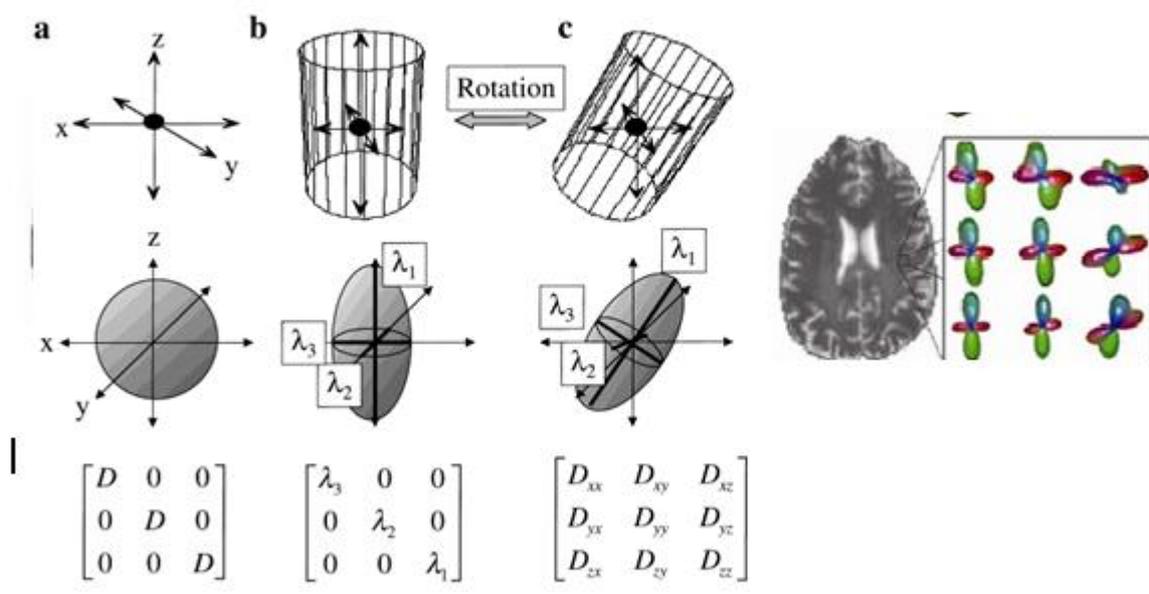


Figura 4.2. Metodo di determinazione del vettore di diffusione, ripresa da Behrens e collaboratori (2007). Uno dei problemi fondamentali di questo metodo è l'impossibilità di trattare l'orientamento di diverse fibre all'interno del singolo voxel.

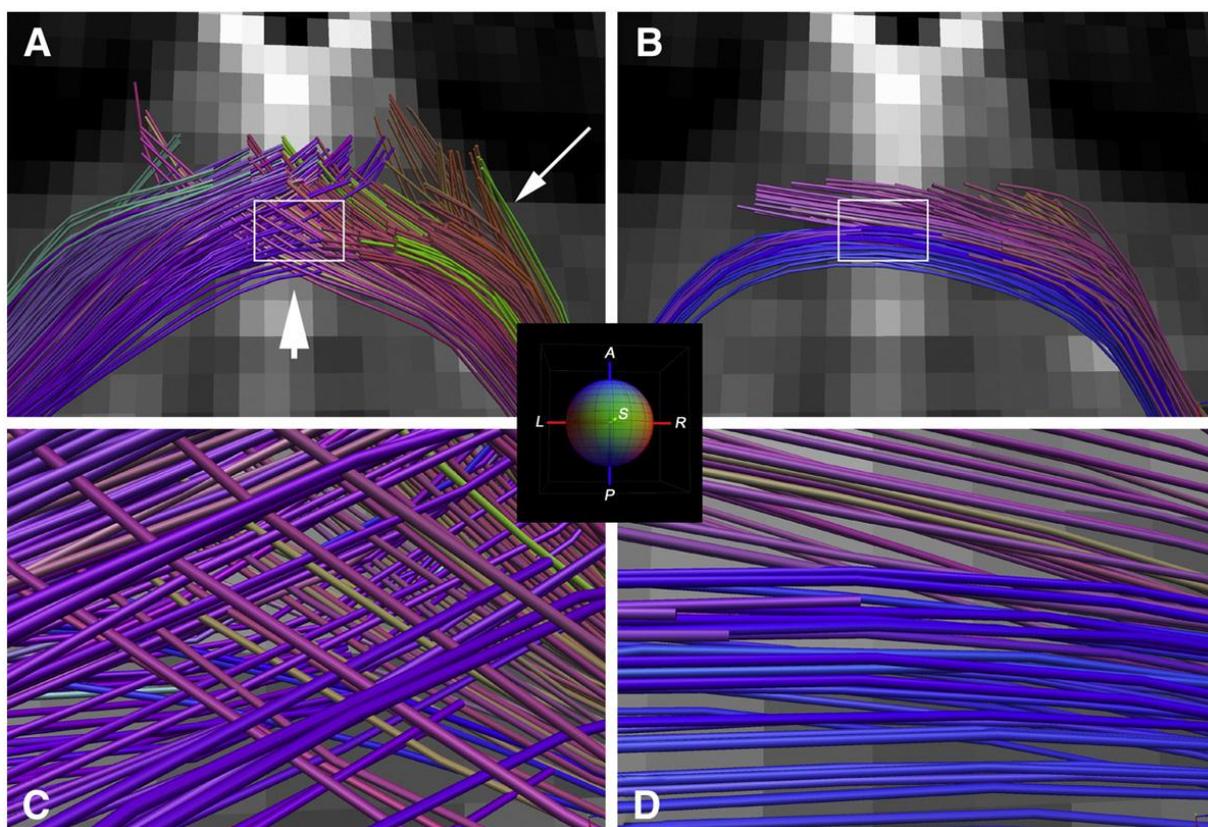


Figura 4.3. Tracciamento delle fibre con differenze significative tra DSI (a sinistra) vs DTI (a destra).

- c. Segmentare le aree corticali e sub-corticali, aventi una risoluzione maggiore rispetto a quelle diffuse;
- d. Determinare la matrice di adiacenza, per individuare quali aree del cervello sono connesse dalle fibre determinate al passo b);
- e. Tracciare e analizzare la rete ottenuta dalla matrice di adiacenza.

Più in dettaglio, per passare dal tensore di diffusione alla costruzione di una matrice di connessione, è necessario seguire i seguenti passi (vedi Figura 4.4):

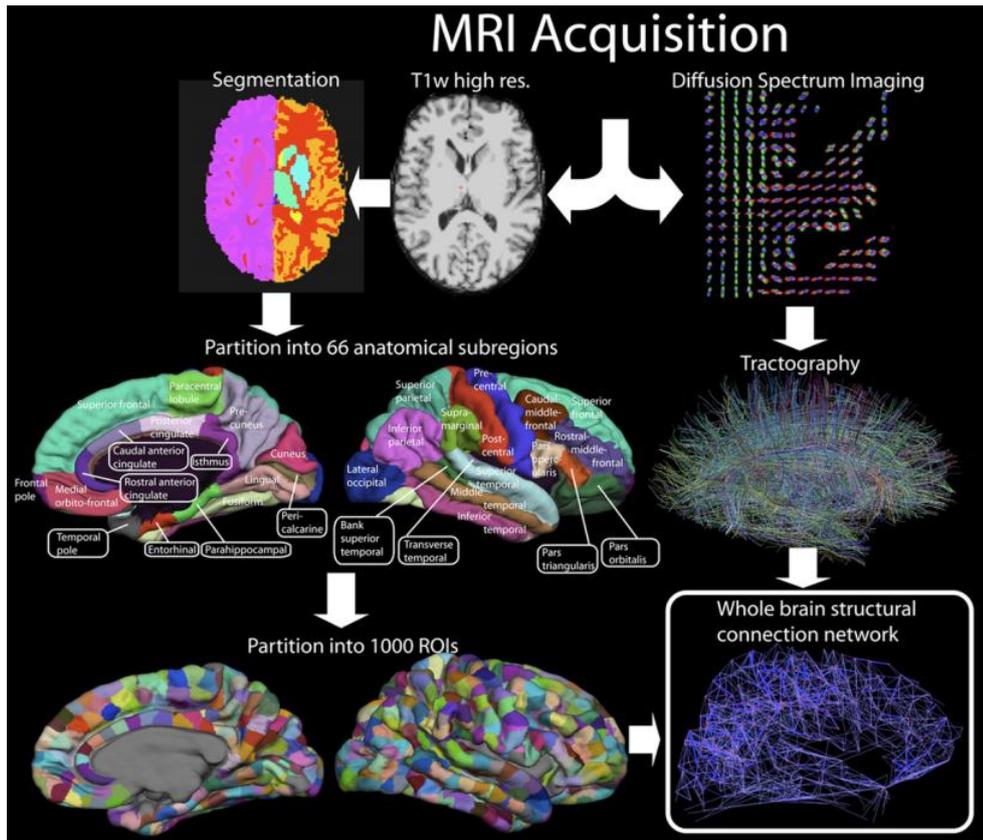


Figure 4.4. Figura riassuntiva di tutti i passaggi del metodo della Connettomica (ripresa da Hagmann , 2010).

1. l'acquisizione di DTI da Immagini di Risonanza Magnetica attraverso T1 pesate ad alta risoluzione. I principali parametri che determinano il segnale di Immagini di Risonanza Magnetica sono la densità protonica e i tempi di rilassamento T1 e T2. Il parametro T1 è definito come il tempo necessario per il recupero del 63% della magnetizzazione macroscopica longitudinale. Il parametro T2, il tempo di rilassamento, viene definito come il tempo necessario per la cancellazione del 63% della magnetizzazione trasversale macroscopica.
2. Segmentazione della sostanza bianca del cervello. La Trattografia della Materia Bianca (White Matter Tractography) comprende un insieme di tecniche che consentono di stimare la connettività e l'orientamento delle fibre cerebrali, attraverso l'uso di immagini ottenute per diffusione

Risonanza Magnetica Diffusiva. Gli algoritmi utilizzati sono numerosi e si dividono in sistemi di ottimizzazione deterministici, probabilistici e globali. Algoritmi deterministici identificano una traiettoria unica dal punto di partenza selezionato, e il risultato è un percorso che collega regioni discrete del cervello.

3. Trattografia della Materia Bianca. Algoritmi probabilistici generano traiettorie molto più probabili da un unico punto di partenza, attraverso il collegamento a voxel diversi e assegnando ad ogni traiettoria un indice di probabilità diversa. Alla fine, algoritmi di ottimizzazione globale generano il percorso ottimale tra due regioni del cervello.

Tutti i tipi di algoritmi seguono fasi di processo comuni, tra cui la scelta del punto di partenza, la fine della traiettoria e la strategia per la scelta delle fibre. Ad ogni passo, la traiettoria segue una direzione stimata e la propagazione si arresta in base a vari criteri, tra cui le intersezioni con voxel con anisotropia bassa o grande curvatura locale della traiettoria.

Bassa anisotropia frazionaria è utilizzata per escludere dall'analisi le zone di materia grigia nella fibra. I termini del tensore descrivono il coefficiente di diffusione apparente (ADC) di molecole di acqua in tutte le direzioni. Calcolando gli autovalori del tensore e gli autovettori corrispondenti, è possibile ottenere informazioni su qual'è la direzione del moto preferita di molecole nel tessuto cerebrale. La direzione preferita è rappresentata da un corrispondente autovettore con modulo di autovalore maggiore. L'algoritmo più semplice e più diffuso è la trattografia Streamline (STM), secondo il quale ogni fibra del cervello è disegnata in modo che sia tangente alla direzione principale di diffusione. È efficace in regioni con anisotropia elevata, ma non riesce dove le fibre si intersecano. Sono stati proposti diversi schemi di acquisizione in grado di rilevare "the fiber-crossing problem", come riportato in Figura 4.5.

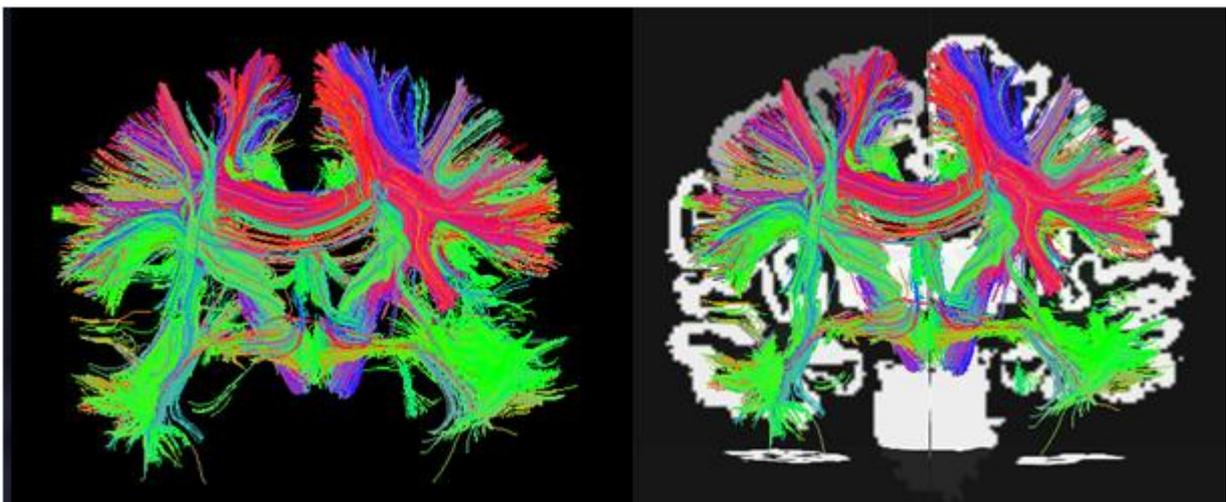


Figura 4.5. Risultato del processo di trattografia da Risonanza Magnetica Diffusiva, dove è possibile vedere come si incrociano le fibre (Wedeen et al., 2008).

4. Segmentazione in regioni di interesse (ROI) (Hagmann , 2010). Brodmann (1909), propose di separare la corteccia in 47 settori, ciascuno dei quali effettua diversi compiti funzionali (Figura 4.6). L'intuizione di Brodmann fu poi riconosciuta in un gran numero di studi funzionali, a dimostrazione della specializzazione regionale o raggruppamenti locali delle funzioni cerebrali. Ad esempio, gli input visivi sono gestiti nel lobo occipitale, stimoli uditivi nei lobi parieto-temporale di giunzione, mentre compiti motori sono condensati nel lobo frontale, e il raggruppamento funzionale si basa su connessioni cortico-corticali (Petersen et al., 1989). Collegamenti funzionali tra le strutture corticali e sub-corticali più isolate sono stati però dimostrati, introducendo il concetto di flussi di processo gerarchici, dove avviene la decomposizione o combinazione sensoriale o motoria.

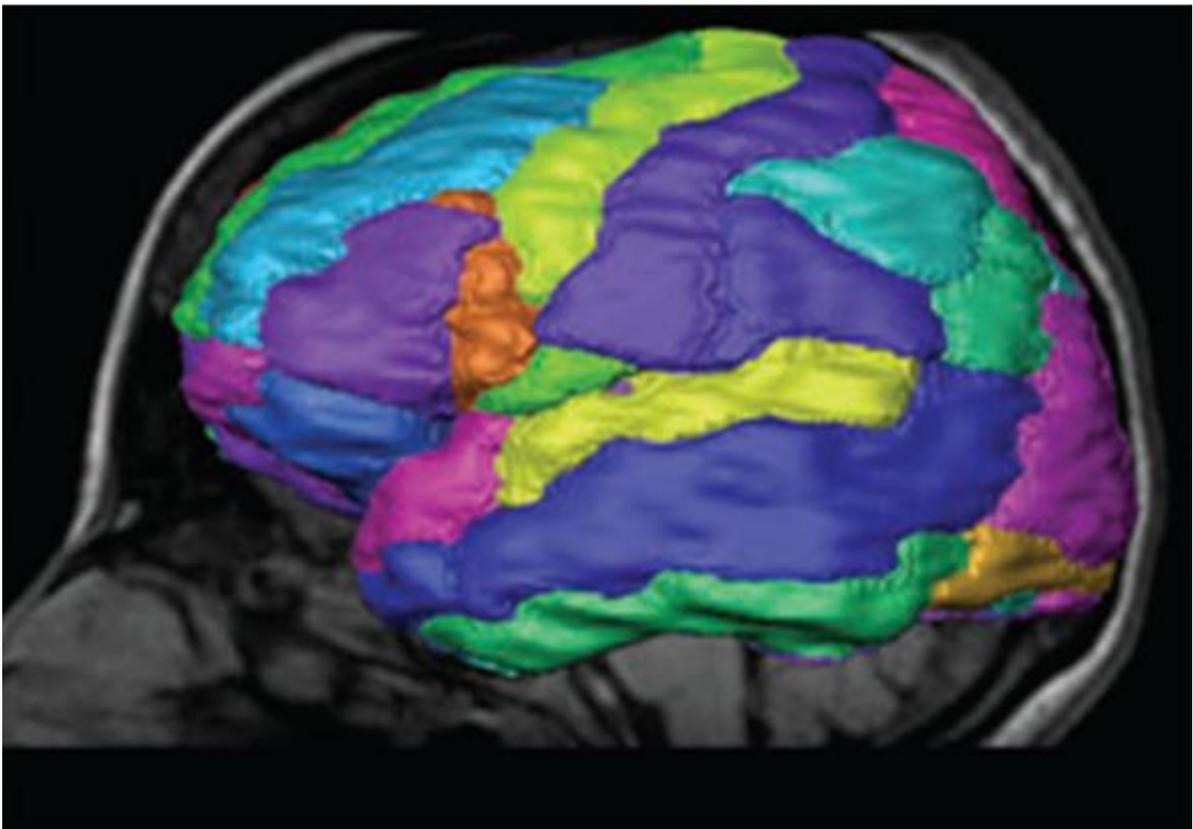


Figura 4.6. Immagine ottenuta dal software free surfer (www.freesurfer.org), che permette la segmentazione del cervello in aree distinte. Hagmann (2010) utilizza tale software per partizionare la corteccia in 1015 Regions Of Interest (ROI) omogenei, di circa 1.5 cm^2 che daranno luogo ad una rete di 1015 nodi..

5. La scelta dei nodi di una rete è effettuata secondo atlanti differenti che consentono di effettuare una segmentazione del cervello per diverse aree (regione di interesse). Questi atlanti codificano diverse aree del cervello relative a specifiche caratteristiche morfologiche o funzionalità.

Segmentazioni cerebrali diverse in regioni di interesse più o meno definite, corrispondenti a differenti strutture strutturali e funzionali, produrranno reti con 83 o 1015 nodi (Figura 4.7). Questa è una scelta molto delicata, e dipende dallo scopo dell'analisi, come indicato in Sporns (2005). La mancanza di un modello universalmente accettato per la segmentazione è un problema ben noto. Ai fini degli studi specifici descritti nel capitolo 7, si è scelto di utilizzare la rete a 83-nodi.

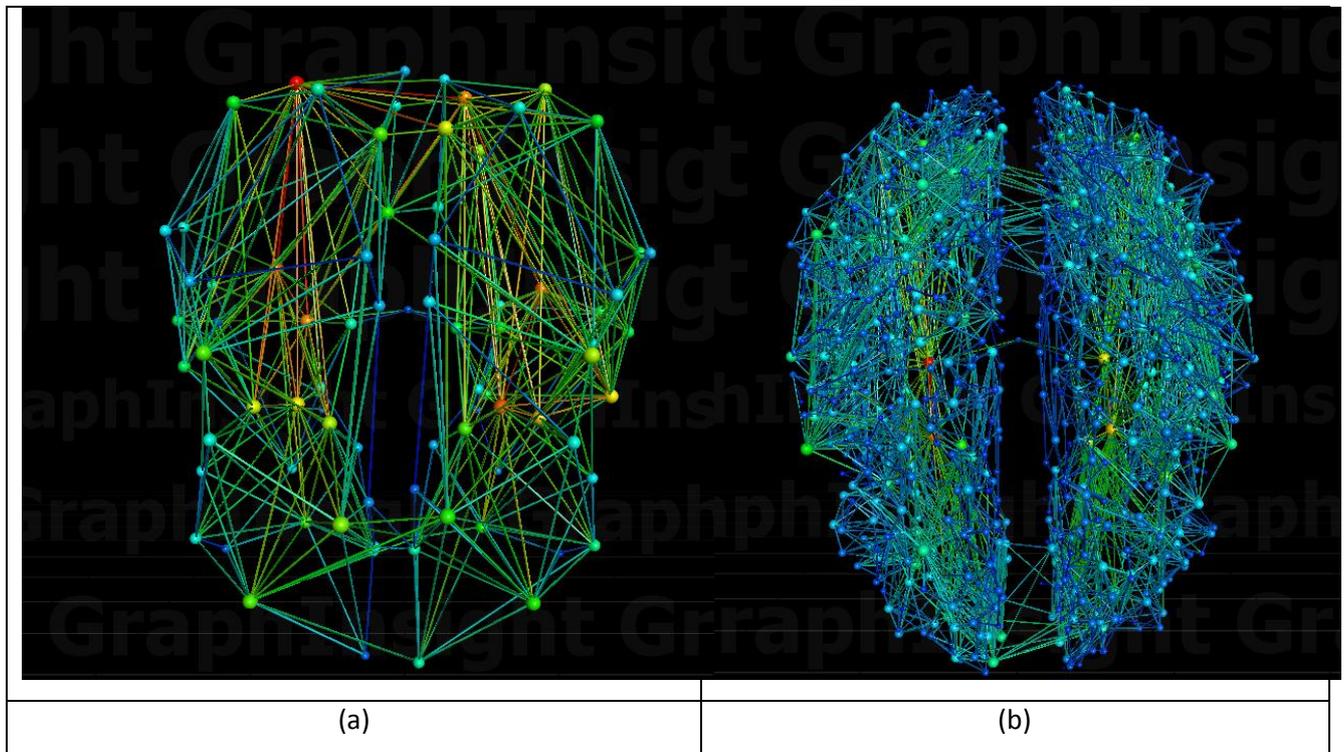


Figura 4.7. Reti realizzati con il software Connectome, rispettivamente con 83 (a) e 1015 (b) nodi.

Le Regioni di Interesse (che individuano le aree cerebrali citate precedentemente) per una rete di 83 nodi sono riportate in Tabella 4.1:

1	ctx-rh-lateralorbitofrontal	43	ctx-lh-parsorbitalis
2	ctx-rh-parsorbitalis	44	ctx-lh-frontalpole
3	ctx-rh-frontalpole	45	ctx-lh-medialorbitofrontal
4	ctx-rh-medialorbitofrontal	46	ctx-lh-parstriangularis
5	ctx-rh-parstriangularis	47	ctx-lh-parsopercularis
6	ctx-rh-parsopercularis	48	ctx-lh-rostralmiddlefrontal
7	ctx-rh-rostralmiddlefrontal	49	ctx-lh-superiorfrontal
8	ctx-rh-superiorfrontal	50	ctx-lh-caudalmiddlefrontal
9	ctx-rh-caudalmiddlefrontal	51	ctx-lh-precentral
10	ctx-rh-precentral	52	ctx-lh-paracentral
11	ctx-rh-paracentral	53	ctx-lh-rostralanteriorcingulate
12	ctx-rh-rostralanteriorcingulate	54	ctx-lh-caudalanteriorcingulate
13	ctx-rh-caudalanteriorcingulate	55	ctx-lh-posteriorcingulate
14	ctx-rh-posteriorcingulate	56	ctx-lh-isthmuscingulate

15	ctx-rh-isthmuscingulate	57	ctx-lh-postcentral
16	ctx-rh-postcentral	58	ctx-lh-supramarginal
17	ctx-rh-supramarginal	59	ctx-lh-superiorparietal
18	ctx-rh-superiorparietal	60	ctx-lh-inferiorparietal
19	ctx-rh-inferiorparietal	61	ctx-lh-precuneus
20	ctx-rh-precuneus	62	ctx-lh-cuneus
21	ctx-rh-cuneus	63	ctx-lh-pericalcarine
22	ctx-rh-pericalcarine	64	ctx-lh-lateraloccipital
23	ctx-rh-lateraloccipital	65	ctx-lh-lingual
24	ctx-rh-lingual	66	ctx-lh-fusiform
25	ctx-rh-fusiform	67	ctx-lh-parahippocampal
26	ctx-rh-parahippocampal	68	ctx-lh-entorhinal
27	ctx-rh-entorhinal	69	ctx-lh-temporalpole
28	ctx-rh-temporalpole	70	ctx-lh-inferiortemporal
29	ctx-rh-inferiortemporal	71	ctx-lh-middletemporal
30	ctx-rh-middletemporal	72	ctx-lh-bankssts
31	ctx-rh-bankssts	73	ctx-lh-superiortemporal
32	ctx-rh-superiortemporal	74	ctx-lh-transversetemporal
33	ctx-rh-transversetemporal	75	ctx-lh-insula
34	ctx-rh-insula	76	Left-Thalamus-Proper
35	Right-Thalamus-Proper	77	Left-Caudate
36	Right-Caudate	78	Left-Putamen
37	Right-Putamen	79	Left-Pallidum
38	Right-Pallidum	80	Left-Accumbens-area
39	Right-Accumbens-area	81	Left-Hippocampus
40	Right-Hippocampus	82	Left-Amygdala
41	Right-Amygdala	83	Brain-Stem
42	ctx-lh-lateralorbitofrontal		

Tabella 4.1. Le 83 aree cerebrali utilizzate per realizzare la rete con 83 nodi, attraverso il software Connectome.

6. Costruzione di una matrice di adiacenza per determinare il collegamento tra le varie aree cerebrali. (Figura 4.8).

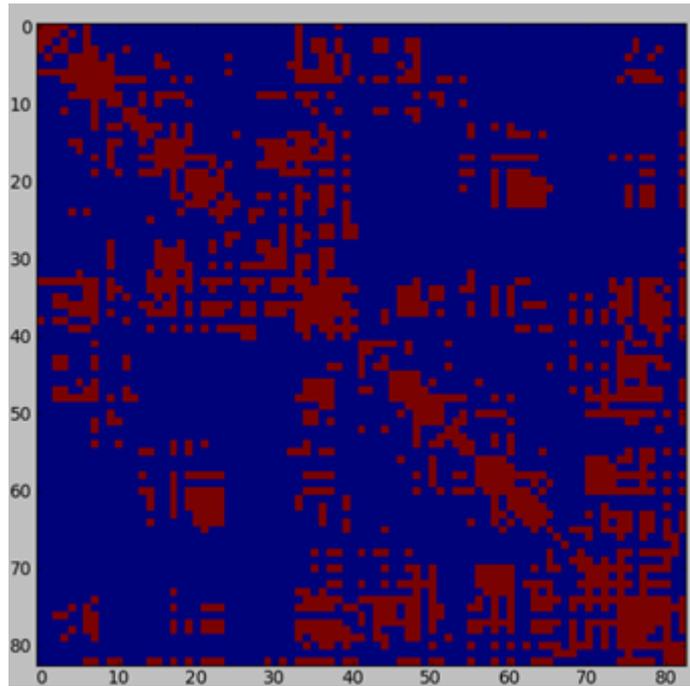


Figura 4.8. La matrice di adiacenza determina il collegamento fra le varie aree cerebrali.

Il sistema utilizza FREESURFER per la segmentazione, FSL per visualizzare i file DICOM, DTK TRACKVIS per determinare il tensore, e trarne le fibre e CONNECTOME per creare la matrice. I seguenti algoritmi per calcolare la matrice di adiacenza:

- (i) NATIVE FREESURFER (83 ROI), realizza una rete con 83 punti, dividendo il file DTI ad alta risoluzione e il file T1 di Risonanza Magnetica pesata in 83 regioni di interesse.
- (ii) LOSANNA 2008 (Hagmann et al., 2010), restituisce una rete con 1015 punti, dividendo il file DTI ad alta risoluzione e file T1 di Risonanza Magnetica pesata in 1015 regioni di interesse. In questo modo, i nodi corrispondono a grandi aree cerebrali o Region of Interest (ROI).

- f. Una volta che la rete è stata ottenuta diverse analisi di rete e le relative statistiche come quelle descritte nel Capitolo 2 possono essere realizzate, per individuare le dinamiche dei circuiti della struttura cerebrale.

Nella parte sperimentale della tesi, in particolare nel Capitolo 7, sarà analizzato un campione di soggetti sani e soggetti con una variante del Parkinson e saranno individuate le dinamiche cerebrali per determinare come si modifica l'architettura cerebrali in soggetti malati.

Capitolo 5. Modellazione delle lingue naturali e del significato

5.1. Background di riferimento sulle reti del linguaggio umano

Un'ipotesi forte sulla natura del linguaggio è posta da Ramon Ferrer i Cancho & Ricard Solé (2002), secondo i quali, avendo il genere *Homo* il bisogno di comunicare e condividere informazioni con gli altri conspecifici, il linguaggio è stato un vantaggio in termini evolutivi. Sviluppandosi da un piccolo insieme di termini iniziali, legati agli istinti primari, e alle emozioni, la rete del linguaggio umano si è sviluppata nel corso dell'evoluzione, con l'aggiunta di altri termini, tramite collegamenti non casuali. I due autori considerano il lessico del linguaggio umano come un grafo, composto da nodi e archi, dove i nodi rappresentano le parole che lo compongono, e gli archi le relazioni che uniscono le parole. Dal punto di vista formale, essi considerano il linguaggio umano come un grafo Ω , costituito dall'insieme di nodi W (words) e di collegamenti E (edges), per cui la rete è espressa da $\Omega = (W, E)$. Il grado di connettività di un nodo o di una parola è il numero di legami che lo connettono con le altre parole del grafo. Gli autori dimostrano che le lingue naturali hanno caratteristiche di piccolo mondo: un alto *coefficiente di clustering*, la rete esprime una forte tendenza dei nodi a raggrupparsi, formando agglomerati di nodi o gruppi di sinonimi, e un bassissimo *grado di separazione* tra i nodi, in proporzione alla estensione del lessico umano. Un altro fenomeno che questi due autori affrontano è la co-occorrenza di due parole nel lessico di una lingua, che secondo loro, non è casuale, ma ubbidisce a leggi precise: due parole possono essere unite da un legame grammaticale (nome e aggettivo), o da altre forme di interdipendenza (verbo, avverbio). La struttura di queste relazioni rende la probabilità di occorrenza di gruppi di parole non casuale, come accadrebbe in una rete random, ma la rete di parole si sviluppa nel tempo, aumentando il numero delle parole che si connettono ad altre parole, a loro volta altamente connesse (Ferrer i Cancho & Solé, 2001). La *distanza media* tra due parole nel lessico indica che queste hanno una grande possibilità di connettersi: in genere la distanza tra un nome e un aggettivo è pari ad 1 (1 solo passaggio è necessario per connettere i due nodi), quella tra un verbo e il suo complemento oggetto può essere pari a 2. (verbo, articolo, nome). Il grado di separazione tra due nodi random nella rete del linguaggio è pari a 3 passaggi per raggiungere un qualunque

elemento, partendo a caso da un elemento qualsiasi (Ferrer i Cancho & Solé, 2001). La distanza tra parole che co-occorrono frequentemente è molto breve. Potremmo dire, facendo un paragone con la musica, che i rapporti di co-occorrenza sono analoghi ai rapporti di consonanza, più sono bassi più sono frequenti.

Sempre secondo i due ricercatori, la rete del linguaggio umano presenta invarianza di scala, segue una legge di potenza, ha numerosi *hub*, e si sviluppa secondo il modello di *crescita preferenziale*.

I nodi non hanno lo stesso grado nella rete. Esistono nodi con un gran numero di link rispetto alla media. La distribuzione dei legami intercetta una legge di potenza. In una lingua, gli hub rinforzano la solidità e l'interconnessione della rete, permettendo il passaggio immediato da un nodo all'altro. Quando la rete cresce, i nodi non si attaccano agli altri nodi in modo casuale. I nodi con grado maggiore, che corrisponde al nodo con più alta frequenza, sono preferiti. Questi nodi (e le parole che sono loro collegate) sorreggono i processi cognitivi di comprensione e produzione, una volta che sono memorizzati i concetti riferiti ai termini lessicali, proprio perché sono molto frequenti. La rete del linguaggio acquisisce nodi nuovi in continuazione.

Mark Steyvers & Josh Tenenbaum (2004), per analizzare la natura del linguaggio umano, considerano tre diversi tipi di reti: a. una rete associativa, realizzata sulla base di libere associazioni ottenute da un campione di circa 6000 soggetti; b. la rete WordNet, dove i nodi rappresentano concetti, collegati fra loro attraverso relazioni di antinomia, iperonimia, sinonimia ecc.; c. il Roget's Thesaurus, che classifica circa 30.000 nomi, in base al criterio delle categorie semantiche. Tutte e tre i tipi di rete hanno caratteristiche di piccolo mondo e invarianza di scala, presenza di hub, alto grado di connettività, distribuzione delle connessioni secondo una legge di potenza, confermando i risultati già ottenuti da Ferrer i Cancho & Solé (2001) per le reti del linguaggio umano. Sulla base di queste analisi sulla natura della lingua, Steyvers e Tenenbaum (2004) suggeriscono, come modello di sviluppo di una rete lessicale, una struttura in crescita, che acquisisce nuovi nodi, tramite criteri di incremento preferenziale. Nel processo di evoluzione della rete, i due ricercatori individuano nell'età del nodo, la caratteristica privilegiata intorno alla quale si accresce la rete lessicale. Gli elementi acquisiti nel primo stadio dello sviluppo del linguaggio sono, probabilmente, quelli più salienti, per cui sono anche i più interconnessi e quelli più resistenti nel caso di malattie neurodegenerative o di perdita di memoria con l'invecchiamento.

Un altro studio per capire la natura del linguaggio umano e dell'organizzazione del lessico è stato realizzato da Sigman & Cecchi (2002). Considerando il linguaggio come uno specchio della mente, gli autori sostengono che uno dei problemi più importanti della scienza contemporanea è la caratterizzazione del lessico, per capire la natura del funzionamento dei processi cognitivi soggiacenti la produzione linguistica. Gli autori studiano le relazioni semantiche relative a antonimia, iperonimia (iponimia e meronimia), sostenendo che, mentre l'antinomia è intuitivamente nota, l'iperonimia e l'iponimia non lo sono per nulla.

Un iponimo è un termine che acquisisce tutte le caratteristiche dell'iperonimo, che in questo senso è un concetto più generico. Nell'esempio di iponimo, che gli autori riportano, il termine "rovere" è un iponimo del termine "albero". Questo modo di procedere, che si chiama rapporto "E' un...", è generalmente utilizzato dai vocabolari, che fanno uso dell'iperonimo per spiegare l'iponimo. La meronimia, invece utilizza il rapporto "è parte di". Per esempio, il termine "ramo" è parte o meronimo del termine "albero". La polisemia, forme di parole comuni che determinano ambiguità nell'estrazione del corretto significato, è una dinamica di collegamento nella rete semantica del linguaggio umano, in relazione alle specializzazioni del significato. Per cui i collegamenti polisemici hanno, secondo gli autori, un profondo impatto nell'organizzazione della rete, realizzando configurazioni di piccolo mondo e invarianza di scala nel lessico, con molti hub, che fungono da connettori con altri significati. Gli autori dimostrano che il sistema WordNet ha proprietà tipiche delle reti e dei sistemi complessi auto-organizzati, e che la polisemia è una caratteristica portante della rete linguistica, che spiega perché questo fenomeno si manifesta in quasi tutte le lingue. Rapporti di antinomia e polisemia presentano simmetrie, mentre iperonimia e meronimia hanno iponimia e omonimia come organizzazioni inverse. Vedremo nei paragrafi successivi come questi rapporti si sviluppano anche come patterns nelle reti dei termini emozionali.

Mentre la maggior parte degli studi sulle reti del linguaggio si concentrano sulle proprietà strutturali delle stesse, altri metodi di identificazione di caratteristiche di raggruppamento in reti di interconnessione sintattica sono rintracciabili in Ferrer i Cancho e collaboratori (Ferrer i Cancho et al., 2007a, b), Mukherjee e colleghi (Mukherjee et al., 2009). Questi ultimi, usano una tecnica puramente strutturale, l'analisi spettrale per individuare le correlazioni globali in Phonet, individuando la co-occorrenza di consonanti in rete, che forgiavano la struttura degli inventari delle consonanti, fornendo anche una misura quantitativa della loro importanza. Questa tecnica può essere applicata per studiare la struttura dei livelli delle lingue naturali. Le strutture principali di piccolo mondo e di invarianza di scala si ripetono anche per lingue orientali come il cinese (Zhou et al., 2008).

Oltre allo studio delle relazioni semantiche e sintattiche un'altra direzione di analisi delle lingue naturali studia le reti di utenti di lingua (ad esempio le reti sociali di persone della comunità linguistica), e il loro ruolo nella evoluzione del linguaggio (Ke, 2007). In questo articolo, l'autore analizza quattro tipi di reti sociali, che presentano due categorie di dinamiche di diffusione sociale di linguaggio. Resta ancora molto da investigare. Ma fondamentale nello studio sulle reti linguistiche è che tipo di modellazione e segmentazione usare nel caso del parlato, uno dei temi fondamentali della ricerca sulle reti linguistiche.

La scienza delle reti permette di capire in profondità la natura e le radici del significato lessicale, nelle sue articolazioni. Si farà riferimento ai concetti operativi delle lingue e dei linguaggi per realizzare la modellazione di lingue naturali in WordNet, per introdurre l'analisi del linguaggio umano attraverso la

scienza delle reti (Barabási & Albert, 1999; Barabási et al., 1999; Barabási & Bianconi, 2001; Barabási, 2004; Caldarelli, 2007).

Numerosi sono i problemi relativi alla modellazione delle relazioni lessicali. Nell'approccio informatico per la realizzazione di database Entità/Relazione, o nei linguaggi di programmazione, per esempio, le relazioni lessicali sono trattate come primitive, quindi date per via assiomatica. Tali approcci, derivati dai modelli di Chomsky e Fodor hanno dato vita alla modellazione semantica del significato, utilizzato nei linguaggi XML (*eXtensible Markup Language*, Linguaggio di marcatura estensibile). Invece, nella Linguistica computazionale, i ricercatori stanno indagando sui seguenti problemi:

- a. quanti tipi di relazioni semantiche esistono (ovvero quante sono e cosa sono queste relazioni primitive)?
- b. è il numero di relazioni fisso o mutabile, e in rapporto a cosa variano?
- c. Possono esperimenti con soggetti umani aiutare a modellare la definizione dei problemi inerenti i confini di categoria concettuale, per creare modelli computazionali intelligenti e distinguere il significato di un concetto da un altro?

Questi interrogativi scientifici sono particolarmente rilevanti per i sistemi di disambiguazione semantica. Per cui il problema che si sono posti questi ricercatori che costruiscono sistemi di linguistica computazionale è come contare le differenze tra un termine e l'altro (Arnold, 1971). Arnold ha dimostrato che questa distanza non è euclidea. In particolare, Arnold ha usato il *Multidimensional Scaling Method* di Kruskal (1964). Tale metodo è utilizzato in Psicologia per rappresentare la forma di un oggetto in più di una dimensione nello stesso tempo, avendo in mente che un soggetto percepisca le differenze con altri oggetti, correlati per somiglianza o differenza all'oggetto stimolo. Questo metodo serve a determinare le scale di distanza cognitiva tra gli oggetti lungo l'asse similarità-differenza, per poterle poi rappresentare su mappe multidimensionali. Arnold ha dimostrato che, dopo aver scalato in modo unidimensionale gruppi di concetti sulle qualità strettamente legate al differenziale semantico 3, ordinando le serie di aggettivi, nomi, e verbi attraverso il metodo di organizzazione multidimensionale, comparando con le risposte di apprezzamento sulle diversità di significato tra un termine e un altro, realizzato da soggetti umani, ha trovato che le distanze semantiche tra significati non sono euclidee. Questo risultato suggerisce che nella formazione di frasi, tutte le dimensioni di diversità sono soppresse tranne la dimensione 1, che discrimina al massimo i membri di una coppia di concetti. Il termine "componente a distanza massima" è stato usato per caratterizzare il modello migliore di distanza. Tutti e tre le serie di differenze hanno permesso l'estrazione di almeno 4 dimensioni (Kruskal, & Wish, 1978). La scala di differenza semantica prediceva abbastanza bene le distanze tra nomi e aggettivi.

In questo capitolo, sarà realizzata un modellazione del sistema WordNet, un database di significati strutturati secondo precise regole psicolinguistiche di organizzazione strutturale. In particolare, si forniranno i passi metodologici per la creazione di un vocabolario, si sceglieranno i termini da ricercare, si modellerà la crescita di reti semantiche. Il Capitolo 8 riporterà i risultati sperimentali di tale approccio. I risultati sperimentali dimostrano che le particelle emergenti nelle reti semantiche hanno più di 4 dimensioni.

5.2. Modellazione di WordNet come rete

Indichiamo con $W = \{w_1, w_2, \dots, w_N\}$ l'insieme di tutti i termini di WordNet. Supponiamo inoltre che un termine di W stabilisca un qualche tipo di relazione $R_1, R_2 \dots R_s$ con altri termini di W , in questo modo possiamo costruire delle *reti locali* e la *rete globale* di WordNet. Banalmente la rete globale R_G avrà come nodi tutti gli elementi di W . Gli archi saranno delle coppie di nodi (w_i, w_j) tra i quali sussista almeno una delle relazioni R_h .

La risorsa linguistica WordNet, scaricata dal sito dell'università di Princeton (<http://WordNet.princeton.edu/>), e successivamente trasformata in rete, presenta 139.999 nodi. Solo alcuni termini sono stati automaticamente eliminati, in quanto presentavano dei problemi di correttezza formale nell'inserimento nel database. Questo database è in continua crescita, per cui il suo contenuto fa riferimento al particolare momento nel quale le risorse vengono scaricate. Le relazioni tra i termini, sulle quali il sistema computazionale è stato costruito, consente di predisporre i dati linguistici secondo l'organizzazione dei significati semantici sopra espressi. In sintesi abbiamo sviluppato i seguenti modelli computazionali per indagare il sistema:

R_1 : Antonimia, o opposizione di significato, che si organizza nei seguenti elementi:

- (i) antinomia, che può essere data da una opposizione graduabile, lungo una scala graduabile (piacere, stato neutro, dispiacere)
- (ii) complementarità di significato, dove i termini non sono organizzati lungo una scala graduata e senza intermedi;
- (iii) inversione, tali termini indicano prospettive opposte, che possono essere diverse e simmetriche (di solito riferentesi a ruoli sociali, di parentela, relazioni-temporali e spaziali).

R_2 : Iperonimia. Questa funzione con l'iponimia crea la struttura gerarchica delle mappe concettuali. Secondo i linguisti queste relazioni svolgono una azione di classificazione tassonomica in tipi e sottotipi, con tassonomie scientifiche o correlate all'organizzazione concettuale di ogni soggetto umano.

Esistono relazioni tra termini che incorporano un rapporto tra significato generale e significato specifico (mobili/sedia-tavolo-cassapanca)

R_3 : Iponimie. Sono tutti i termini correlati attraverso la funzione "E' -UN", con una relazione di inclusione. Se X è iperonimo e Y è un iponimo allora X include Y (Esempio: I mobili includono le sedie).

R_4 : Implicazione. Tale funzione (Entailment nel sistema WordNet), identifica tutti i termini che sono in qualche modo implicati, dal punto di vista concettuale, con il termine di partenza.

R_5 : Similarità. Questa funzione implica una somiglianza fra termini.

R_6 : Sinonimie. La sinonimia o somiglianza di significato che si articola in:

- (iv) condivisione parziale del contenuto semantico;
- (v) condivisione del significato denotativo (il significato primitivo, estratto dal vocabolario), con una differenza nel significato connotativo (o psicologico, relativo al significato associato al termine);
- (vi) intercambiabilità perfetta;

R_7 : questa funzione, (indicata con Related in WordNet), individua i termini correlati al termine scelto per l'indagine. In tal modo, le espressioni che sono legate al concetto che il termine di partenza esprime possono essere indagati con ulteriori ricerche, da parte dei soggetti che usano WordNet. Inoltre, il sistema contiene anche informazioni che sono correlate al termine, anche dal punto di vista del livello morfologico (in poche parole, radice, e desinenze, derivazioni di altri termini dalla radice, derivazioni di nomi da verbi o di nomi di oggetti o di soggetti animati che compiono l'azione che il termine esprime, se il termine è un verbo).

R_8 : Descrizione (nel sistema WordNet indicato con il termine Overview), individua tutti i gruppi di sinonimi (synset) per il termine che si vuole indagare. La funzione restituisce tutti i significati del termine, per tutte le categorie sintattiche.

5.3. Creazione di un vocabolario

I nodi che abbiamo introdotto nella sezione precedente non hanno alcun attributo. Supponiamo ora di voler associare a questi termini un attributo grammaticale come aggettivo, nome etc. Per fare ciò dobbiamo costruire un vocabolario ed associare a ciascun termine della rete un attributo contenuto in tale vocabolario. A tal fine abbiamo costruito un sistema computazionale che opera nel modo seguente:

step 1. Costruzione del vocabolario

Partendo da una serie di vocabolari in linea messi a disposizione nel sito <http://dizionari.corriere.it/> abbiamo costruito un file xml per le due lingue italiano ed inglese. Il vocabolario inglese presenta 73.456 vocaboli e quello italiano 41747 termini.

Il modello utilizzato è riportato nella Figura 5.12.

In tale struttura, un <vocabolario> è formato da un numero di vocaboli (che ci dice la grandezza del vocabolario) e un insieme di elementi detti <vocabolo>. Ogni <vocabolo> è caratterizzato dalla <lingua>

(Italiano, Inglese, Spagnolo, ecc.), <attributo>, <relazione>, <significato>, e dal <termine> stesso. Ogni <attributo>, indica la qualifica grammaticale che il vocabolo può avere.

Le relazioni tra i termini sono state descritte nel paragrafo precedente (sinonimia, antinomia, ecc.).

I Significati indicano i vari significati che il termine può assumere a seconda del contesto. In questo vocabolario, essendo stato estratto in automatico, nel tag <significati> è indicato solo il link alla definizione del vocabolo. Il termine indica il vocabolo stesso.

```
<?xml version="1.0"?>
<vocabolario>
  <numero_vocaboli>73456</numero_vocaboli>
  - <vocabolo>
    <lingua>eng</lingua>
    - <attributi>
      <attributo>n.</attributo>
    </attributi>
    <relazioni/>
    <significati>http://dizionari.corriere.it/dizionario_inglese/Inglese/A/a_1.shtml</significati>
    <termine>a</termine>
  </vocabolo>
  - <vocabolo>
    <lingua>eng</lingua>
    - <attributi>
      <attributo>art.indeterm.</attributo>
    </attributi>
    <relazioni/>
    <significati>http://dizionari.corriere.it/dizionario_inglese/Inglese/A/a_2.shtml</significati>
    <termine>a</termine>
  </vocabolo>
  - <vocabolo>
    <lingua>eng</lingua>
    <attributi> </attributi>
    <relazioni/>
    <significati>http://dizionari.corriere.it/dizionario_inglese/Inglese/A/a_3.shtml</significati>
    <termine>a</termine>
  </vocabolo>
  - <vocabolo>
    <lingua>eng</lingua>
    - <attributi>
      <attributo>a.</attributo>
    </attributi>
    <relazioni/>
    <significati>http://dizionari.corriere.it/dizionario_inglese/Inglese/A/a_.shtml</significati>
    <termine>a.</termine>
  </vocabolo>
  - <vocabolo>
    <lingua>eng</lingua>
    - <attributi>
      <attributo>agg.</attributo>
    </attributi>
    <relazioni/>
    <significati>http://dizionari.corriere.it/dizionario_inglese/Inglese/A/A_2-.shtml</significati>
    <termine>A</termine>
  </vocabolo>
  .
  .
  .
```

Figura 5.12. Come si presenta il file xml del vocabolario inglese.

Step 2: Associazione di un colore per identificare l'attributo grammaticale di un termine

Nel file xml è possibile anche associare un particolare colore all'attributo grammaticale. In particolare colori differenti sono associati ai nomi, ai verbi, agli aggettivi ed altri avverbi (Figura 5.13), secondo codici colore, definiti di volta in volta dall'utente, ma appartenenti al sistema RGB.

```

<?xml version="1.0"?>
- <attributi>
  - <attributo>
    <colore b="139" g="28" r="208"/>
    <id>1</id>
    <nome>name</nome>
    - <riferimenti>
      <gramm>name</gramm>
      <gramm>s.</gramm>
      <gramm>s.m.</gramm>
      <gramm>s.f.</gramm>
      <gramm>m.</gramm>
      <gramm>f.</gramm>
      <gramm>pl.</gramm>
      <gramm>sost.</gramm>
      <gramm>pron.</gramm>
    </riferimenti>
  </attributo>
  - <attributo>
    <colore b="211" g="174" r="234"/>
    <id>2</id>
    <nome>adjective</nome>
    - <riferimenti>
      <gramm>adjective</gramm>
      <gramm>agg.</gramm>
    </riferimenti>
  </attributo>
  - <attributo>
    <colore b="177" g="222" r="190"/>
    <id>3</id>
    <nome>verb</nome>
    - <riferimenti>
      <gramm>verb</gramm>
      <gramm>v.</gramm>
      <gramm>v.tr.</gramm>
      <gramm>v.intr.</gramm>
      <gramm>v.rifl.</gramm>
    </riferimenti>
  </attributo>
  - <attributo>
    <colore b="38" g="172" r="77"/>
    <id>4</id>
    <nome>adverb</nome>
    - <riferimenti>
      <gramm>adverb</gramm>
      <gramm>avv.</gramm>
    </riferimenti>
  </attributo>
</attributi>

```

Figura 5.13. File xml che associa colori in formato RGB agli attributi grammaticali.

Step 3. Associazione dell'attributo ai termini di una rete

Per poter associare l'attributo grammaticale ai nodi della rete, abbiamo dovuto progettare e realizzare un sistema computazionale, in corso di sperimentazione, denominato *Semantic Network Creator*. Questo sistema, dopo aver individuato, tramite le impostazioni, i file xml relativi al vocabolario e alle definizioni

degli attributi, prima riportato, consente di importare una rete e di associare sia attributi grammaticali sia un codice colore ai nodi della rete stessa (Figura 5.14).

The screenshot shows the SemanticNetworkCreator application window. On the left, there is a sidebar with a tree view containing 'File Rete', 'File Vocabolario', and 'File Attributi'. The main area displays a table with the following data:

Id	Label	Attribute	Color
1	president		
2	yeltsin		
3	consolidated	adjective	234; 174; 211
4	his	adjective	234; 174; 211
5	authority		
6	yesterday	adverb	77; 172; 38
7	by		
8	persuading		
9	russia's		
10	highest		
11	parliament		
12	to		
13	shelve	verb	190; 222; 177
14	an		
15	attempt		
16	trim	verb	190; 222; 177
17	powers		
18	and		
19	defusing		
20	a		
21	dangerous	adjective	234; 174; 211
22	row		
23	with		
24	ukraine		
25	over		
26	the	adverb	77; 172; 38
27	black	adjective	234; 174; 211
28	sea		

On the right side of the window, there is a button labeled 'ASSOCIA ATTRIBUTI VOCABOLARIO'.

Figura 5.14. Associazione dei nodi della rete degli attributi presenti nel vocabolario e ad un colore.

Il software funziona nel seguente modo. Dopo aver caricato i due file `attribute_definition.xml` e `vocabolario.xml`, è possibile importare la rete alla quale si vogliono associare gli attributi e premere il pulsante "Associa attributi vocabolario". Il software caricherà per ogni termine, se trovato nel vocabolario, il suo attributo ed il colore corrispondente ad esso (specificati nel file `attribute_definition`).

Per i termini che non sono presenti all'interno del vocabolario verranno caricati tutti gli attributi a disposizione.

È possibile, inoltre, fare un uso libero del sistema, associando di volta in volta i significati e il codice colore, all'interno di relazioni semantiche stabilite di volta in volta dall'utente. Selezionando un nuovo attributo, per i termini che non sono presenti all'interno del vocabolario, verrà creato il vocabolo all'interno del vocabolario contenente l'attributo appena specificato.

Una volta terminata la procedura è possibile esportare la rete nel formato "graphml" tramite File->esporta. Tale formato è riconosciuto dai principali software di visualizzazione di reti. In Figura 5.15, una rete ottenuta dal British National Corpus (BNC), un database di testi scritti e orali di lingua Inglese, visualizzata tramite il software *Gephi*. Si possono notare i colori dei nodi. La dimensione dei nodi è invece calcolata automaticamente, ponendo il raggio proporzionale al grado del nodo. I termini non presenti nel vocabolario, manualmente aggiornati dall'utente potrebbero essere utili nello studio filologico di corpora storici. L'approccio utilizzato può essere esteso, realizzando appropriati vocabolari, una lista di termini con i loro attributi, ad altri corpora, comprese raccolte di file musicali.

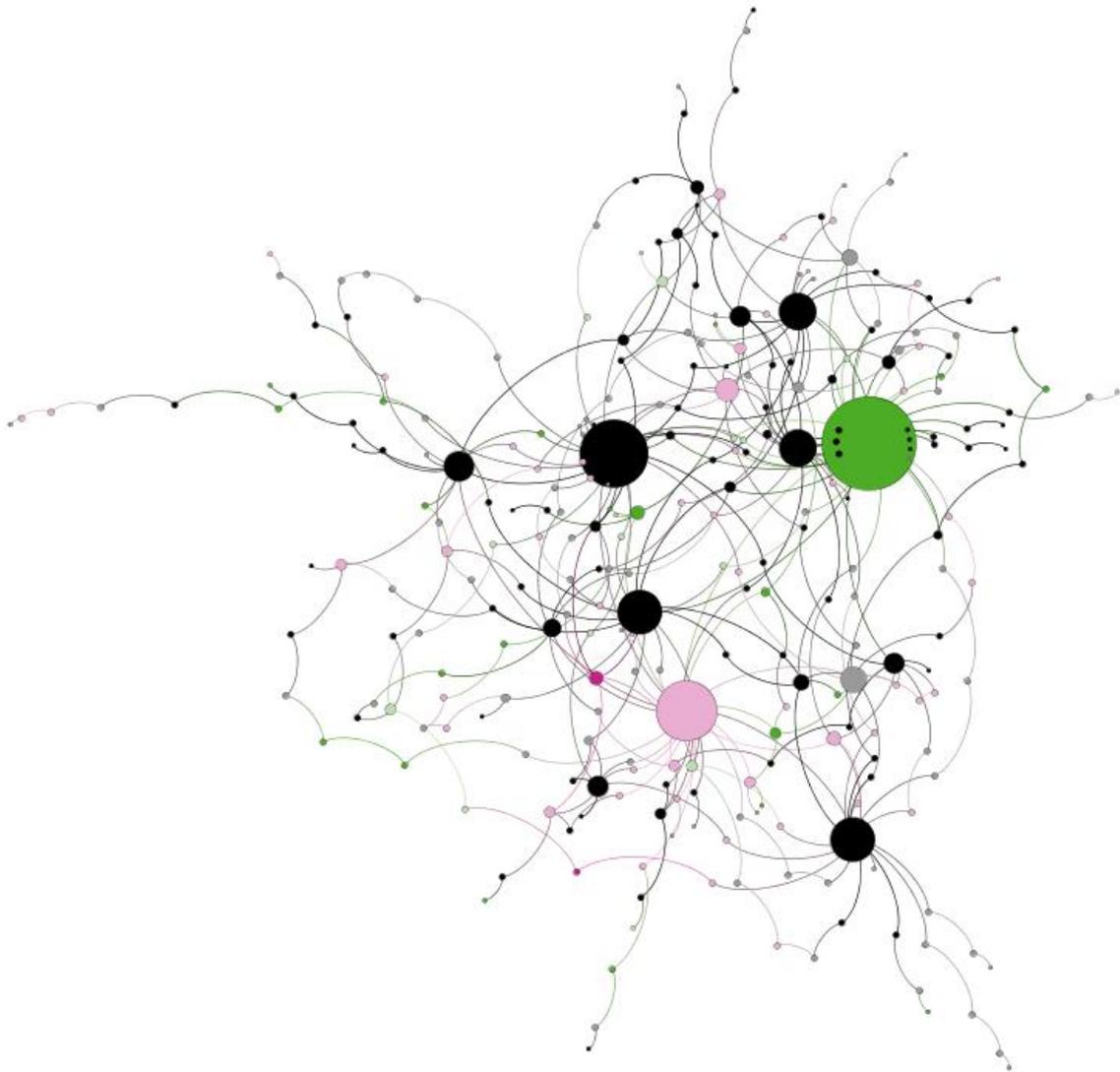


Figura 5.15. La prima rete del termine "dangerous" estratta da BNC, e visualizzata con il software Gephi.

5.4. Il vocabolario delle emozioni

Di seguito riportiamo il vocabolario delle parole emozionali per indagare la creazione di reti semantiche in WordNet. Sono state scelte le parole emozionali, perché hanno un grande valore dal punto di vista evolutivo (Darwin, 1871). Tutta l'analisi è in lingua Inglese. La motivazione di questa scelta risiede nel fatto che il database WordNet (rispetto agli altri database esistenti, generati da questo originario progetto per moltissime altre lingue, compresa la lingua Italiana) è il più vasto ed esteso contenendo circa 140.000 termini e quasi un milione di collegamenti fra termini. I termini emozionali sono stati scelti in base al modello di Plutchkin (1980), che classifica le emozioni in base agli stimoli che le determinano, la comprensione cognitiva della situazione da parte del soggetto di quello che sta avvenendo, lo stato emozionale in cui si trova, il comportamento che viene scatenato dal punto di vista motorio, e gli effetti che tali emozioni provocano sul corpo.

Stimulus event	Cognition	Feeling state	Behavior	Effect
Threat	"Danger"	Fear	Escape	Safety
Obstacle	"Enemy"	Anger	Attack	Destroy obstacle
Gain of valued object	"Possess"	Joy	Retain or repeat	Gain Resources
Loss of valued object	"Abandonment"	Sadness	Cry	Reattach to lost object
Member of one's group	"Friend"	Acceptance	Groom	Mutual support
Unpalatable object	"Poison"	Disgust	Vomit	Eject poison
New territory	"Examine"	Expectation	Map	Knowledge of the territory
Unexpected event	"What is it?"	Surprise	stop	Gain time to orient

Tabella 1. le otto emozioni primarie di Plutchik (1980), "paura", "sorpresa", "tristezza", "disgusto", "rabbia", "aspettativa", "gioia", "accettazione", insieme con gli eventi stimolanti, la valutazione cognitiva, lo stato emozionale, il comportamento, gli effetti che gli stimoli e le emozioni determinano nel corpo.

Per cui le teorie psicologiche considerano le emozioni un processo a molte componenti, considerando diversi fattori che lo costituiscono: cognitivo, fisiologico, biologico, sociale, comportamentale e culturale.

Esiste poi una interpretazione più articolata, sempre realizzata da Plutchkin (2003), che prevede lo schema riportato in Tabella 2. In questa Tabella, le emozioni primarie, sono articolate lungo la dimensione dell'intensità, passando da meno intense a più intense.

Less Intense	Primary Emotion	More Intense
Interest	Expectation	Attention
Serenity	Joy	Ecstasy
Approval	Acceptance	Admiration
Apprehension	Fear	Terror
Distraction	Surprise	Amazing
Thoughtfulness	Sadness	Pain
Nay	Disgust	Disgust
Discomfort	Anger	Ire

Tabella 2. Articolazione delle emozioni da Plutchik (2003), lungo la dimensione dell'intensità, con una serie di termini collegati all'emozioni di base che si articolano lungo la dimensione della minore o maggiore intensità.

Lo stesso autore parla di composizionalità delle emozioni, un processo che permette di unire due emozioni di base, per dar luogo a nuove dimensioni dell'emozioni messe insieme. Nella Tabella 3, sono riportati una serie di esempi.

Basic Emotion	Union	New emotion
Expectation	Joy	Optimism
Expectation	Anger	Aggression
Joy	Acceptance	Trust
Acceptance	Fear	Submission
Fear	Surprise	Awe
Surprise	Sadness	Disappointment
Sadness	Disgust	Remorse
Disgust	Anger	Contempt

Tabella 3. Le emozioni hanno le stesse regole di composizionalità delle parole?

Lazarus (Lazarus, 1991a, b) introduce il concetto di valutazione cognitiva dell'emozione (Tabella 4). Secondo questo autore, è l'interpretazione dell'avvenimento piuttosto che l'avvenimento stesso l'elemento fondamentale nel determinare il tipo di risposta emozionale. In questa ottica si può spiegare come in diversi contesti culturali lo stesso evento può far scaturire reazioni differenti.

Valutazione dello stimolo	Emozione
Sei stato offeso e umiliato	Rabbia
Ti senti minacciato	Ansia
Hai vissuto una perdita	Tristezza
Hai infranto una regola morale	Senso di colpa
Desideri qualcosa che appartiene ad un'altra persona	Invidia
Sei vicino a qualcosa di repellente	Disgusto
Stai per raggiungere un obiettivo agognato	Felicità
Hai ottenuto un risultato ambito	Orgoglio
Sei stato trattato bene da un'altra persona	Gratitudine
Desideri affetto da un'altra persona	Amore
Sei commosso dalla sofferenza di qualcuno	Compassione

Tabella 4. Valutazione dell'emozione secondo R. Lazarus (1991a, b).

Ortony e collaboratori (1988), propongono un modello con 24 tipi di emozioni, aggiungendo una classe sovraordinata che comprende la tipologia "emozione", vista in generale. una classe per la mancanza di emozioni, e sette classi di termini in relazione alle emozioni legate a processi mentali, risultanti in un elenco di 33 distinzioni tassonomiche. Nell'analisi delle parole emozionali, sono stati individuati una serie di Motivi strutturali e funzionali (come in Milo et al., 2002). Definiamo come strutturali i motivi, considerati al di fuori del loro significato. In questo caso, rileviamo la presenza di pattern ben organizzati che richiamano i solidi stellati di Platone e altre figure geometriche multidimensionali. Ad ogni modo l'analisi che è stata fatta realizza prima reti strutturali, dentro le quali sono stati individuati sotto-reti con motivi funzionali. Queste strutture sono state chiamate *particelle emergenti*. Si veda il Capitolo 8, per una rassegna esaustiva dei risultati.

5.5. Crescita, simmetrie e particelle emergenti in una rete semantica

Nella sezione precedente sono state indagate la creazione e le caratteristiche di una rete a partire da un lessico dato. In questa sezione si riporta come indagare i processi di crescita di una rete semantica. Per introdurre la crescita di significati, o la produzione linguistica, sono utilizzati i concetti desunti dagli studi di Psicologia cognitiva, che accenniamo di seguito brevemente. La *rete semantica* (descritta precedentemente, paragrafo 5.2) è una forma di rappresentazione della conoscenza. La modellazione di tale rete come un grafo permette di individuare le strutture complesse di una lingua. Tali strutture sono composte da nodi (dove ad ogni nodo si associa un'entità concettuale) e da archi, che servono per collegare i nodi e costituiscono le relazioni tra le diverse entità concettuali (Eysenck & Keane, 2006). Le strutture di

memoria e di conoscenza, elaborate per la stiva dei contenuti semantici di una lingua, generalmente sono anche indicate con *schema*, *frame* o *script* o *sistemi di produzione*. Lo *schema* o *script* viene utilizzato per organizzare una conoscenza complessa. È costituito da un insieme di concetti basato sulla conoscenza generica e può essere utilizzato per rappresentare eventi, sequenze di eventi, ecc. Schank & Abelson (1977) e Schank (1991) hanno proposto il termine “script” come un sistema di memorizzazione che comporta una sequenza di eventi, (per esempio nello script “andare al ristorante”, la sequenza di eventi riguarda: entrare-ordinare-mangiare-pagare-andare via); una sequenza di comportamenti anche motori collegati ad ogni azione; una sequenza di dialoghi, appropriati al contesto delle varie azioni, una sequenza emozionale.

Uno *schema* o un *frame* (o *cornice*) è una specie di micro-universo semantico, all’interno del quale le parole sono comprensibili sullo sfondo di relazioni fra tutti i termini che compongono il frame stesso. Questo termine, introdotto da Minsky (1975) indica la traccia che l’esperienza passata, accumulata nella memoria di ciascuno, lascia nella conoscenza e che permette di collegare le frasi tra di loro, riempire le lacune di informazione e dare senso alle informazioni. Ad esempio, si comprende il significato di “lunedì” facendo riferimento al frame dei giorni della settimana. I *sistemi di produzione* infine, usano regole dalla forma SE...ALLORA. In un sistema di produzione è contenuta una memoria a lungo termine, con regole dalla forma “se...allora”, e una memoria di lavoro che comprende tutte le informazioni. Il sistema di produzione opera confrontando il contenuto della memoria di lavoro e la parte SE della regola, eseguendo poi la forma ALLORA (per esempio SE è si accende la luce del semaforo, ALLORA è possibile attraversare la strada) (Eysenck & Keane, 2006).

Come si fa a produrre significati concatenati nel parlato? Qual è il sistema di produzione che di solito utilizziamo. Nella letteratura sulla teoria delle reti (vedi Capitolo 2), a partire da Eulero, sono stati individuati vari modelli di creazione o crescita di grafi (Newman et al., 2001; Newman, 2003; Newman, 2010). In particolare i più famosi sono: il modello di grafo casuale di Erdős (RG) (Erdős & Rényi, 1959, 1960), il modello di reti piccolo mondo (Small-World Networks o SWN) di Watts & Strogatz (1998) e il modello Scale-Free (SF che segue una legge di potenza) di Barabási & Albert (1999), con la sua crescita ad invarianza di scala e l’attaccamento preferenziale dei nodi.

Tutti questi modelli, pur intercettando molto bene aspetti di sistemi reali, sono sistemi nei quali ogni nodo è equivalente a ciascun altro nodo. Solo nel modello di Barabási & Albert, si utilizza il grado del nodo come attrattività per creare nuovi archi, connessi al nodo di partenza con alto grado. Nei modelli esistenti nella letteratura sulle reti, ad ogni nodo non è assegnata alcuna proprietà. L’originalità dell’approccio adottato in questo capitolo risiede invece nell’aver attributi per ogni nodo. Per cui, i nodi delle reti linguistiche non omogenei hanno diverse caratteristiche semantiche, che sono i diversi significati o significati leggermente diversi. Tali proprietà considerano caratteristiche si può dire *intrinseche* o *connaturate* del nodo, indipendentemente dal grado. In particolare, se il termine di un nodo è un nome, un verbo, ecc., è possibile

che le dinamiche dei suoi intorno e la crescita delle reti manifesti proprietà di crescita imprevedibili, rispetto alla crescita preferenziale del modello Scale Free e/o del modello di crescita casuale.

5.6. Reti con nodi non omogenei e particelle emergenti

Esiste un modello di crescita di una rete semantica che tenga conto degli attributi di ogni nodo? Si ipotizza che, dato un insieme $W = \{w_1, w_2, \dots, w_N\}$ di termini, esista una proprietà $P(w_i)$ che fornisce la capacità del termine di generare legami; cioè $k_i = P(w_i)$. Questo termine allora avrà un intorno G_i , la cui cardinalità (il numero di elementi che appartengono all'insieme) sarà pari a $k_i + 1$. Ad esempio, il termine "dangerous" di WordNet, presenta una rete indiretta di 9 nodi e 14 archi (9 nodi e 18 archi come rete diretta). In Figura 5.2a, è riportata la rete indiretta, visualizzata con Gephi ed il motivo della rete stessa. In questo caso $k_{dangerous} = 8$.

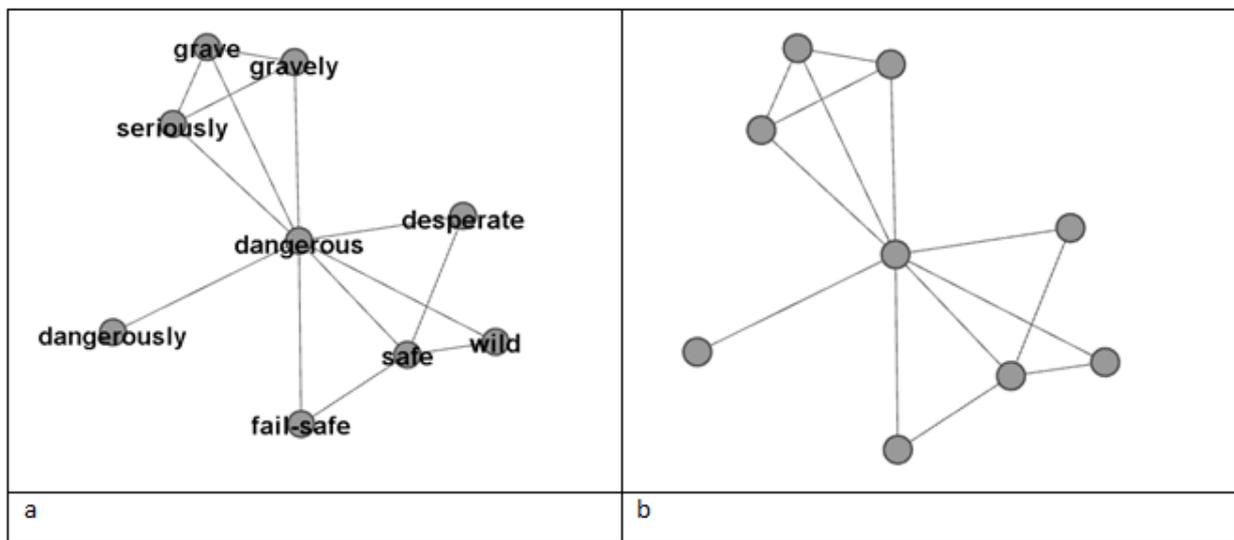


Figura 5.2. (a) L'intorno 1 del termine "dangerous". (b) il motivo corrispondente.

E' interessante notare che i termini "dangerous", "grave", "gravely" e "seriously" formano un grafo completo.

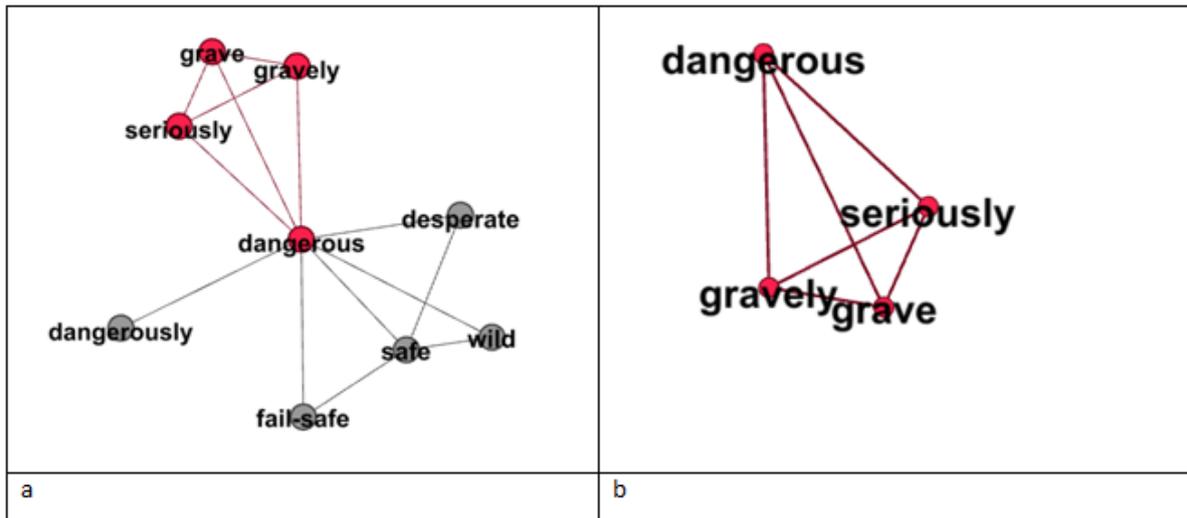


Figura 5.3. (a) I termini “dangerous”, “grave”, “gravely” e “seriously” sono indicati in rosso in $G_{dangerous}^1$; una volta estratti formano un grafo completo (b).

Si indicano col nome *particelle emergenti* raggruppamenti di termini che formano grafi completi. Tale definizione non è esaustiva della molteplicità di manifestazioni che sono state rilevate. Di seguito si riporta qualche esempio e si rileva (dai dati empirici analizzati nel Capitolo 8) una prima non esaustiva classificazione delle particelle emergenti, intorno a due concetti fondamentali:

- a. persistenza, deformazione, evoluzione e, in generale modifica della struttura geometrica del grafo perfetto;
- b. modello di crescita generativo, che procede con modalità completamente diverse da quelle indicate in letteratura (Reti Random, Small World o reti a Invarianza di Scala, con Crescita Preferenziale), realizzando alle volte singoli collegamenti ad un unico nodo (livello locale della crescita), alle volte collegando strutture ad altre di strutture (livello multidimensionale della crescita).

Se si considera la rete diretta di $G_{dangerous}^1$, si osserva che esistono delle sostanziali differenze tra le due reti (Figura 5.4) nella quantità di legami che i due grafi realizzano.

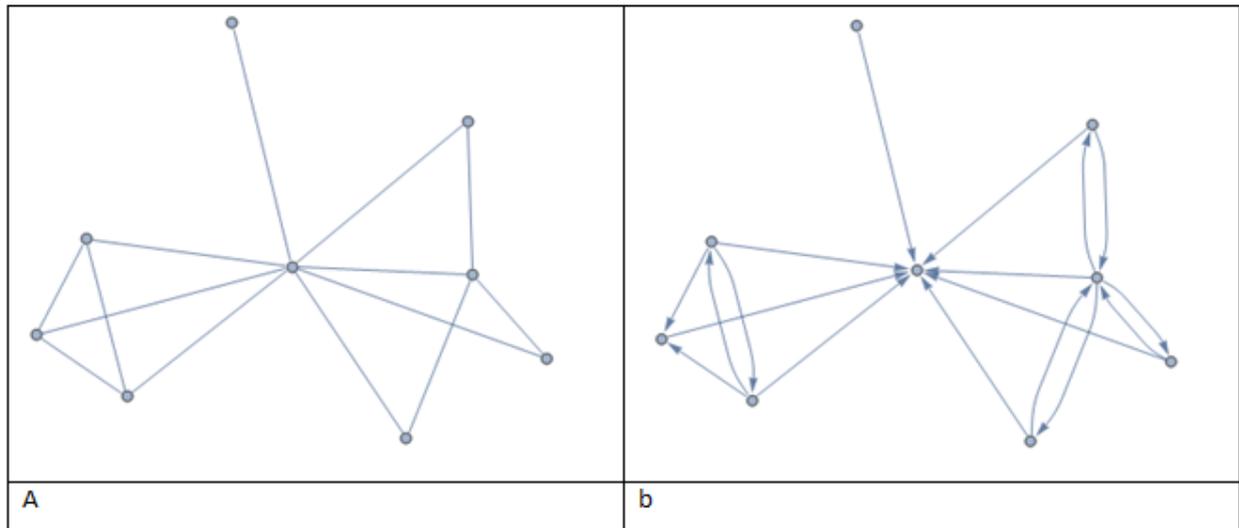


Figura 5.4. Differenze tra reti indirette (a) e reti dirette (b).

Il passaggio da grafo indiretto a grafo diretto comporta sostanziali differenze nei motivi organizzativi della rete semantica, come si nota dalla Figura 5.5.

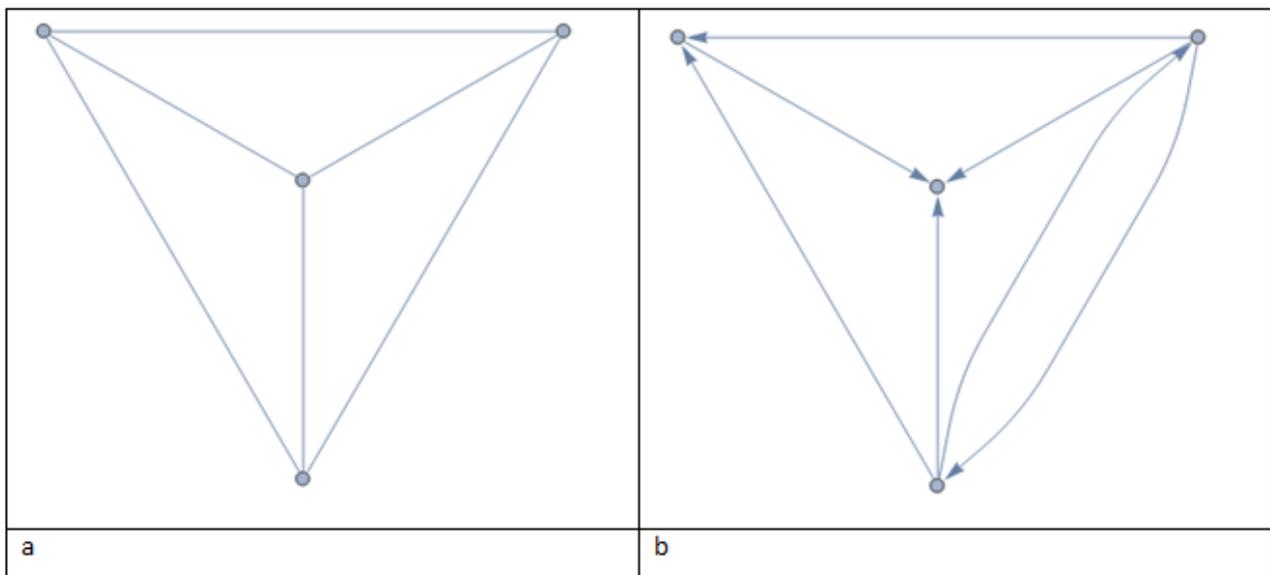


Figura 5.5. La differenze tra reti indirette e reti dirette si traduce in una differenza tra particelle emergenti (a) (b).

Il fatto che le reti possano essere dirette, cioè per qualche i e j può accadere che $a_{ij} \neq a_{ji}$ è che le relazioni $R_1 \dots R_8$ non sono invertibili (tranne alcune, per esempio le relazioni antonimiche). Altre differenze sono dovute al modo in cui i termini e le loro relazioni sono organizzati nel lessico e nella semantica. Se si considera il grafo completo di Figura 5.5a come grafo diretto (vedi Figura 5.6a), si osserva che possiede le stesse simmetrie del quadrato, cioè è invariante per rotazione di 90° , 180° e 270° , è invariante per riflessione, ecc. Alcune di queste simmetrie non sono rilevabili se si considera la rete diretta. In particolare,

questo accade per le simmetrie di rotazione, mentre solo una simmetria di riflessione (quella rispetto all'asse orizzontale, passante per il centro del quadrato) è mantenuta (Figura 5.6c). In un certo senso, la Figura 5.5b risulta da una rottura di alcune delle simmetrie della Figura 5.5a o 5.6b.

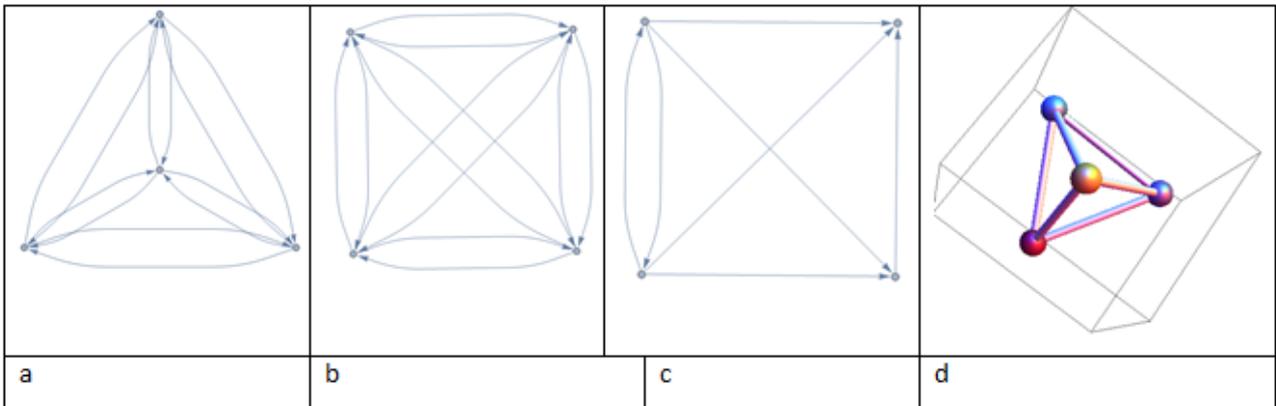


Figura 5.6. (a) il grafo completo della particella emergente composta da 4 termini e le sue differenti manifestazioni attraverso i differenti legami che sono realizzati fra termini (a, b, c). In 3 dimensioni la sua struttura è quella di un tetraedro (d). In due dimensioni il nodo centrale non sembra avere le stesse caratteristiche dei nodi periferici. In realtà questo effetto è dovuto solo alla visualizzazione in quanto le caratteristiche dei nodi del grafo dipendono solo ai legami che intrattengono con gli altri nodi.

Si consideri il grafo della Figura 5.2. Questa rete è composta da un nodo padre P ("dangerous") e 8 nodi figli (F_1, \dots, F_8). I nodi sono del tipo (P, F_i) . Il rank del nodo P è k_P . L'approccio usato è per reti indirette, ma è estendibile anche alle reti dirette. Se si analizza il secondo livello di crescita, la rete $G_{dangerous}^2$ si presenta molto più complessa della precedente, avendo 224 nodi e 563 archi (Figura 5.7). In tale rete, esistono 6 archi tra coppie di figli (F_i, F_j) .

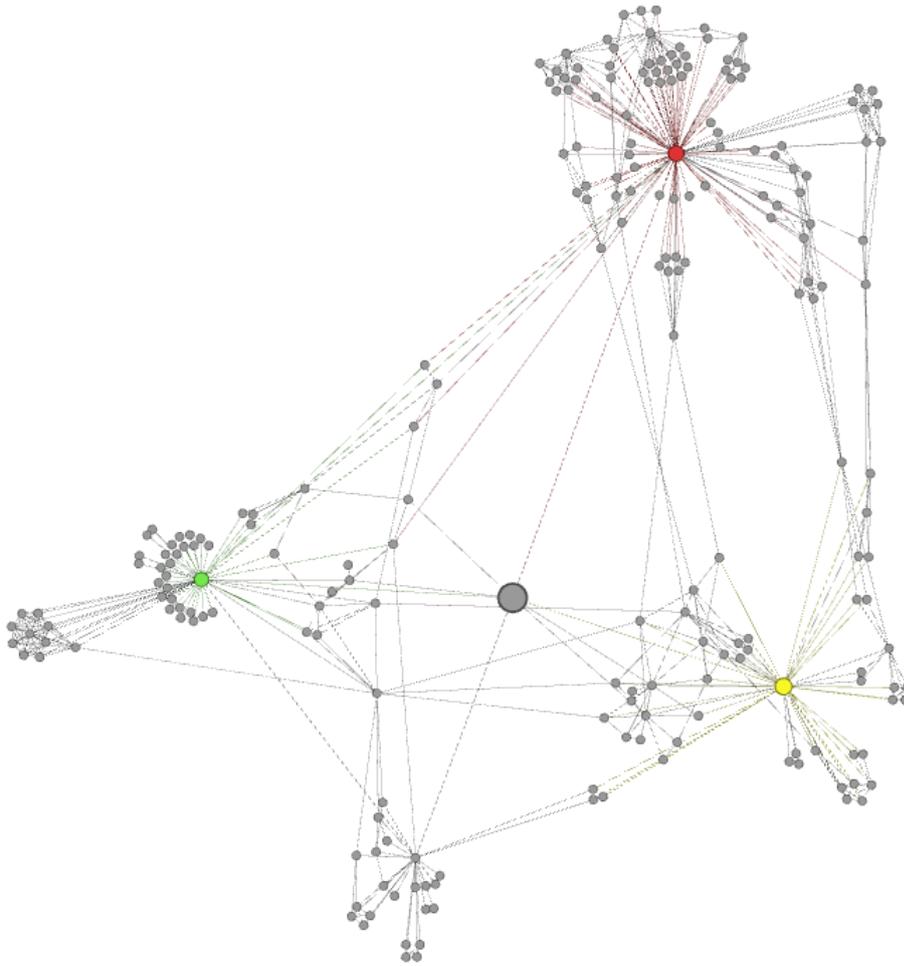


Figura 5.7. Il secondo intorno della rete del termine "dangerous". Il termine è rappresentato dal cerchio più grande. I termini "wild" (in rosso), "saf" (in verde) e "grave" (in giallo) sono perfettamente visibili.

Sebbene il termine "dangerous" continui ad essere il più centrale, non è più quello con il maggior grado (Tabella 5.3).

Termine	Grado
wild	99
safe	53
grave	45
tame	27
desperate	27
savage	17
gravely	14
wilderness	12
good	12
wilderness	11
rubber	11
seriously	11
safety	10

contraceptive	10
prophylactic_device	10
preventive	10
condom	10
prophylactic	10
contraceptive_device	10
birth_control_device	10
preventative	10
tomb	10
gentle	9
natural_state	9
state_of_nature	9
crazy	9
barbarian	9
idle	9
secure	9
heavy	9

Tabella 5.3. Parole con grado maggiore, nel secondo livello di crescita della rete che parte dal termine "dangerous".

Nella Tabella 5.3, sono presentati 30 termini , il cui grado è maggiore di quello del termine "dangerous", considerato il termine *padre* di tale rete. In particolare "wild", rappresentato in rosso nella Figura 5.7, ha ben 99 archi, "safe" (rappresentato in verde), ha 53 archi, e "grave" (rappresentato in giallo), ne possiede 43. La natura della rete è completamente cambiata, in quanto la variabile di Betweenness Centrality è massima per "wild" (vale 15.158), è alta per "safe" (9.707), ma è molto più bassa per "dangerous" (5.670) . Se si indicano con (N_1, \dots, N_{215}) i nodi che distano 2 passi dal termine "dangerous", sono detti nodi *nipote*. Gli archi sono del tipo (N_i, N_j) . Ad esempio, nel grafo di Figura 5.8, ci sono nodi nipoti, nodi figli dello stesso figlio *F*, che realizzano grafi completi di 9 particelle. Ogni nodo ha 8 collegamenti.

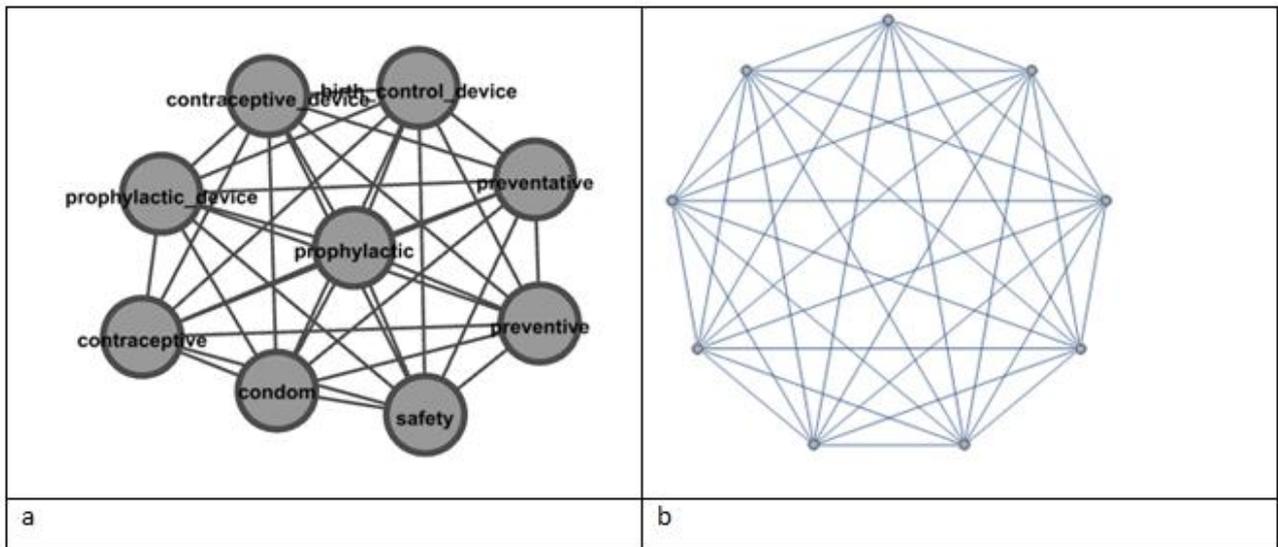


Figura 5.7. Figli dello stesso figlio che realizzano legami tra loro, creando una particella emergente, composta da un grafo completo di 9 nodi (con e senza etichette a e b rispettivamente).

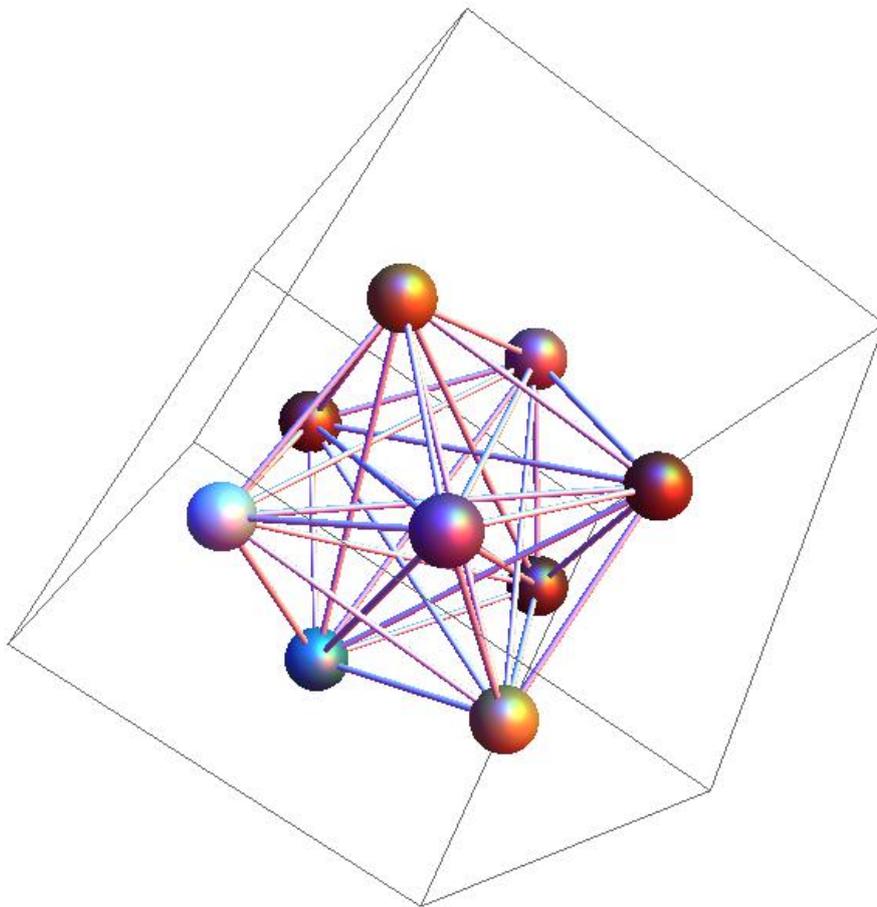


Figura 5.8. Rappresentazione 3d della particella emergente composta da un grafo diretto di nove nodi. Le particelle trovate seguono differenti modelli organizzativi.

Per cui, in un ipotetico *modello di crescita* che tenga conto di un attributo dei nodi, si ha un nodo di partenza P , x nodi *figlio* F , y nodi *nipote* N_F . Tra i nodi *figlio* si realizzano legami del tipo (F,F) ; tra i nodi nipote, si realizzano altri tipi legami (N_F, N_F) , così come altri legami si possono stabilire tra *figli* dello stesso *figlio* o tra *figli di figli* differenti. Ulteriori legami si possono creare tra nipoti che realizzano più legami con figli differenti. Tutte queste forme sono presenti nella rete $G_{dangerous}^2$. Questo processo gerarchico può essere ulteriormente codificato nella rete $G_{dangerous}^3$, che ora possiede 4881 nodi e 27077 archi. Definendo questa serie di probabilità è possibile introdurre un nuovo modello di crescita.

In Figura 5.8, è riportata una visualizzazione 3d della particella composta da un *grafo completo* di nove nodi, rilevata al terzo step di evoluzione della rete, derivata dal termine “*dangerous*”.

Seguendo uno schema definito nei lavori di Bilotta & Pantano (Bilotta et al., 2011; Bilotta & Pantano, 2011a, b, c) per autoriproduttori in Automi Cellulari, le Particelle Emergenti (PE) sono state classificate in 4 categorie principali: *grafi perfetti*, indicati con la sigla PE_di tipo_A, *grafi imperfetti a causa di rotture di simmetrie*, indicati con la sigla PE_di tipo_B, *grafi con simmetrie complesse*, indicati con la sigla PE_di tipo_C e *particelle complesse, risultanti da particelle più elementari*, indicati con PE_di tipo_D.

Nel Capitolo 8, che riporta un ampio report sperimentale, una rassegna di particelle emergenti sono individuate, utilizzando le 4 categorie di classificazione, risultante dall’investigazione di WordNet. Per cui si avrà, riprendendo l’esempio di questo paragrafo, *dangerousPE_di tipo_A_1*, oppure *dangerousPE_di tipo_D_6*, ecc.

5.7. I sistemi software sviluppati

Sono stati realizzati sistemi automatici che incorporano le funzioni che sono servite per modellare una lingua naturale. Il primo sistema si chiama SWAP, già utilizzato per lo sviluppo di reti di piccolo mondo da database musicali (Campolongo & Vena, 2005, 2006). Dopo aver scaricato il database WordNet, sono stati implementati tutti gli algoritmi di ricerca, sopra menzionati: R_1 : Antonimia, R_2 : Iperonimia, R_3 : Iponimie, R_4 : Implicazione, R_5 : Similarità, R_6 : Sinonimie, R_7 : Related e R_8 : Overview.

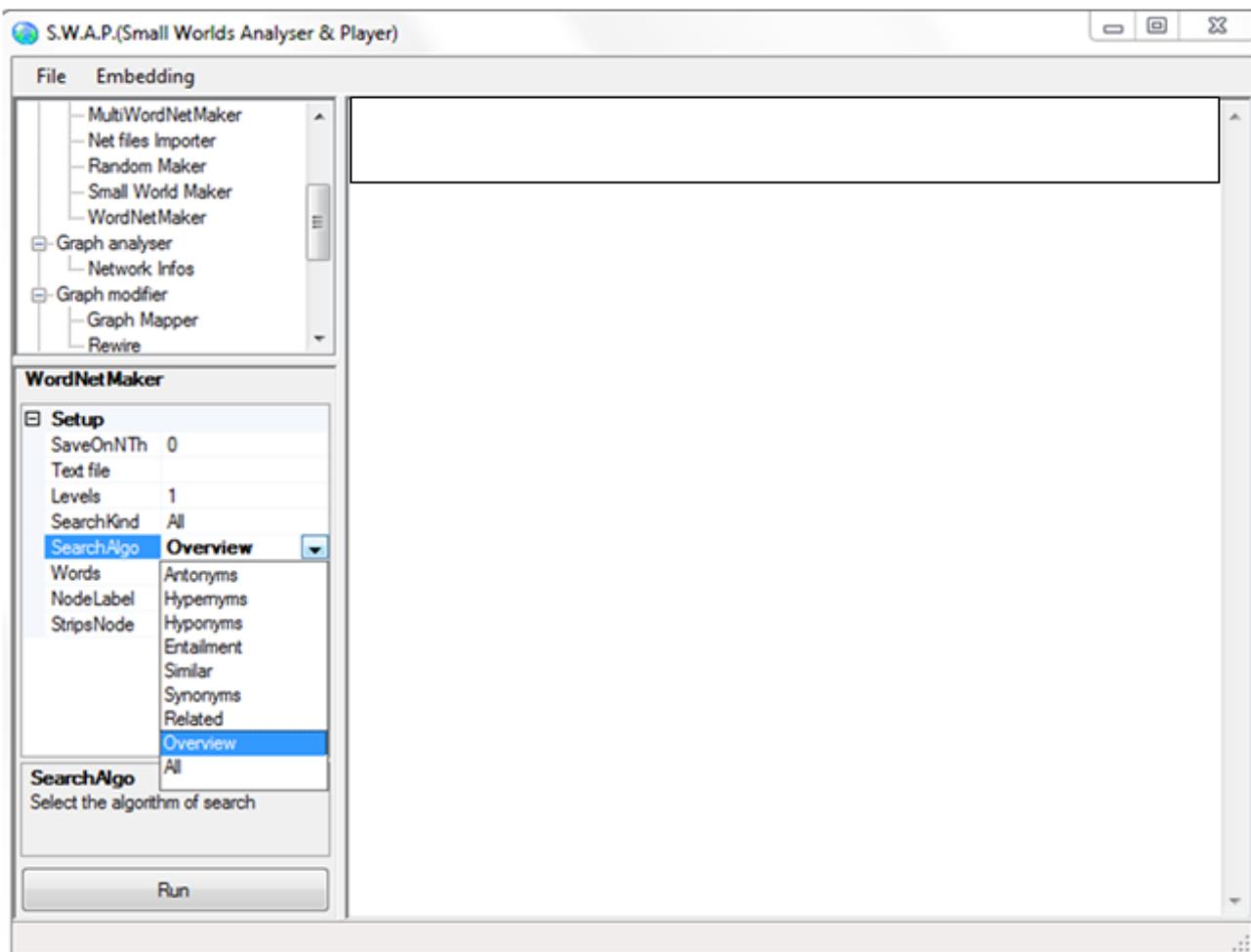


Figura 5.9. Interfaccia di SWAP. Il sistema è posizionato sulla funzione di WORDNET MAKER. L'immagine mostra tutti gli algoritmi di ricerca che sono stati sviluppati. Il tipo di ricerca All rappresenta che il sistema cerca un termine considerando tutte le categorie grammaticali (nomi, aggettivi, verbi, avverbi).

Se si inserisce un termine emozionale, per esempio "love", è possibile organizzare le ricerche tali algoritmi, per indagare il sistema WordNet. Nell'immagine di Figura 5.10, si sceglie la funzione R_8 : Overview, e si inserisce il termine "love". Tale modalità di ricerca sarà utilizzata per la parte sperimentale del Capitolo 8.

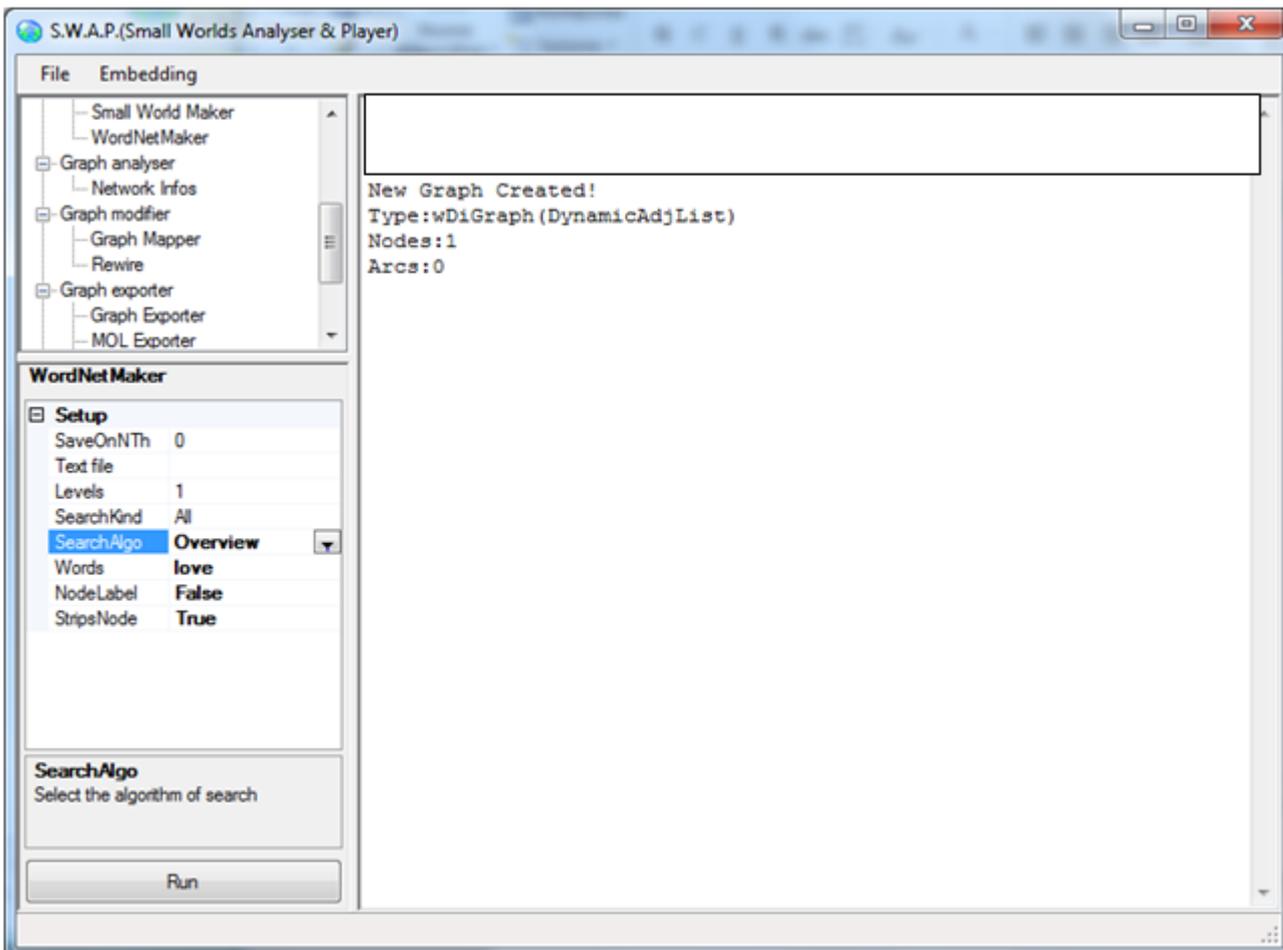


Figura 5.10. Interfaccia di SWAP. Il sistema è posizionato sulla funzione di WORDNET MAKER. Al primo livello di indagine la rete presenta, per il termine “love”, ricercato con l’algoritmo Overview 1, un solo termine, considerando tutte le categorie grammaticali.

Inoltre, il livello di ricerca può essere variato, per il tipo di algoritmo, per il tipo di profondità di ricerca, e per il tipo di elemento grammaticale che si intende analizzare.

In Figura 5.11, è visualizzata un’altra immagine dell’interfaccia per il livello 5, sempre per l’algoritmo di ricerca Overview. Come si può notare, cambiando il livello di analisi, le reti crescono a dismisura.

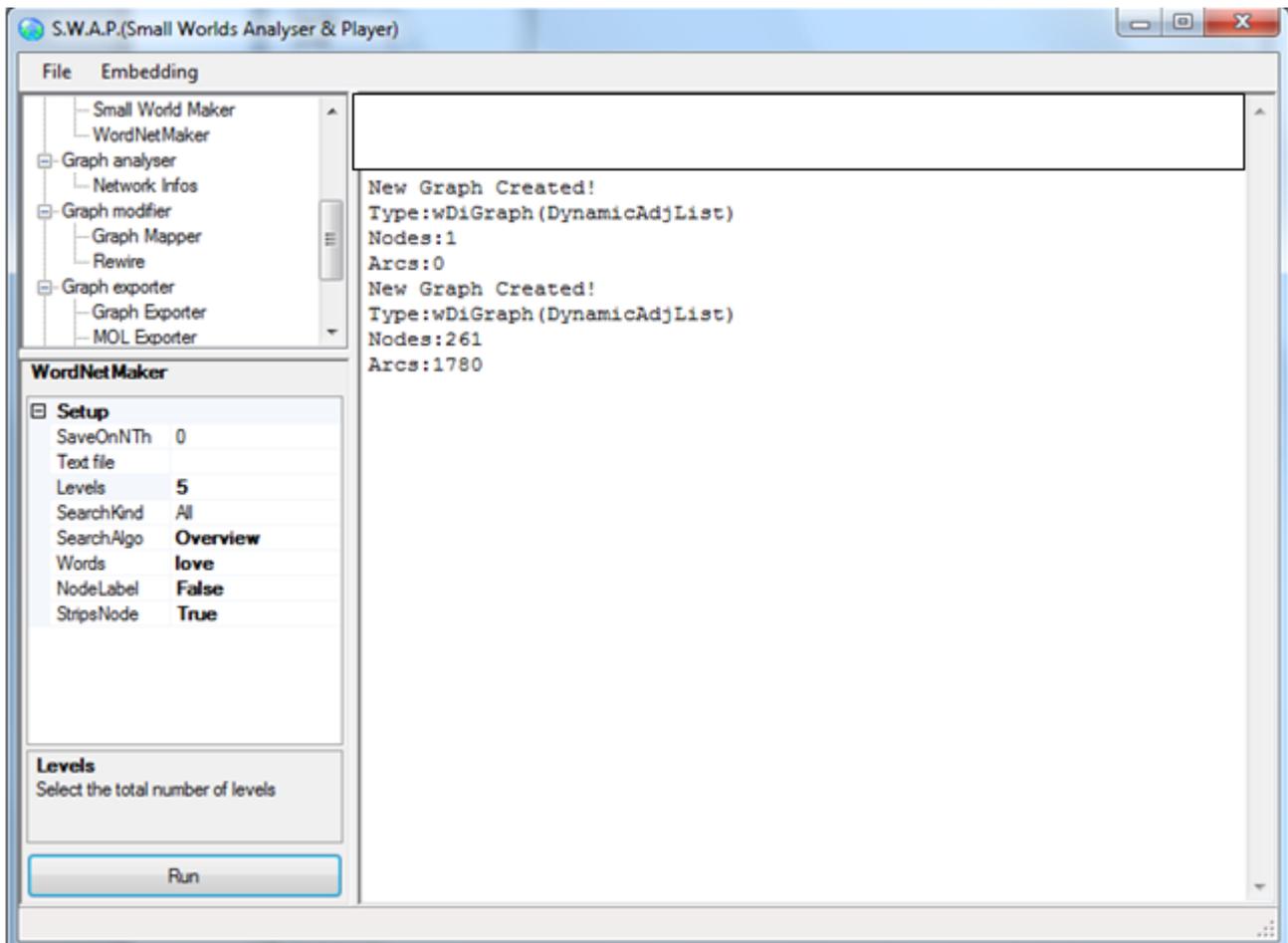
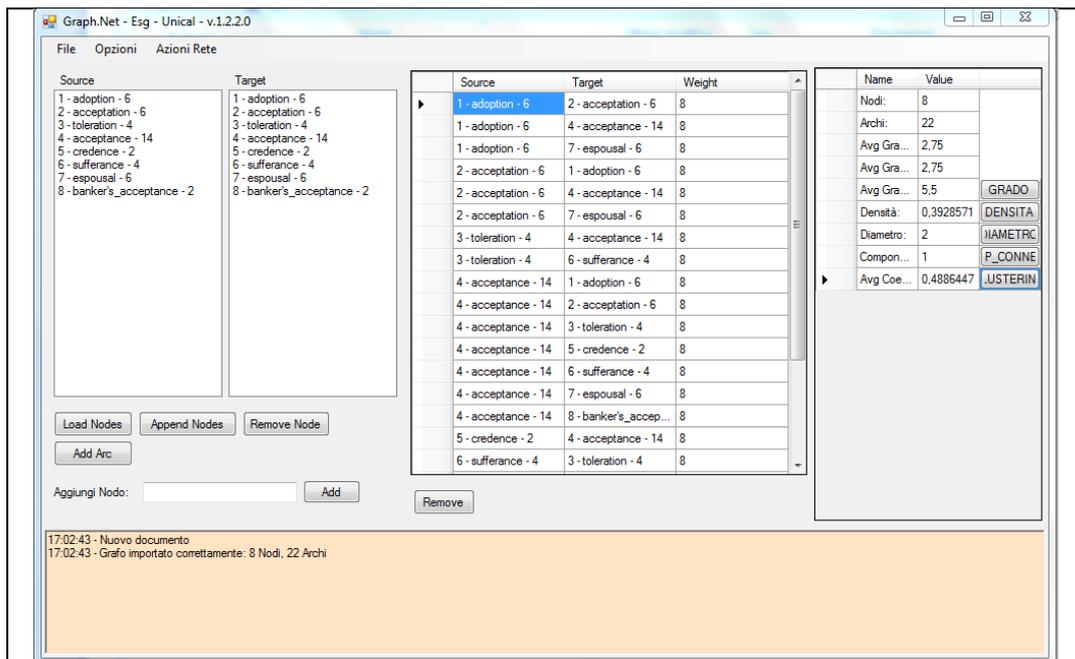


Figura 5.11. Nuova ricerca nel sistema WordNet. Al quinto livello di indagine, la rete per il termine emozionale “love” è diventata di 261 nodi e 1780 collegamenti, considerando tutte le categorie grammaticali.

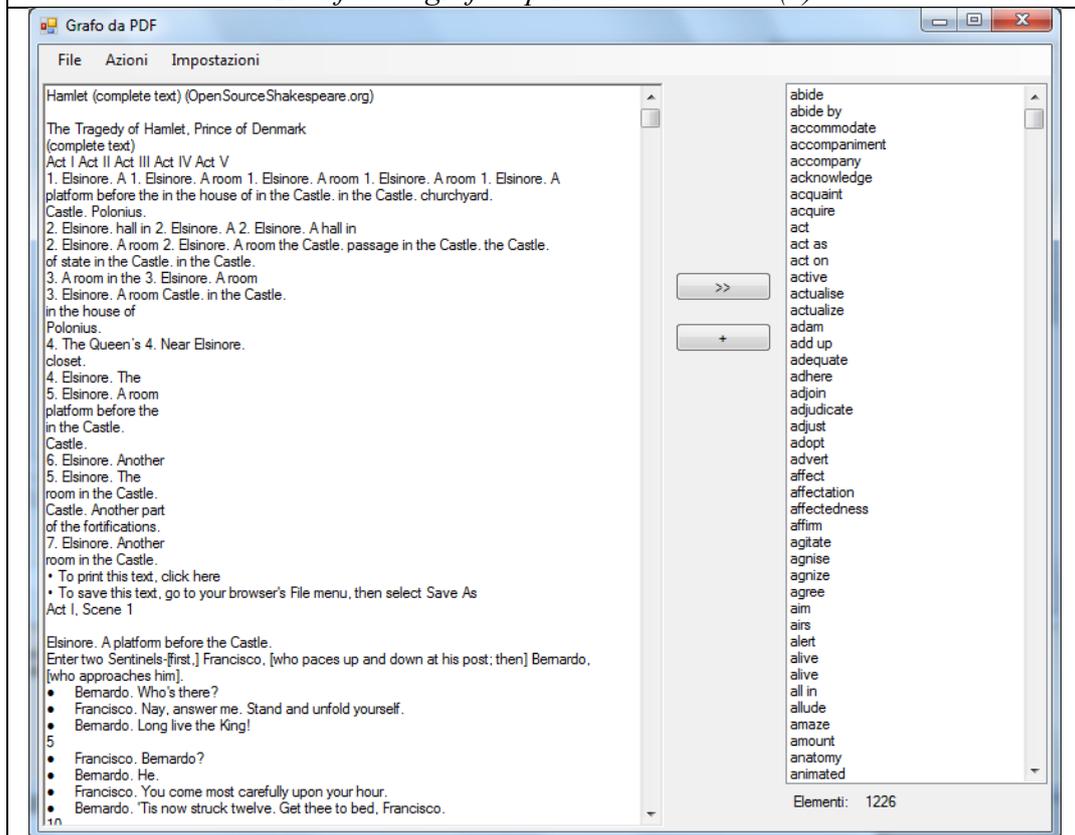
Si passa da un singolo nodo al livello 1, a 261 nodi e 1780 archi del livello 5. Questo meccanismo di crescita sarà oggetto di analisi nel Capitolo 8. La crescita crea le particelle emergenti.

Altri software sviluppati sono *Semantic Network Creator*, di cui si è parlato prima, che serve a creare reti semantiche con l’ausilio di un vocabolario e la possibilità di inserire diversi tipi di attributi, e *GraphNet*, che è un software completo per creare, manipolare, analizzare e realizzare grafi partendo da testi scritti (come file di tipo .txt o .pdf).

La Figura 2.12, riporta una visualizzazione dell’interfaccia sviluppata.



Interfaccia grafica per l'analisi di reti (a)



Interfaccia per creare reti da un vocabolario, usando un file pdf (b), in questo esempio l'Hamlet di Shakespeare.

Figura 2.12. Interfaccia del software GraphNet. Il software serve a analizzare reti e a manipolarle in vario modo(a), e serve anche, importando una lista di termini che funge da vocabolario a creare una rete da un file di testo.

Capitolo 6. Reti musicali in Bach

6.1. Introduzione

In questo capitolo, si indaga in modo esaustivo le melodie di Bach, attraverso la misura dell'entropia, presentata nel Capitolo 3. Le melodie sono successivamente rappresentate tramite le reti, che sono investigate dal punto di vista strutturale al livello della singola traccia, della collezione, dell'intera melodia. I risultati, a diversi livelli di complessità, fanno emergere una straordinaria rete di relazioni, con aggregazione tipiche dei modelli di reti di piccolo mondo (Strogatz, 2001) e delle reti con caratteristiche scale-free (Albert et al., 1999; Albert & Barabási, 2002, Barabási & Albert, 1999), anche se con la presenza di alcune eccezioni. Infine si osserva come in queste reti emergono strutture organizzate come *clique* e strutture ancora più complesse come insieme di *clique*. La musica come le lingue naturali condivide processi organizzativi di base della stessa natura gerarchica e produttiva. Le grammatiche emergenti rappresentano questa base comune. Si rilevano inoltre dinamiche di organizzazione di tipo diverso che possono essere riferite ai cicli e i modelli di crescita musicale.

6.2. L'entropia nelle melodie di Bach

Esistono numerose versioni Midi delle melodie di Bach. La più completa si trova all'indirizzo web <http://www.bachcentral.com/BachCentral/index.html> dove un repertorio di 241 melodie può essere scaricato liberamente. Ad ogni melodia possono corrispondere più tracce, per un totale complessivo di 698 tracce. Queste melodie raggruppate in 13 Collezioni, Arte della Fuga (AOF), Concerti di Brandeburgo (Brandeburgo Concertos, BC), Cantate (Cantatas, C), Suite per Cello (Cello suites, CS), Fughe (Fugues, F), L'Offerta Musicale, anche detta collezione delle Variazioni Goldberg (Musical Offering, MO), Invenzioni (Inventions, I), Lavori per Organo (Organ Works, OW), Corali per Organo (Organ Chorales, OC), Sonate e Partite (Sonatas and Partitas, SP), Sinfonie (Sinfonias, S), Clavicembalo Ben Temperato, Libri I & II (The Well Tempered Clavier, WTCBKI-WTCBKII), sono riportate con il loro nome, in Inglese e con una sigla che le contraddistingue (Bertacchini et al., sottomesso, a).

Categorie	Numero di Melodie	Numero di di tracce
Art Of Fugues (AOF)	18	62
Brandeburgo Concertos (BC)	9	79
Cantatas (C)	2	9
Cello suites (CS)	13	13
Fugues (F)	2	8
The Musical Offering (Goldberg Variations) (MO)	32	98
Inventions (I)	15	30
Organ Works (OW)	29	109
Organ Chorales (OC)	12	56
Sonatas & Partitas (SP)	12	54
Sinfonias (S)	15	45
The Well-Tempered Clavier Book I (WTC, BKI)	48	111
The Well-Tempered Clavier Book II (WTC, BKII)	24	24

Tabella 6.1. Composizioni selezionate del repertorio di Bach, numero di composizioni e tracce considerate.

Può essere interessante indagare la dimensione dello spazio musicale M in questo insieme di tracce di melodie di Bach. Per ogni traccia abbiamo contato il numero di note che compongono la melodia e la dimensione dello spazio M (numero di note differenti presenti in una melodia). Il risultato è evidenziato in Figura 6.1.

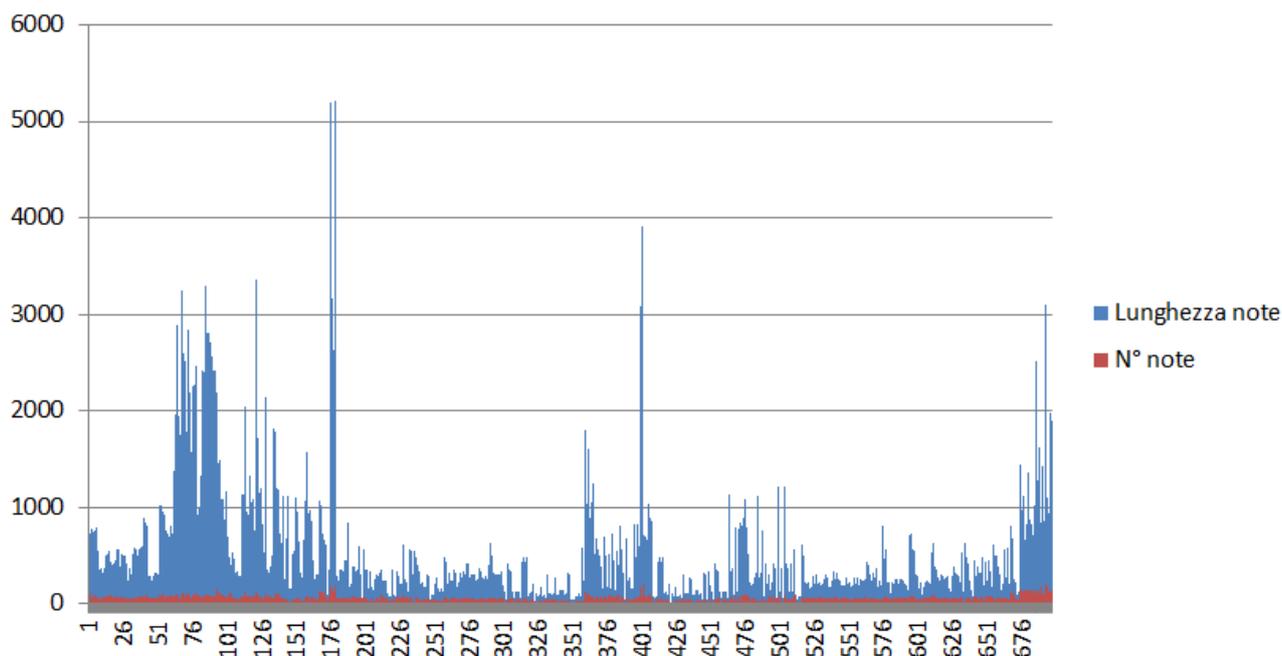


Figura 6.1. Numero di note presenti nella melodia (in azzurro) e dimensione dello spazio musicale (numero di note non ripetute) (in rosso).

La lunghezza delle varie tracce è molto variabile e oscilla da un minimo di 17 note (traccia 2, melodia catech3, della Collezione *Organ Works, OW*), ad un massimo di 5213, traccia 1, della Collezione *The Musical Offering (Goldberg Variations (MO))*. Anche lo spazio musicale varia enormemente, da un massimo di 210 (traccia 1, melodia toccata 2, della Collezione *Organ Works, OW*) ad un minimo di 10 note relativo alla traccia 2, melodia catech3, della Collezione *Organ Works, OW*. La traccia 1, della Collezione *The Musical Offering (Goldberg Variations (MO))*, che ha la lunghezza maggiore, ha però uno spazio musicale pari a 194.

La correlazione tra lunghezza delle melodie e la dimensione dello spazio musicale è intorno a 0,706421224, per cui esiste una buona correlazione tra le due variabili. Questa relazione implica che il numero di note differenti, utilizzate da Bach, è abbastanza correlato alla lunghezza della traccia melodica. Più la melodia diventa più lunga, più cresce il numero di note utilizzate, anche se sono state rilevate notevoli differenze tra un caso e l'altro, dato il coefficiente di correlazione trovato.

Se si considerano le entropie relative alle singole note, per le 698 tracce delle 241 melodie di Bach, si ottiene il grafico di Figura 6.2.

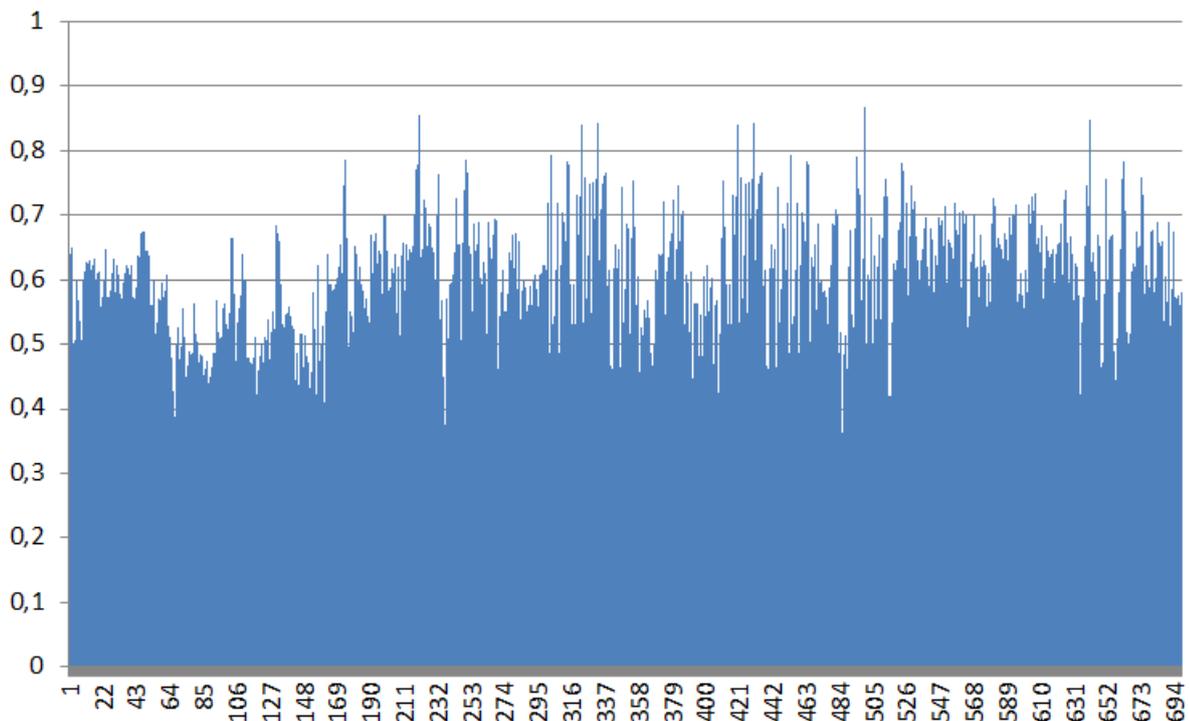


Figura 6.2. Entropie riferite alle 698 tracce delle 241 melodie di Bach.

E' possibile osservare che si passa da un valore minimo di 0,363280014, per la terza traccia della melodia *gig1* del raggruppamento *Partitas (P)*. ad un valore di 0,8681153. per la terza traccia della melodia *ron2*, sempre della stessa Collezione. Si noti che mentre la traccia con l'entropia più bassa possiede un elevato numero di note (1124), a fronte di uno spazio musicale di dimensione 26, quella più grande, al contrario, possiede un basso numero di note (46), con una dimensione dello spazio musicale pari a 31. Anche le tracce con i più alti valori di entropia hanno un comportamento analogo.

Si possono indagare le correlazioni tra il numero di note, la lunghezza melodica e l'entropia. Il valore relativo alla correlazione tra lunghezza della traccia e l'entropia è pari a -0,506588614, che indica che c'è una debole correlazione tra la lunghezza della traccia e l'entropia (nel senso che l'entropia diminuisce, in media, seppur debolmente, all'aumentare del numero di note). Nessuna correlazione sembra esistere tra la dimensione dello spazio musicale e l'entropia (il valore in questo caso è pari a -0,134). Inoltre, si può estendere l'uso del concetto di entropia, considerando non solo singole note, ma anche blocchi di più note, per vedere se ci sono ripetizioni. Il risultato va in questa direzione: per esempio, la quinta traccia della melodia *sin2* della Collezione *Partitas(P)* ha un valore minimo per l'entropia di coppie di note, pari 0,408. E' interessante notare che questa traccia ha un lunghezza complessiva di 75 note, e uno spazio musicale di dimensione 13. Questo non è un fenomeno generale, perché la terza traccia della melodia *var27c9*, della Collezione *The Musical Offering Goldberg Variations (MO)* ha un valore pari 0,517 per l'entropia di coppie di note, mentre la lunghezza della traccia è pari a 278 e la dimensione dello spazio musicale è pari a 13. Anche

l'entropia di terne di note si mantiene mediamente su valori bassi (0,589 e 0,640 rispettivamente) per queste composizioni.

In Tabella 6.2, si riporta il valore medio dell'entropia per ogni Collezione della musica di Bach.

Categorie	Entropia media
Art Of Fugues (AOF)	0,607694986
Brandeburgo Concertos (BC)	0,519005145
Cantatas (C)	0,478183605
Cello suites (CS)	0,641207049
Fugues (F)	0,654801707
The Musical Offering (Goldberg Variations) (MO)	0,632729174
Inventions (I)	0,591566791
Organ Works (OW)	0,615858495
Organ Chorales (OC)	0,650248698
Sonatas & Partitas (SP)	0,605463879
Sinfonias (S)	0,665884498
The Well-Tempered Clavier, Book I (WTC, BI)	0,633343333
The Well-Tempered Clavier, Book II (WTC, BII)	0,609553698

Tabella 6.2. Entropia media per ogni Collezione considerata.

La Tabella 6.2 indica che l'entropia per le Collezioni considerate risulta mediamente intorno a 0,6 ad eccezione delle categorie *Brandeburgo Concertos* e *Cantatas* dove vale rispettivamente 0,519 e 0,478.

6.3. Analisi delle reti

Come descritto nel Capitolo 3, le tracce dell'intera opera di Bach sono state trasformate in reti. Il grafico di Figura 6.3. In questa figura, si indica ogni singola rete con un simbolo, nel piano Oxy (piano nodi-archi).

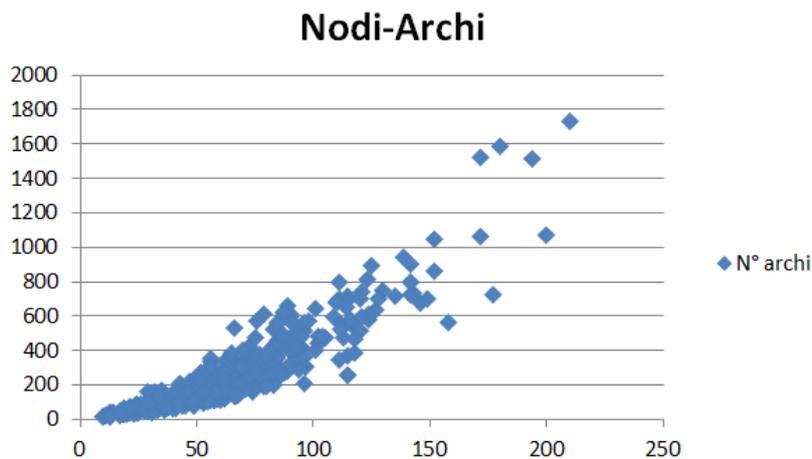


Figura 6.3. Distribuzione del numero di archi vs. il numero di nodi.

Il grafico manifesta un alto grado di correlazione tra il numero di nodi delle reti e il numero di archi (questo coefficiente è pari a 0,9045). In questo caso, si ha una maggiore correlazione tra il numero di note utilizzate e le relazioni che esse intrattengono, rispetto al confronto tra durata-numero di note. Interessante è il modo in cui si distribuiscono gli archi. Essi vanno da un minimo di 14 ad un massimo di 1728. In Figura 6.4, si riporta la distribuzione degli archi a partire dal più grande al più piccolo. A destra la versione logaritmica di questa distribuzione. Da questi grafici, si rileva il tipico andamento di una legge di potenza, con piccole differenze ai valori estremi.

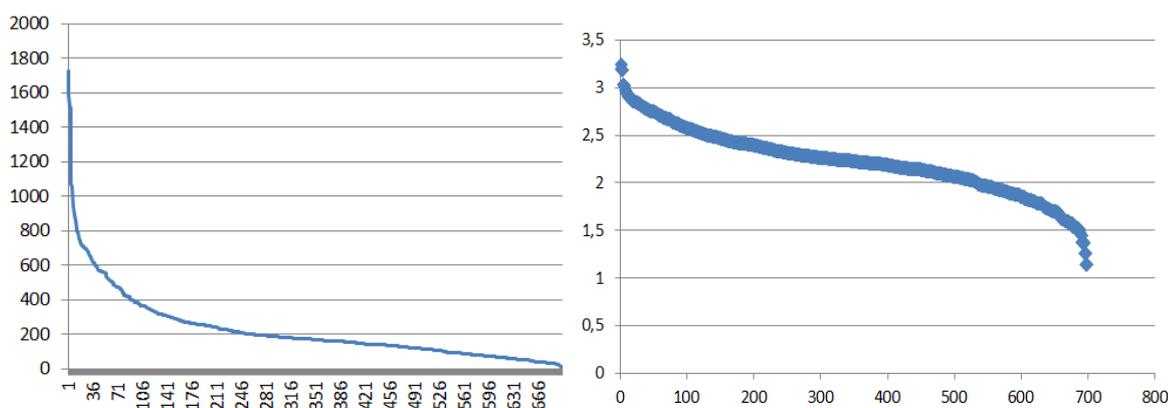


Figura 6.4. Distribuzione della frequenza degli archi.

In Figura 6.5 è possibile osservare che il diametro delle reti musicali va da un minimo di 4 ad un massimo di 18. Si noti che questo parametro non dipende dal numero di nodi.

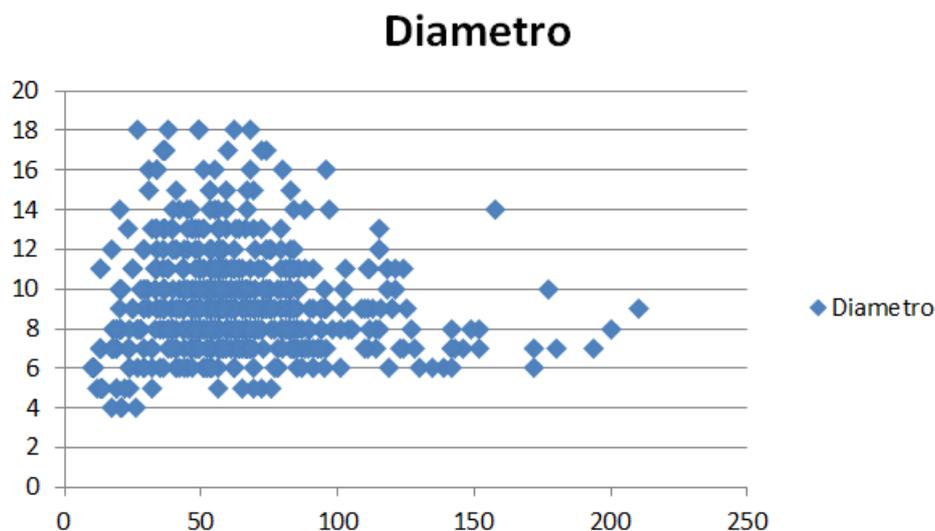


Figura 6.5. Diametro delle reti vs. il numero di nodi delle stesse.

6.3.1. Statistica a livello della singola traccia

Per ogni rete risultante da ogni melodia, sono state calcolate le seguenti grandezze statistiche, introdotte nel Capitolo 2:

1. Numero totale di nodi,
2. Numero totale di archi,
3. Grado medio,
4. Coefficiente di clustering,
5. Efficienza,
6. Lunghezza media,
7. Diametro della rete,
8. Densità,
9. Caratteristiche di piccolo mondo e invarianza di scala.

In Tabella 6.3 si riportano i valori medi di queste grandezze, considerando le varie collezioni.

Collezioni	Nodi	Archi	Grado	Clustering	Efficienza	Cammino	Diametro	Densità
Art Of Fugues (AOF)	71,096	291,161	8,069	0,181	0,355	3,307	8,66	0,058
Brandeburgo Concertos (BC)	83,113	417,189	9,885	0,216	0,387	3,025	7,860	0,062
Cantatas (C)	35,555	117,444	6,250	0,242	0,385	3,111	7,888	0,098
Cello suites (CS)	54,153	187,769	7,107	0,217	0,371	3,282	8,461	0,078
Fugues (F)	84	293,875	6,590	0,153	0,316	3,825	9,75	0,044

(Goldberg Variations) (MO)	54,826	182,530	5,633	0,156	0,321	3,908	10,214	0,057
Inventions (I)	53,666	158,166	5,906	0,169	0,330	3,623	9,166	0,057
Organ Works (OW)	50,862	193,945	6,295	0,181	0,349	3,452	8,944	0,077
Organ Chorales (OC)	35	91,857	4,973	0,170	0,335	3,578	9,303	0,087
Sonatas & Partitas (SP)	50,666	141,555	5,189	0,182	0,326	3,904	10,240	0,063
Sinfonias (S)	58,022	145,755	5,037	0,137	0,298	4,127	10,711	0,045
The Well-Tempered Clavier, Book I (WTC, BI)	58,558	175,909	5,904	0,175	0,323	3,775	9,990	0,054
The Well-Tempered Clavier, Book II (WTC, BII)	136,416	727,333	10,675	0,253	0,371	3,019	7,541	0,040

Tabella 6.3. Valori medi dei parametri di rete, relativi alle varie collezioni.

Dalla Tabella 6.3 si possono osservare significative differenze tra melodie che appartengono a differenti Collezioni del repertorio di Bach. Per esempio, le reti della Collezione *The Well-Tempered Clavier, Book II (WTC, BII)*, possiedono un numero medio di nodi significativamente più grande, rispetto alle altre collezioni. Negli *Organ Chorales (OC)*, questo numero è molto basso (attorno a 35). Cruciale è anche il ruolo che ciascuna nota gioca come una sorgente di connessione con altre note, dove queste connessioni sono maggiori nel *The Well-Tempered Clavier, Book II* rispetto alle altre Collezioni. Questo fenomeno è legato al grado medio più alto (10,68), rispetto alla Collezione *Organ Chorales (OC)* (4,93). Un numero maggiore di archi e nodi non implicano una perdita significativa di efficienza nel *The Well-Tempered Clavier, Book II*. Questo parametro, infatti, è intorno a 0,3. Una decrescita significativa si manifesta nella Collezione *Sinfonias (S)*, con un valore più basso intorno a 0,3. Il valore del diametro delle reti, nella Collezione *The Well-Tempered Clavier, Book II*, è molto basso (attorno a 7,5), indicando che due nodi possono essere raggiunti con un cammino di meno di 8 archi. Invece più di 10 passi sono necessari per unire in media due note nelle collezioni *Sonatas & Partitas (SP)* e *Sinfonias (S)*. La lunghezza media è intorno a 3 nella Collezione del *The Well-Tempered Clavier, Book II*, mentre è più di 4 nelle *Sinfonias (S)*. Il clustering è mediamente maggiore di 2 per il *The Well-Tempered Clavier, Book II*, *Cello suites (CS)*, *Brandeburgo Concertos (BC)* e *Cantatas (C)*, ma si abbassa significativamente per le Collezioni *Sinfonias (S)* e *Art Of Fugues (AOF)*.

6.3.2. Strutture di Piccolo Mondo, a livello delle singole collezioni

Come è stato discusso nel Capitolo 2, le reti di piccolo mondo sono caratterizzate da un alto coefficiente di clustering e lunghezza media comparabile, rispetto a quella di reti random con lo stesso numero di nodi e di archi. I parametri γ , λ , σ , già introdotti nel capitolo 2, possono essere utilizzati per caratterizzare queste reti. In accordo con Watts e Strogatz (1998), una rete è di piccolo mondo se $\gamma \gg 1$, $\lambda \approx 1$ e $\sigma > 1$. Se si considerano 1000 reti random, per ciascuna delle 698 reti ottenute dalle melodie del repertorio di Bach, e si calcolano i valori γ , λ , σ , si ottengono i risultati della Tabella 6.4.

Collezioni	γ	λ	σ	Small Word Networks %
Art Of Fugues (AOF)	1.423269	1.099818	1.295044	88.7%
Brandeburgo				
Concertos (BC)	1.090133	1.073125	1.012467	51.9%
Cantatas (C)	1.22533	1.108479	1.100033	44.4%
Cello suites (CS)	1.088962	1.113497	0.979206	30.7%
Fugues (F)	1.977567	1.154518	1.683503	75%
The Musical Offering (Goldberg Variations) (MO)	1.24215	1.126251	1.098583	62.2%
Inventions (I)	1.395495	1.126479	1.24254	66.7%
Organ Works (OW)	1.071933	1.088371	0.978943	50.4%
Organ Chorales (OC)	0.974791	1.086736	0.893401	41.1%
Sonatas & Partitas (SP)	1.255708	1.130341	1.130274	64.2%
Sinfonias (S)	1.454684	1.137012	1.271267	71.7%
The Well-Tempered Clavier, Book I (WTC, BI)	1.363462	1.129722	1.212645	70.3%
The Well-Tempered Clavier, Book II (WTC, BII)	1.316291	1.024282	1.285528	91.7%

Tabella 6.4. Valori medi di γ , λ , σ e percentuali di reti di piccolo mondo per ciascuna collezione..

Analizzando accuratamente la Tabella 6.4, si osserva che i valori di σ non sono molto alti ed addirittura in due categorie sono minori di 1. Inoltre il numero di reti di piccolo mondo è molto alta nel caso *Art Of*

Fugues (AOF) e della Collezione *The Well-Tempered Clavier, Book II (WTC, BII)*, mentre è molto basso nel caso della collezione *Cello suites (CS)* e nelle composizioni degli *Organ Chorales (OC)*. Questo risultato dimostra che la caratteristica di piccolo mondo non è una proprietà generale per questo tipo di reti. Il fenomeno è ancora più evidente se si osserva la Figura 6.6 dove molte reti della Collezione *Art Of Fugues (AOF)* hanno $\sigma > 1$ (Figura 6.6a), mentre altre reti degli *Organ Chorales (OC)* hanno $\sigma < 1$ (Figura 6.6b).

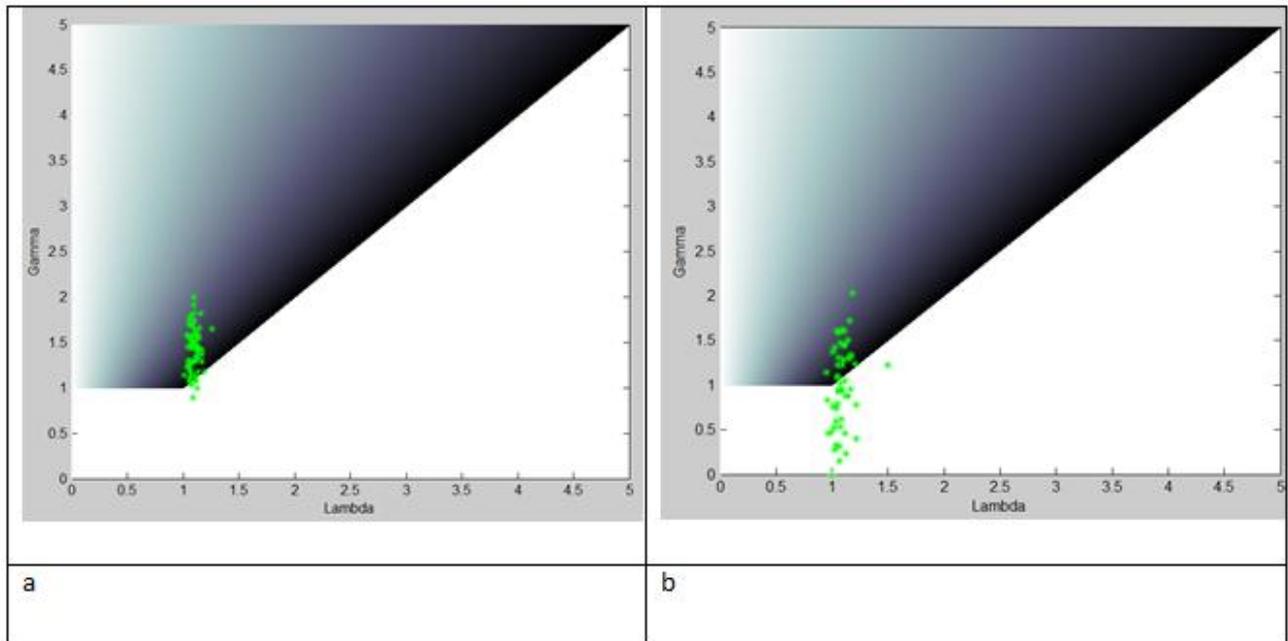


Figura 6.6. Distribuzione dei valori dei parametri γ, λ, σ , per le reti che appartengono alle collezioni *Art Of Fugues (AOF)* (a) e *Organ Chorales (OC)* (b). Questi diagrammi individuano come si distribuiscono le reti nello spazio dei parametri γ, λ . Si noti che più i punti si trovano nelle zone più scure, maggiori sono le caratteristiche di piccolo mondo delle reti.

6.3.3. Reti con Invarianza di Scala

Sebbene i dati sperimentali di ciascuna rete della collezione siano un campione insufficiente, si può investigare comunque la presenza di strutture scale-free. Solo 19 reti su 698 seguono una legge di potenza per la distribuzione dell'In-degree e 29 per l'Out-degree. Questi dati, in generale, sono intercettati in modo più appropriato da *Leggi di Potenza Troncate* (Strogatz, 2001; Albert & Barabási, 2002), dove le distribuzioni sono date seguendo la formalizzazione riportata in (Cluset et al. 2009):

$$(6.1) \quad P(k) = ae^{(-\frac{k}{kc})}k^b$$

ed in qualche caso da leggi esponenziali

$$(6.2) \quad P(k) = ae^{(-\frac{k}{kc})}$$

Se si considera la legge di potenza $P(k) = ab^k$, nelle Figure 6.7a e 6.7b, è possibile ottenere curve che intercettano l'andamento dei dati sperimentali (curve di fitting) per l'In-Degree per le rete B20 e WBI87

dove $a = 1.105$ e $b = -0.641$ per la B20, e $a = 1.027$ - 0.969 e $b = -0.969$ per WBI87, con $R^2 = 0.897$ e 0.977 , rispettivamente. Si ricorda che quanto più R^2 si avvicina a 1, tanto più le curve intercettano i dati sperimentali (Kvålseth, 1985, Cluset et al. 2009).

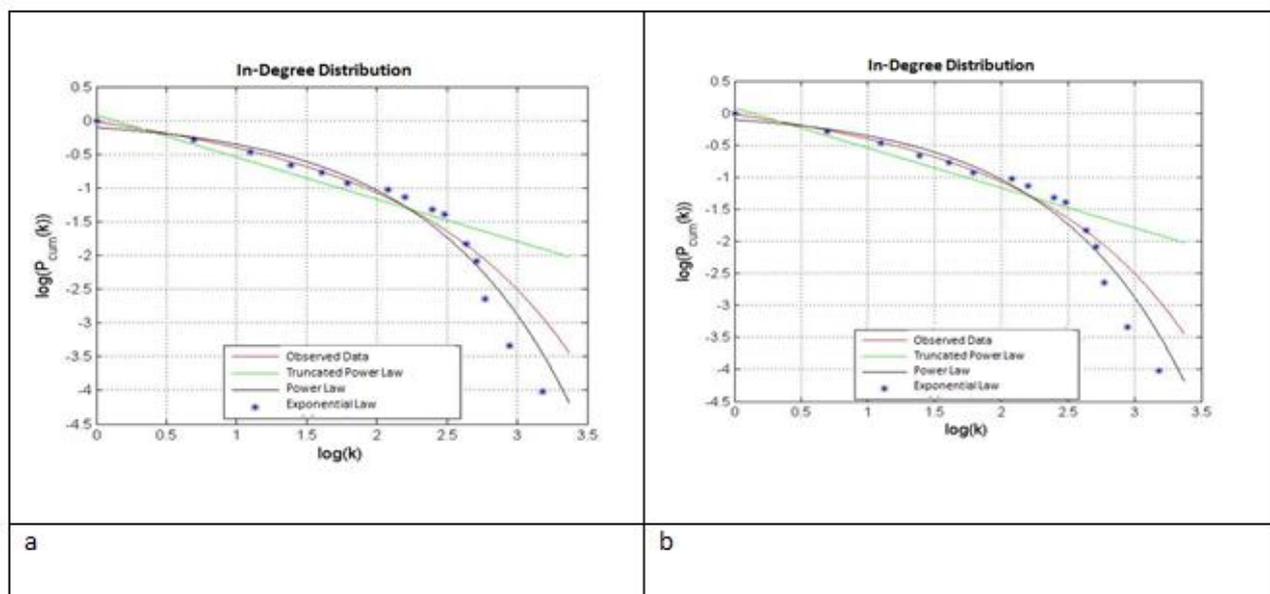


Figura 6.7. Due grafici con curve di fitting per le reti B20 e WBI87 rispettivamente. Come è possibile notare, alcuni dati sperimentali non intercettano la legge di potenza, ma la legge di potenza troncata (Bertacchini et al., sottomesso, a)

6.4. Statistiche di rete a livello delle Collezioni

Consideriamo ora le reti emergenti a livello di collezioni. Queste possono essere ottenute unendo le singole reti che compongono ogni collezione. Ovviamente in questo modo otteniamo reti con un maggior numero di nodi e di archi rispetto alle singole reti. Se calcoliamo le grandezze statistiche delle 13 reti, otteniamo i valori della Tabella 6.5.

Collezioni	Nodi	Archi	Grado	Clustering	Efficienza	Cammino	Diametro	Densità
Art Of Fugues (AOF)	269	5381	40,007	0,411	0,431	2,643	7	0,074
Brandeburgo Concertos (BC)	423	8945	42,293	0,425	0,449	2,475	9	0,050
Cantatas (C)	124	701	11,306	0,298	0,365	3,155	8	0,045
Cello suites (CS)	154	1445	18,766	0,278	0,407	2,777	8	0,061
Fugues (F)	195	1319	13,528	0,305	0,344	3,306	9	0,034
The Musical Offering (Goldberg)	282	4727	33,524	0,401	0,450	2,451	6	0,059

Variations) (MO)								
Inventions (I)	245	2644	21,583	0,300	0,368	2,949	8	0,044
Organ Works (OW)	243	2459	20,238	0,261	0,369	3,047	8	0,041
Organ Chorales (OC)	363	8788	48,418	0,365	0,440	2,561	14	0,066
Sonatas & Partitas (SP)	314	3774	24,038	0,313	0,383	2,986	10	0,038
Sinfonias (S)	282	3672	26,042	0,258	0,391	2,891	8	0,046
The Well-Tempered Clavier, Book I (WTC, BI)	310	8177	52,754	0,351	0,468	2,399	7	0,085
The Well-Tempered Clavier, Book II (WTC, BII)	298	10878	73,006	0,389	0,519	2,127	7	0,122

Tabella 6.5. Grandezze statistiche per le reti ottenute dall'unione di tutte le reti, per tutte le collezioni.

La media dei parametri delle 13 reti ha 270 nodi e 4840 archi, con un coefficiente di correlazione pari a 0,805. Il grado medio è attorno a 32,731, valore che indica un alto grado di correlazione, pari a 0,961. Invece il valore di correlazione tra grado medio e nodi è molto basso e pari a 0,645. A dispetto del basso numero di note che compongono lo spazio musicale delle Collezioni nel loro insieme, le note hanno un alto grado, emergente da strutture ben correlate con un gran numero di note musicali connesse da molti archi, sia in ingresso che in uscita. Le reti con un alto numero di archi possiedono anche un alto numero di strutture organizzate in cluster. I nodi con alto grado sono più facilmente individuabili nelle zone più dense delle reti, come si può osservare dalle Figure 6.8 e 6.9. La Figura 6.8 è una rappresentazione della rete relative alla Collezione *The Musical Offering, Goldberg Variations, (MO)*, mentre la Figura 6.9 corrisponde alle reti delle Collezioni *Fugues (F)* (a), *Organ Works (OW)* (b), *Inventions (I)* (c), *Cello Suites (CS)* (d) e *The Well-Tempered Clavier, Book II (WTC, BII)* (e).

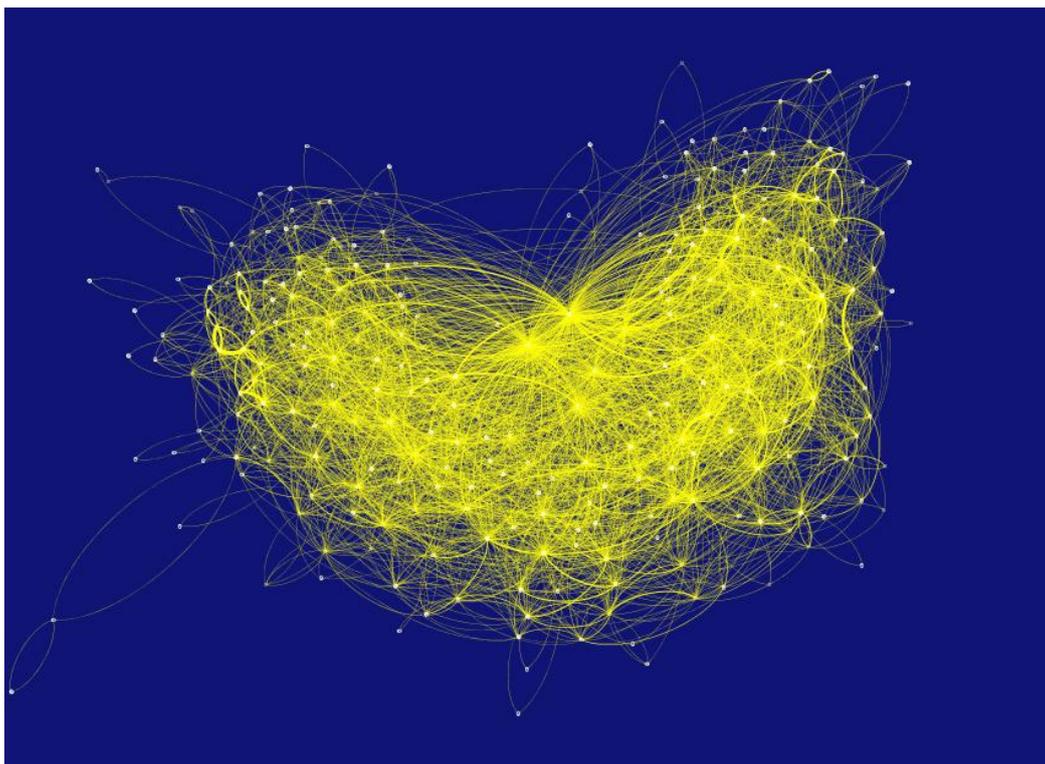
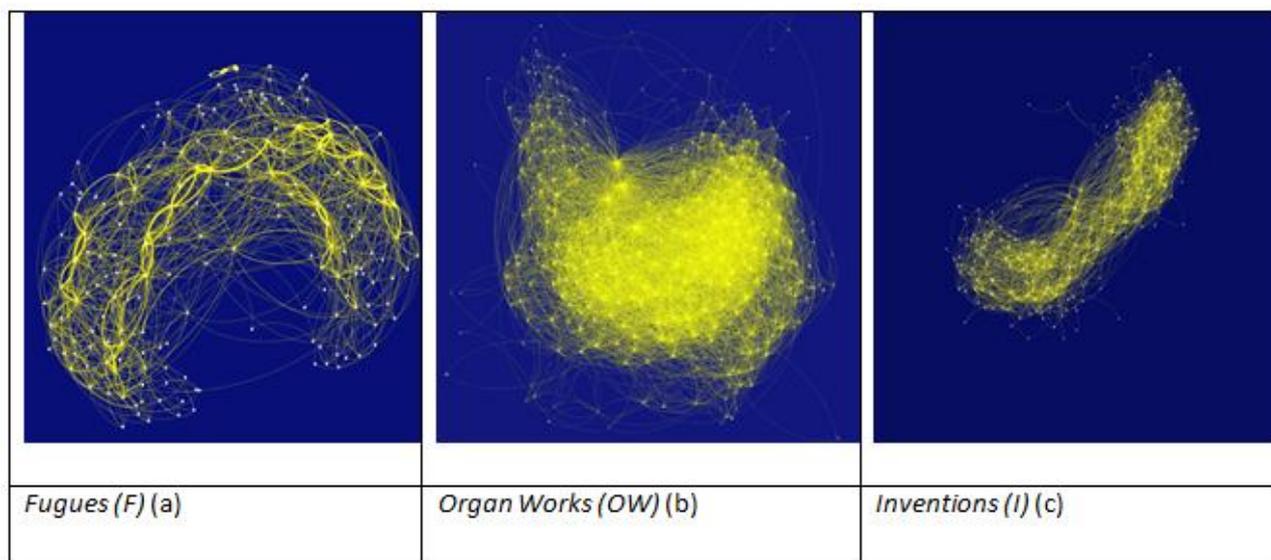


Figura 6.8. Una rappresentazione della rete relativa alle Goldberg Variations (GV), realizzata con la visualizzazione Force Atlas del software Gephi.



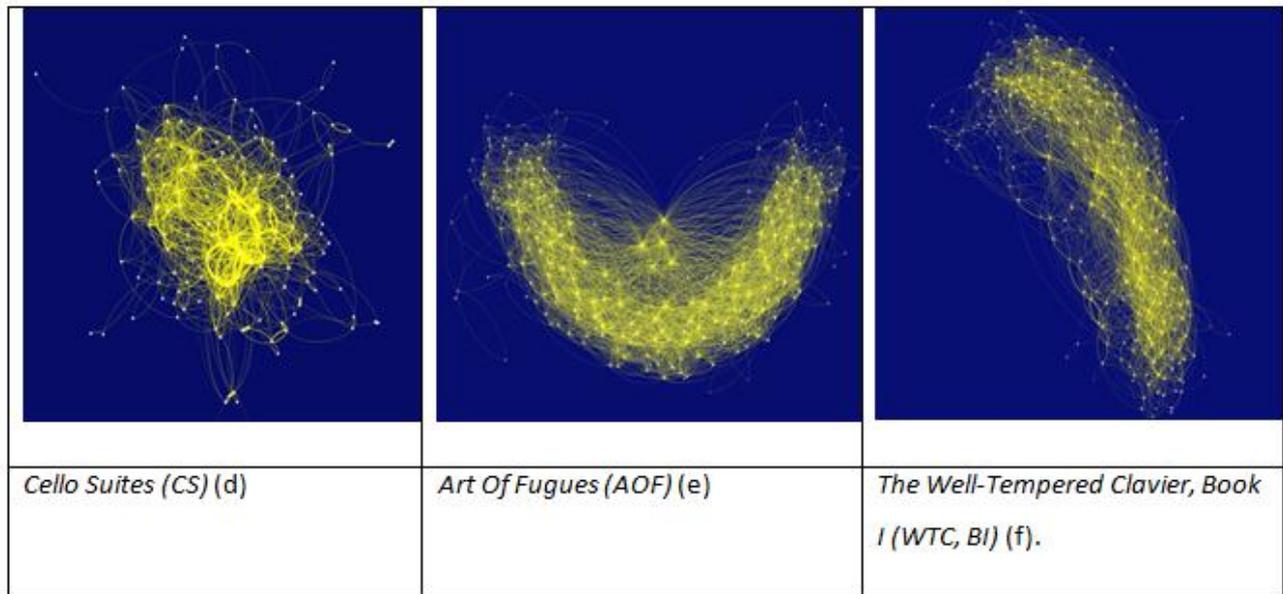
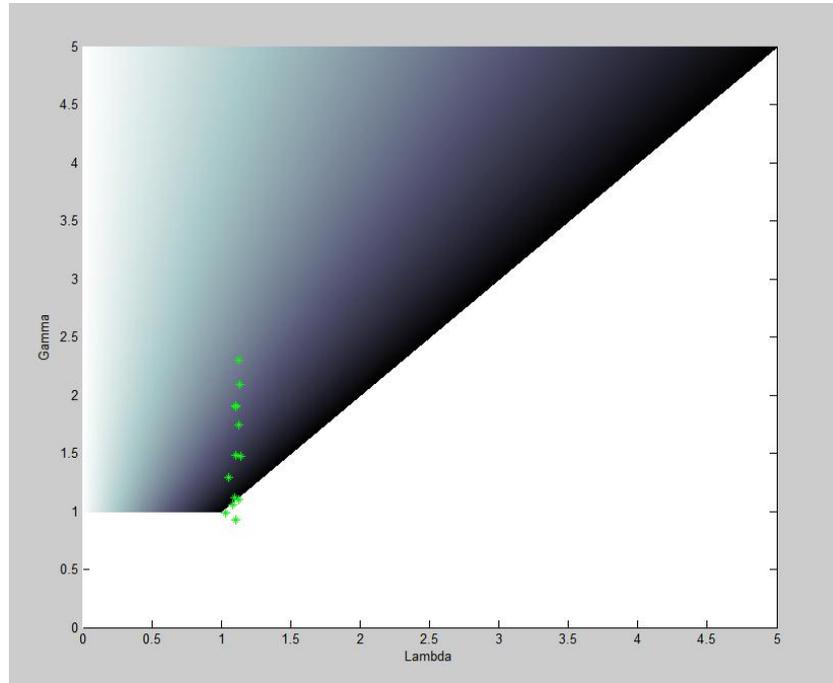


Figura 6.9. Galleria di reti. Si noti la similarità nei pattern che sembrano costellazioni, con una depressione centrale (realizzata da pause), e strutture ricorsive che formano i principali motivi melodici, particolarmente evidenti nelle reti delle fughe e dei canoni.

6.4.1. Strutture di piccolo mondo

I parametri di piccolo mondo per le 13 collezioni sono riportati nella Tabella 6.6. Quasi tutte le reti hanno $\gamma \gg 1, \lambda \approx 1$ and $\sigma > 1$, con alcune differenze. Tre eccezioni sono le reti della Collezione *The Well-Tempered Clavier, Book II (WTC, BII)*, che hanno valori di γ e σ minori di 1, e le reti delle Collezioni *Brandeburgo Concertos (BC)* e *Organ Chorales (OC)*, che hanno valori di σ minori di 1.

Collezioni	γ	λ	σ
Art Of Fugues (AOF)	1,744711	1,128996	1,545364
Brandeburgo Concertos (BC)	1,054725	1,079801	0,976778
Cantatas (C)	1,90977	1,108018	1,72359
Cello suites (CS)	1,119413	1,096083	1,021284
Fugues (F)	2,30319	1,130129	2,037989
The Musical Offering (Goldberg Variations) (MO)	1,293364	1,05534	1,225543
Inventions (I)	1,898888	1,10838	1,713209
Organ Works (OW)	2,09208	1,131709	1,848603
Organ Chorales (OC)	1,099823	1,127566	0,975395
Sonatas & Partitas (SP)	1,476001	1,141443	1,2931
Sinfonias (S)	1,740093	1,128442	1,542031
The Well-Tempered Clavier, Book I (WTC, BI)	1,48444	1,106901	1,341078
The Well-Tempered Clavier,	0,987168	1,030574	0,957881

Tabella 6.6. Valori medi γ , λ , σ per tutte le reti complessive di ogni collezione.Figura 6.10. Distribuzione dei valori dei parametri γ , λ , σ per le reti globali delle 13 collezioni.

Il diagramma di Figura 6.10 evidenzia la distribuzione delle reti nel piano dei parametri γ , e λ . Analogamente a quanto visto per le singole tracce, più i punti sono nell'area scura più le caratteristiche di piccolo mondo sono presenti. Si noti che tali caratteristiche sono maggiormente marcate, raggruppando tutte le reti per tutte le collezioni, di quando si considerano le reti al livello delle singole melodie.

6.4.2. Reti ad Invarianza di scala

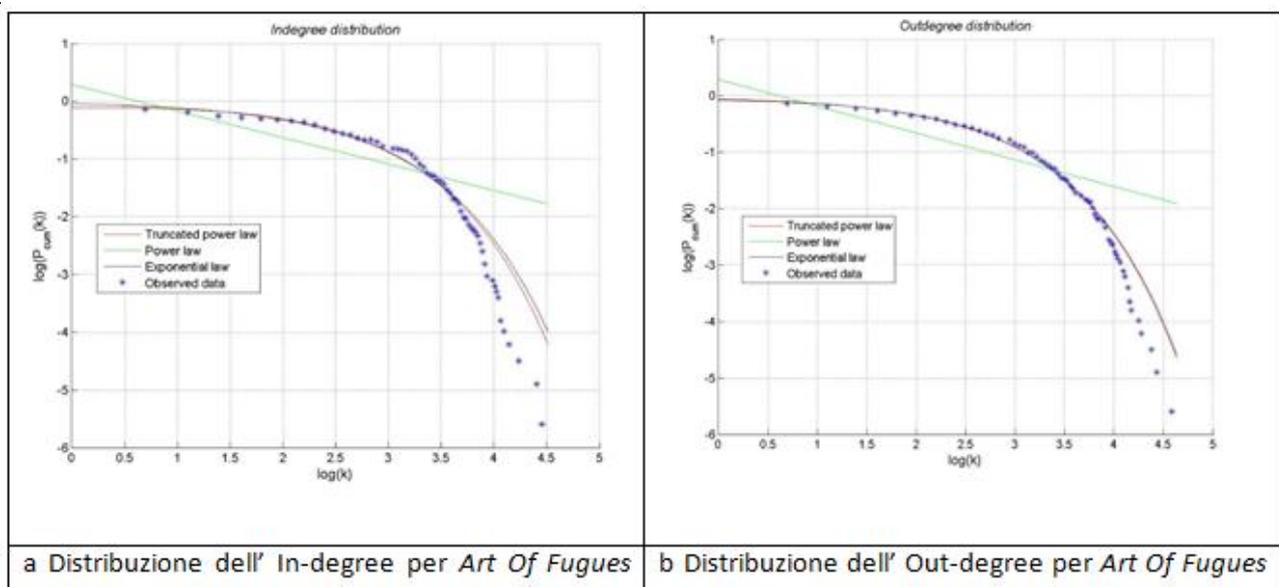
Per le reti generate dalle collezioni, si calcolano i parametri di fitting per le leggi di potenza, leggi di potenza troncata e legge esponenziale. I risultati sono evidenziati nella Tabella 6.7.

Collezioni	a	kc	b	R2	a	b	R2	a	kc	R2	Scale free?
Art Of Fugues (AOF)	0,951	44,738	0,0160	0,990	1,473	-0,384	0,718	0,977	46,232	0,989	NO
Brandeburgo Concertos (BC)	1,072	57,907	-0,103	0,998	1,886	-0,494	0,841	0,567	0,075	-1,535	NO
Cantatas (C)	1,193	12,641	-0,093	0,993	1,911	-0,733	0,896	0,450	0,083	-1,311	NO
Cello suites (CS)	1,111	27,364	-0,142	0,989	1,337	-0,490	0,845	0,974	19,989	0,974	NO
Fugues (F)	1,108	20,084	-0,165	0,990	1,277	-0,548	0,862	0,987	14,311	0,975	NO

The Musical Offering (Goldberg Variations) (MO)	1,018	36,793	-0,024	0,997	1,897	-0,504	0,820	0,015	0,043	-1,626	NO
Inventions (I)	1,057	32,329	-0,118	0,984	1,326	-0,464	0,823	0,926	25,161	0,973	NO
Organ Works (OW)	1,038	20,386	0,011	0,996	1,403	-0,481	0,764	1,051	20,829	0,996	NO
Organ Chorales (OC)	1,039	58,911	-0,053	0,998	1,933	-0,462	0,811	0,940	52,799	0,996	NO
Sonatas & Partitas (SP)	1,058	29,598	-0,084	0,996	1,394	-0,483	0,810	0,962	24,852	0,992	NO
Sinfonias (S)	1,029	27,419	-0,003	0,997	1,452	-0,456	0,757	1,024	27,222	0,997	NO
The Well-Tempered Clavier, Book I (WTC, BI)	0,967	50,362	0,048	0,9907	1,614	-0,367	0,683	1,056	55,524	0,988	NO
The Well-Tempered Clavier, Book II (WTC, BII)	1,011	82,902	-0,013	0,993	1,595	-0,339	0,720	0,984	80,287	0,993	NO

Tabella 6.7. I parametri di fitting per le reti relative alle 13 collezioni. Più R^2 si avvicina ad 1, maggiore è l'attendibilità della legge che interpola le reti.

A differenza del caso precedente, nessuna legge di potenza sembra ben rappresentare le distribuzioni per l'in-degree e l'out-degree dei nodi. Nella Figura 6.11, sono riportate le distribuzioni per l'in-degree e l'out-degree dei nodi le Collezioni *Art Of Fugues (AOF)* e *The Musical Offering (Goldberg Variations) (MO)*.



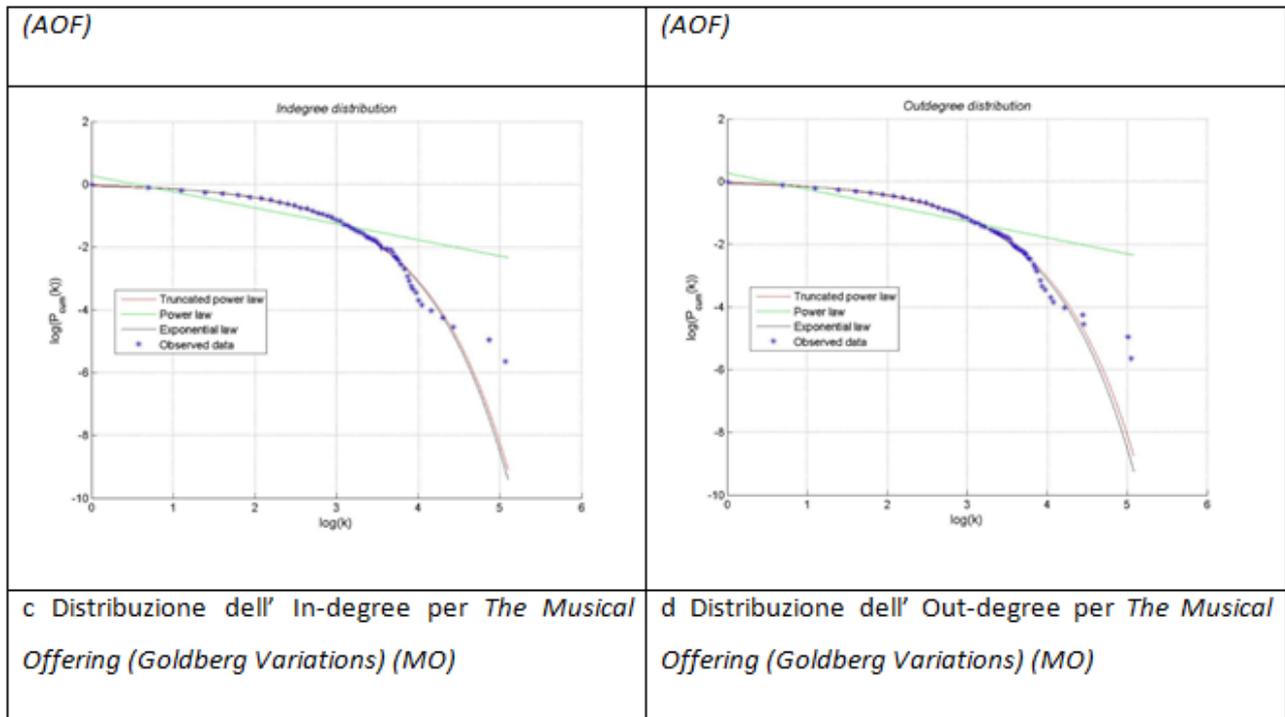


Figura 6.11. Distribuzione per l'In-degree (a)-(c) e per l'out-degree (b)-(d) dei nodi per le reti relative alle Collezioni Art Of Fugues (AOF) e The Musical Offering (Goldberg Variations) (MO). I dati sperimentali non sono intercettati bene da una legge di potenza, ma da leggi esponenziali e leggi di potenza troncata.

6.5. Statistiche di rete a livello dell'intera collezione di Bach

Se si considera la rete ottenuta dall'unione delle reti di tutte le collezioni, si può osservare che essa sembra possedere la stessa struttura delle reti delle singole Collezioni, mantenendo un'invarianza di pattern, attraverso differenti livelli di visualizzazione, nonostante questa visualizzazione dipenda fortemente dal tipo di algoritmo utilizzato. Le immagini di Figura 6.12 sono realizzate utilizzando l'algoritmo della visualizzazione *Force Atlas* del software *Gephi*. La rete dell'intero campione presenta 467 nodi e 26561 archi. Il grado medio è 56,876, mentre la lunghezza media è 2,3. Il diametro della rete è 9, mentre la densità è 0,122. La rete di tutte le Collezioni non presenta strutture di di piccolo mondo in quanto: $\gamma=0,92252$, $\lambda=1,102525$ e $\sigma=0,836734$.

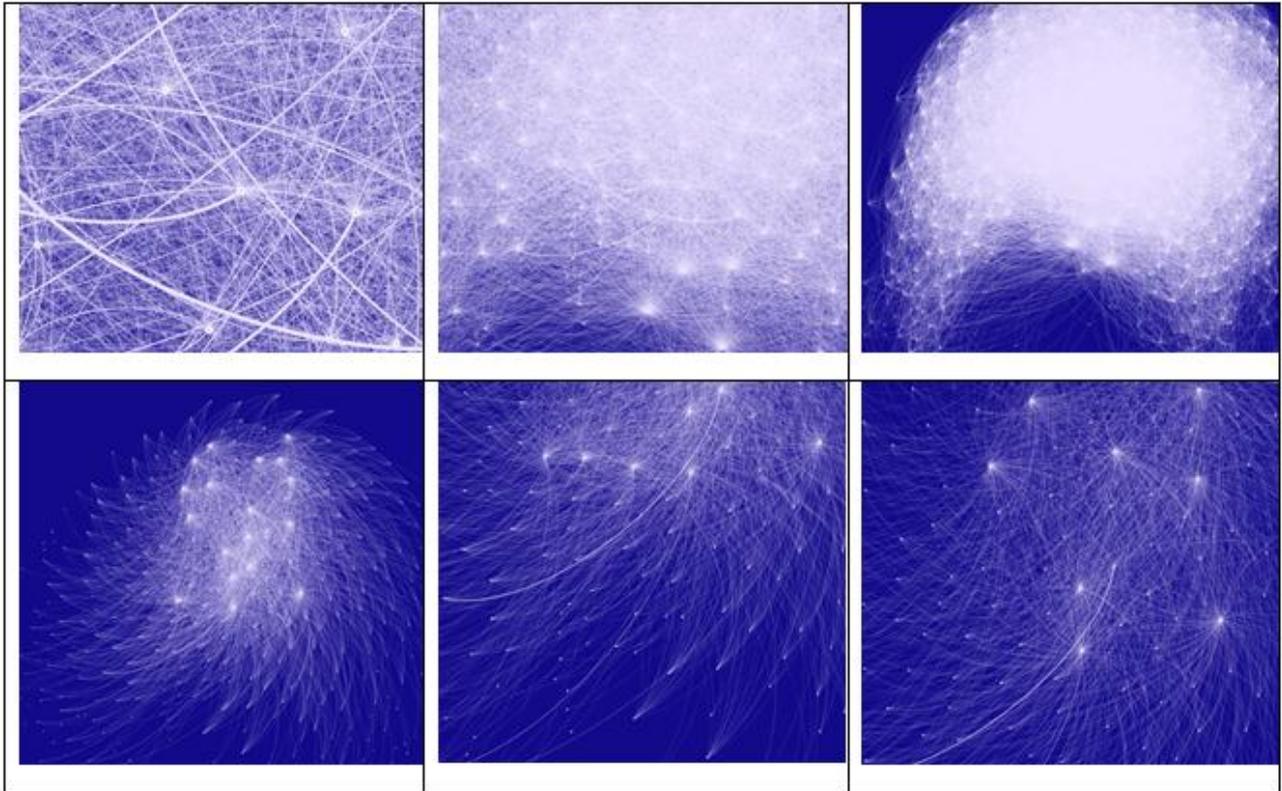


Figura 6.12. Immagini della rete realizzata unendo tutte le reti di tutte le Collezioni del repertorio musicale di Bach. Il pattern è stato ottenuto utilizzando l’algoritmo Force-Atlas di Gephi, ripreso in diversi momenti dell’evoluzione della rete stessa.

Questa rete musicale non presenta invarianza di scale, com’è evidente dal valore dei parametri riportati nella Tabella 6.8. Anche in questo caso, i dati osservativi sembrano meglio intercettati da una curva di potenza troncata che da una legge di potenza (Figura 6.13).

a	kc	b	R2	a	b	R2	a	kc	R2
0,902	156,173	-0,021	0,981	1,803	-0,35	0,745	0,9203	0,052	-2,455

Tabella 6.8. Parametri di invarianza di scala della rete relative all’intero campione delle Collezioni di Bach.

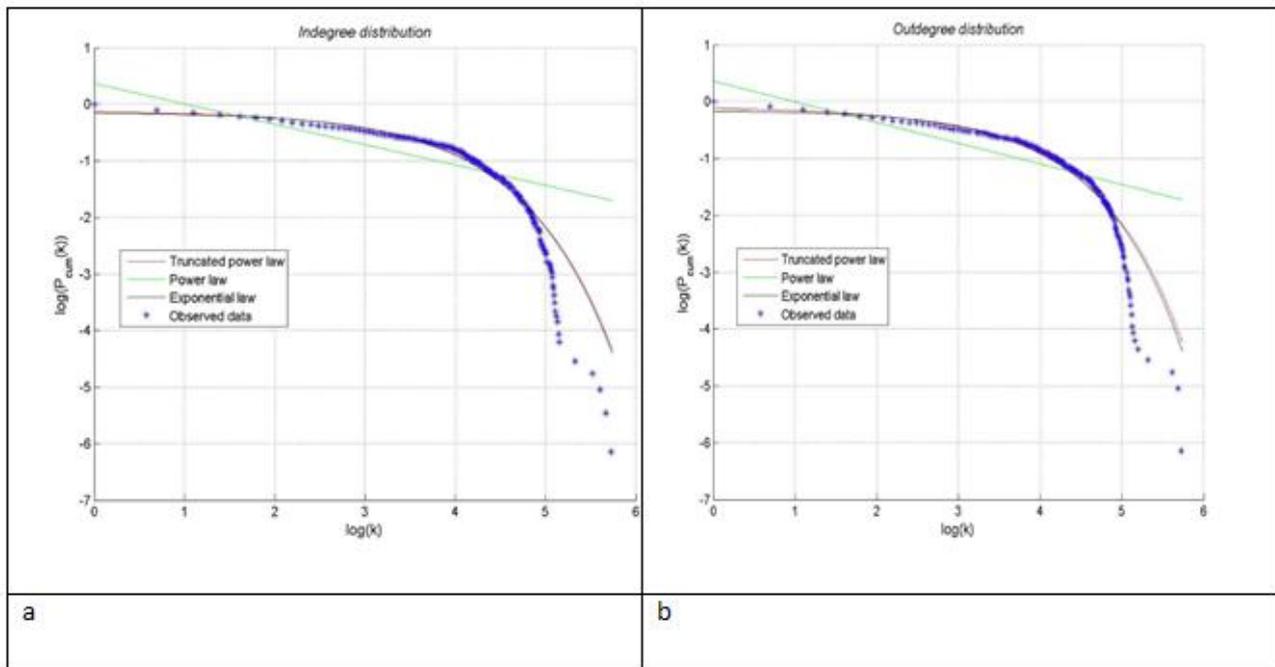


Figura 6.13. Distribuzioni per l' In-degree (a) e l'out-degree (b) dei nodi per la rete dell'intero campione delle melodie di Bach. Come si può osservare questi dati sono meglio intercettati da una legge di potenza troncata.

6.6. Motivi e Grammatiche musicali emergenti

In questo paragrafo si indaga se le reti musicali hanno una organizzazione realizzata da motivi più piccoli che, unendosi, possono realizzare livelli gerarchici superiori di strutturazione. Per tale analisi si analizza la prima traccia della prima melodia della collezione relativa alla Collezione *Art of Fugues (AOF)*.

Nella rete, tre terne di nodi con le etichette: $\{(5,0,G,1/16)(5,0,F,1/16)(5,0,E,1/16)\}$ e $\{(5,0,E,1/16)(5,0,D,1/16)(5,1,C,1/16)\}$. I nodi rappresentano tre note discendenti sulla quinta ottava, di eguale valore, che si ripetono 7 e 6 volte rispettivamente. Queste note sono date nel codice *.MET*, introdotto nel Capitolo 3. Si ricorda che il primo numero rappresenta l'ottava, il secondo numero se è presente (1) o assente (0) la deformazione *diesis* o *bemolle* della nota, il terzo e il quarto numero il nome in Inglese e il tempo della nota rispettivamente. In Figura 6.14, si presenta una visualizzazione del *motivo comune* che realizzano all'interno della melodia, attraverso le rappresentazioni con etichette (a , b) e senza (c).

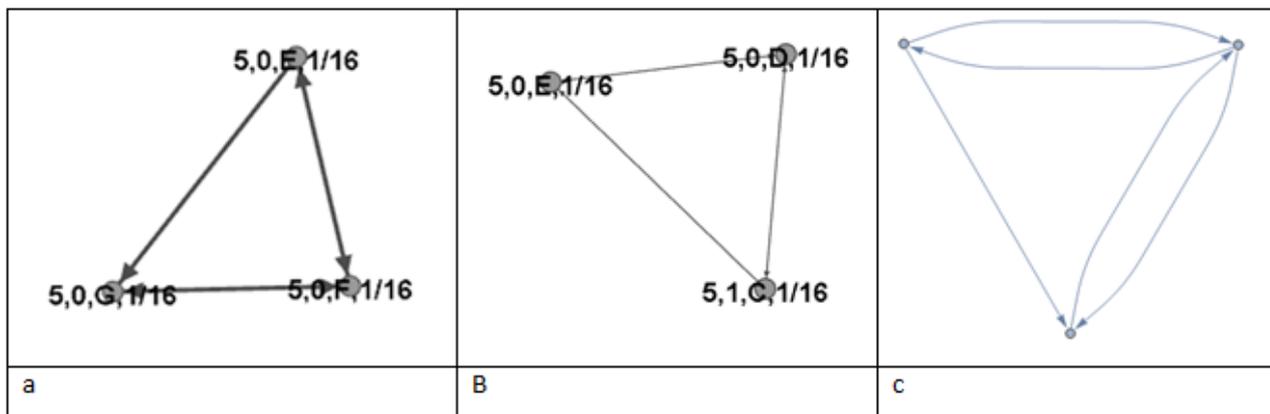


Figura 6.14. Due terne di note (a)-(b) come presenti nella rete della prima traccia della prima melodia di AOF, ed il motivo che realizzano (c).

In questo lavoro, i motivi con le etichette sono considerati come le occorrenze fisicamente definite delle note. Il grafico senza etichette, il modello matematico più generale, che contiene un'organizzazione astratta che determina le occorrenze fisiche delle note nella melodia. Il motivo (c) non è l'unico presente nella rete. Per esempio la terna $\{(4,0,F,1/16)(4,0,E,1/16)(4,0,D,1/16)\}$ è ripetuta 5 volte e realizza il motivo di Figura 6.15.



Figura 6.15. Motivo realizzato dalla terna di note $\{(4,0,F,1/16)(4,0,E,1/16)(4,0,D,1/16)\}$.

Il motivo (c) individua un cluster ed è tra i responsabili delle caratteristiche di piccolo mondo della rete, mentre il motivo di Figura 6.15 non porta a organizzazioni in cluster che determinano le caratteristiche di piccolo mondo. I motivi del tipo (c) sono presenti in molte reti musicali. Ad esempio, è presente con un valore pari al 2,3%, nell'undicesima traccia della Collezione *Art of Fugues (AOF)*, valore molto più alto dell'1%, percentuale che compare in una rete casuale. Il fenomeno è stato evidenziato tramite il software FANMOD, che consente di scoprire i motivi presenti in una rete (Figura 6.16).

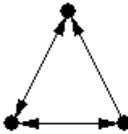
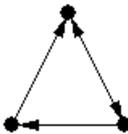
ID	Adj	Frequency [Original]	Mean-Freq [Random]	Standard-Dev [Random]	Z-Score	p-Value
174		2.2941%	1.0657%	0.0032079	3.829	0
102		2.2941%	1.3743%	0.0034124	2.6953	0.01

Figura 6.16. Motivi ottenuti tramite il software FANMOD nella melodia 11 della Collezione *Art of Fugues* (AOF).

Sempre facendo riferimento alla prima traccia della prima melodia della Collezione *Art of Fugues* (AOF), si esplora l'esistenza di altri motivi, con organizzazioni diverse. Ad esempio, il blocco di 4 note $\{(5,0,G,1/16)(5,0,F,1/16)(5,0,E,1/16)(5,0,D,1/16)\}$ è ripetuto 5 volte nella melodia e individua il motivo di Figura 6.17.

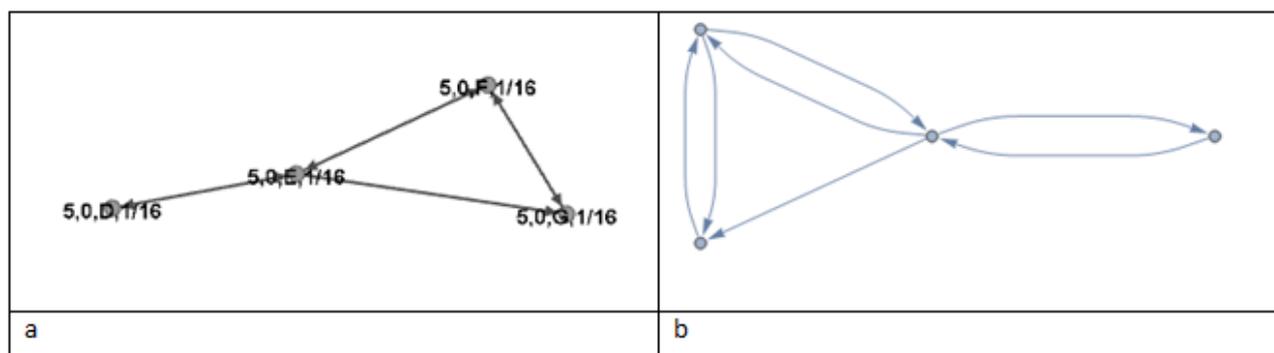


Figura 6.17. Il motivo di 4 note maggiormente ripetuto nella melodia, con le etichette (a) e il motivo matematico corrispondente (b).

Come si può osservare, quest'ultimo motivo è formato da un sotto-motivo, uguale al motivo (c) introdotto nella Figura 6.14, e un altro sotto-motivo, equivalente alla struttura di Figura 6.15. In un certo senso, questo motivo può essere considerato come un pattern che i due motivi precedentemente illustrati. Questo processo potrebbe continuare ulteriormente, cercando ripetizioni di blocchi di 5 note. Nella melodia considerata, il blocco $\{(5,0,G,1/16)(5,0,F,1/16)(5,0,E,1/16)(5,0,D,1/16)(5,1,C,1/16)\}$ è ripetuto 5 volte. Esso individua un nuovo motivo, rappresentato in Figura 6.18.

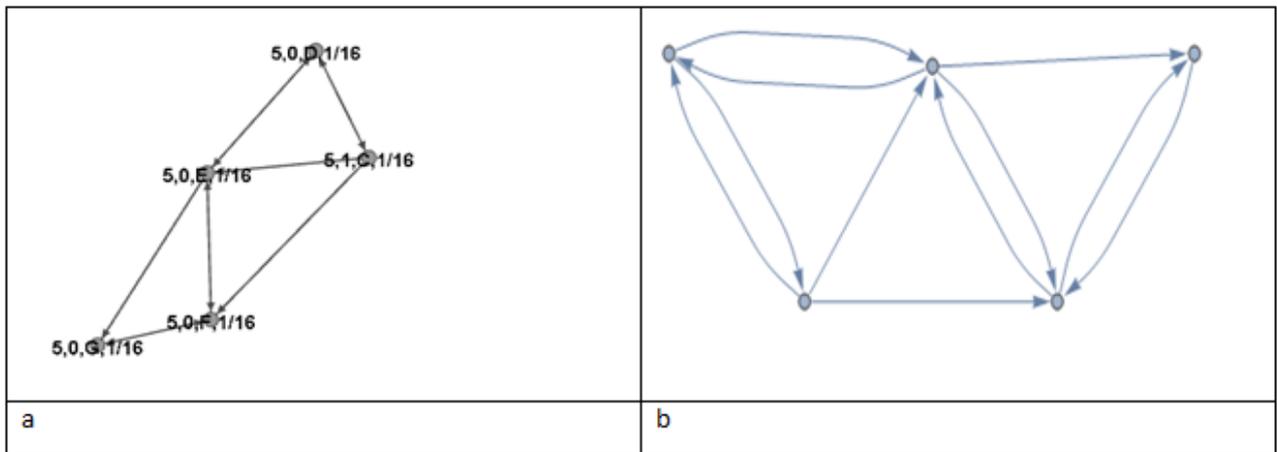


Figura 6.18. Il blocco di 5 note maggiormente ripetuto nella melodia (a) e il motivo corrispondente (b).

Si noti che il motivo visualizzato nella Figura 6.17 è di nuovo un sotto-motivo di quello della Figura 6.18. L'aggiunta di un nodo alla Figura 6.17 per ottenere il motivo di 5 note, comporta un drastico aumento di strutture clusterizzate di piccolo mondo. Il motivo di Figura 6.18 presenta inoltre interessanti simmetrie e rotture di simmetrie rispetto ad altri motivi. Anche questo motivo può essere ottenuto come somma di due motivi precedenti, come evidenziato nella Figura 6.19.

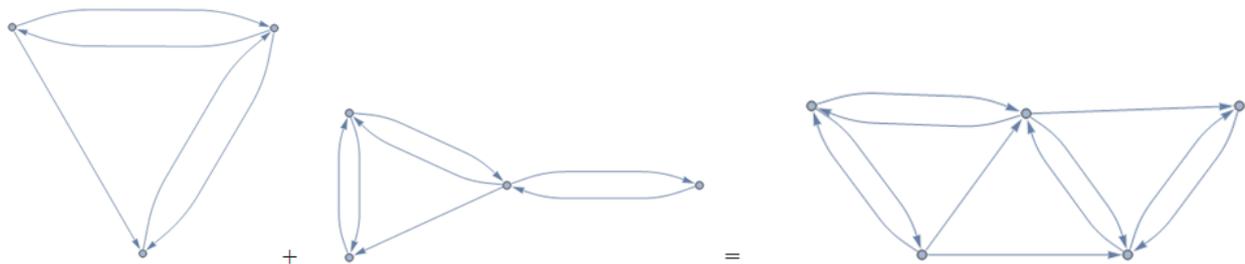


Figura 6.19. Un motivo di 6 nodi ottenuto come somma/composizione di un motivo di 3 più un motivo di 4 nodi.

Questi motivi possono essere considerati come le parti costituenti di una grammatica generativa, analoga a quella introdotta nel capitolo 1, per le lingue naturali e i linguaggi formali. In questo caso, si può pensare che le note siano analoghe ai fonemi e che esse, accoppiandosi tra loro, formino motivi di tre note. Tali strutture costituiscono la base dei morfemi che accoppiandosi a loro volta costituiscono parole, che si uniscono insieme per formare elementi dell'ordito melodico (il lessico), che si ritrovano come strutture ricorrenti nella rete.

La presenza di tali raggruppamenti (motivi) è un'evidente dimostrazione della presenza di numerose simmetrie. A livello sociale, i motivi sono rappresentati da *clique*, che rappresentano gruppi di individui interagenti (Travers & Milgram, (1969). Le *clique*, già introdotte nel capitolo 2, sono formate da nodi che presentano legami diretti con tutti gli altri nodi, per cui esse sono grafi perfetti. Nella melodia considerata (

Figura 6.20a), la *clique* di dimensione massima è composta dalla terna di nodi indicata con le etichette: $\{(4,0,F,1/8) (4,0,E,1/16) (4,0,G,1/16)\}$, riportata in Figura 6.20b).

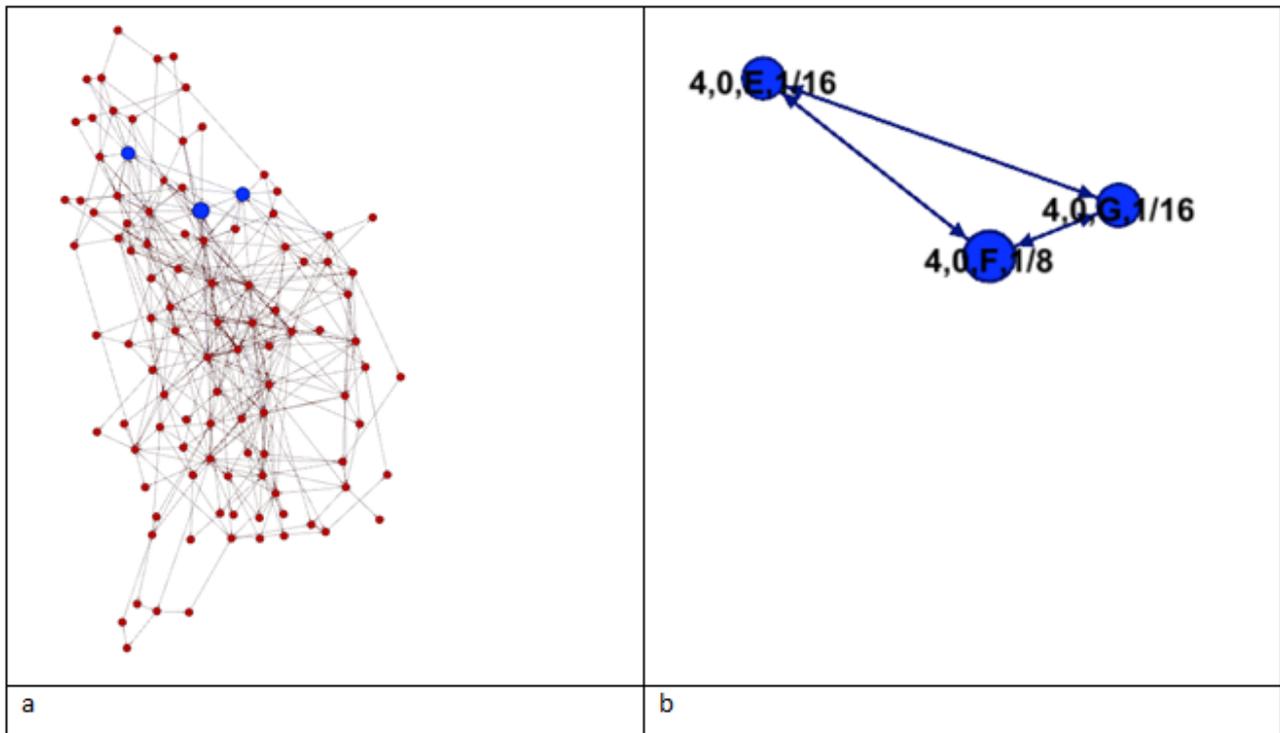


Figura 6.20. (a) La rete della prima traccia della prima melodia appartenente alla Collezione Art of Fugues (AOF). In azzurro è rappresentata la terna che individua la *clique*. In (b) il dettaglio delle note che compongono la struttura.

In un certo senso, i motivi di tre nodi descritti in precedenza, possono essere visti come delle rotture di simmetrie, rispetto alla simmetria perfetta della *clique*.

Considerando la rete di Figura 6.20a come grafo indiretto, si osserva la presenza di una *clique* di 6 nodi, il cui motivo è un esagono perfetto (Figura 6.21b).

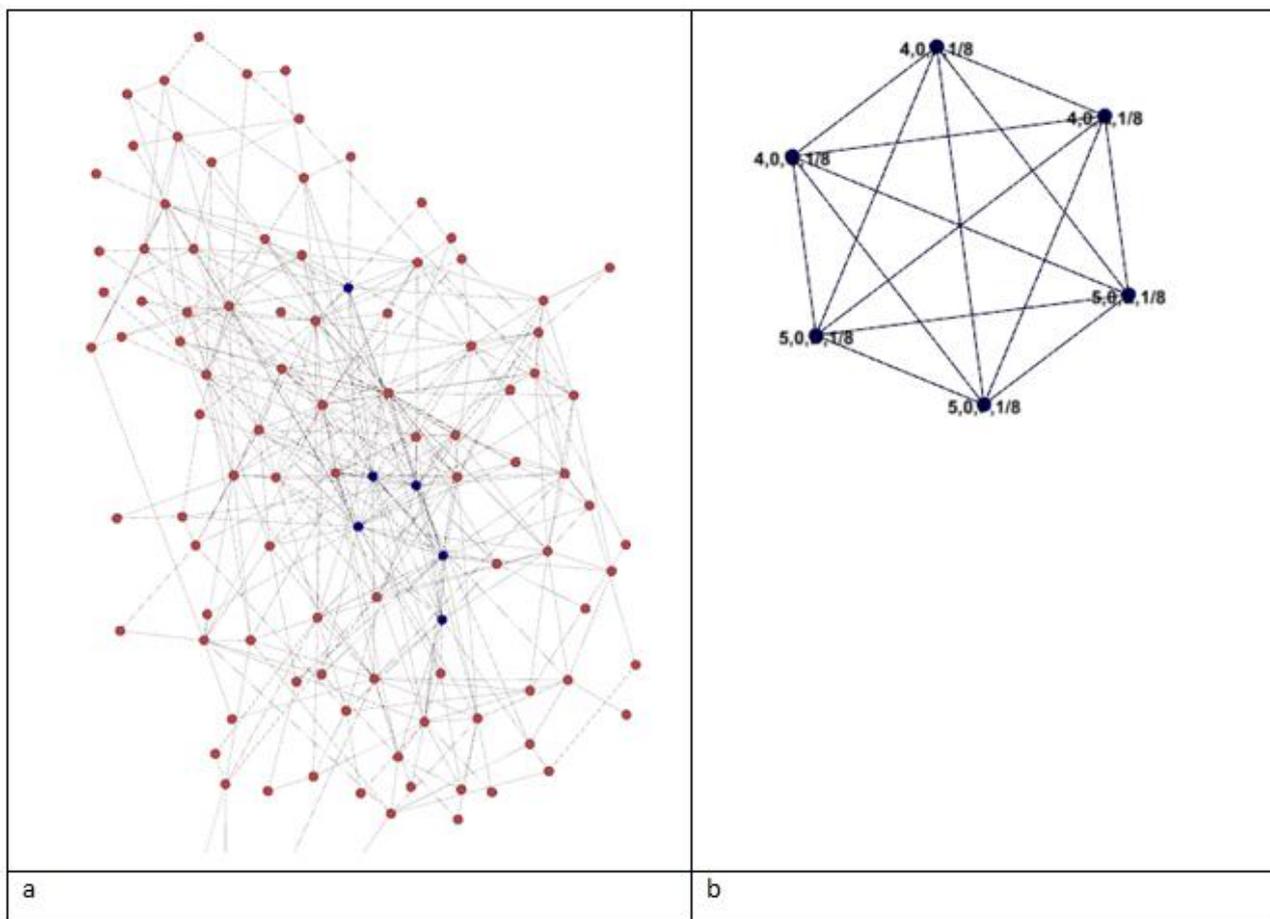


Figura 6.21. In questa immagine, la melodia è rappresentata come un grafo indiretto (a) e la clique di dimensione massima è composta da 6 nodi (b).

Questa *clique* presenta degli elementi di interesse musicale. Per esempio, tre note appartengono ad un'ottava, mentre le altre tre ad un'altra, sono presenti intervalli di quinta perfetta o consonanze di ottava e altre caratteristiche tipiche della composizione e dell'armonia musicale. Le strutture sembrano contenere elementi di organizzazione generali o ad un livello profondo, che si manifestano con diverse occorrenze pratiche a livello della superficie melodica, realizzate da note diverse, che sfruttano però la stessa organizzazione.

6.7. Cicli e crescita delle reti musicali

Altre e più complesse organizzazioni, anche se meno evidenti, sono presenti nella rete. Tali strutture sono sequenze organizzate di frasi, composte dalle parole sopra individuate (motivi o grafi perfetti) che sono stati indicati col nome di *cicli*. I cicli possono essere considerati cammini all'interno della rete, per cui contengono una dinamica. Nell'accezione formale utilizzata in questa tesi, si parte da un nodo e si percorre la rete, con regole ben precise. Per esempio, si possono creare cicli, considerando un nodo alla volta, considerando un nodo ogni due, ogni tre e così via.

Analizzando la prima traccia della prima melodia della collezione *Art of Fugues (AOF)*, si consideri il ciclo che inizia con la nota $(4,0,A,1/8)$, valutando un nodo alla volta. Questo ciclo è composto dalle 29 note:

$(4,0,A,1/8)$ $(5,0,F,1/8)$ $(4,0,B,1/8)$ $(5,0,D,1/8)$ $(5,0,C,1/8)$ $(4,1,A,1/8)$ $(5,0,G,1/8)$ $(4,1,A,1/8)$ $(4,0,A,1/16)$ $(5,1,D,1/16)$ $(5,0,D,1/4)$ $(5,0,C,1/8)$ $(4,1,A,1/8)$ $(5,0,G,1/8)$ $(5,0,C,1/8)$ $(5,0,A,1/8)$ $(5,0,D,1/8)$ $(5,1,A,1/8)$ $(5,0,E,1/8)$ $(5,0,G,1/8)$ $(5,0,F,1/8)$ $(5,1,D,1/8)$ $(6,0,C,1/8)$ $(5,1,D,1/8)$ $(5,0,D,1/8)$ $(4,0,D,1/4)$ $(4,0,E,1/8)$ $(4,0,F,1/8)$ $(4,0,A,1/8)$

e si manifesta come un ciclo complesso all'interno della rete. Considerando la Figura 6.20a, questo ciclo a sua volta contiene un altro ciclo che inizia e finisce con la nota $(5,0,D,1/8)$, che a sua volta contiene un sotto-ciclo che inizia e finisce con la nota $(4,1,A,1/8)$. Si noti il carattere ricorsivo della melodia, che comprende cicli che a loro volta comprendono cicli che a loro volta comprendono cicli. Questi cicli si connettono gli uni con gli altri, creando la trama musicale. Le dinamiche si riflettono a livello di rete, dove blocchi di note individuano sotto-reti che rappresentano le parole e le frasi della grammatica musicale individuata. In questa sequenza, si osserva un'esplosione di simmetrie anche di natura frattale. Questo fenomeno evidenzia le infinite possibilità di scelta a disposizione del compositore, che può ricorrere a blocchi di questo tipo per effettuare infinite variazioni sul tema.

In generale, si osservi che, le note più ripetute all'interno di una melodia, sono anche quelle che presentano il grado più alto e quelle con valori di centralità maggiore. Nella Figura 6.22, si osserva questa forte correlazione tra grado delle note e numero di note in una melodia, configurazione che si mantiene inalterata anche se si considera il tempo delle note. Nella melodia considerata, la nota con grado più alto è $(4,0,A,1/8)$.

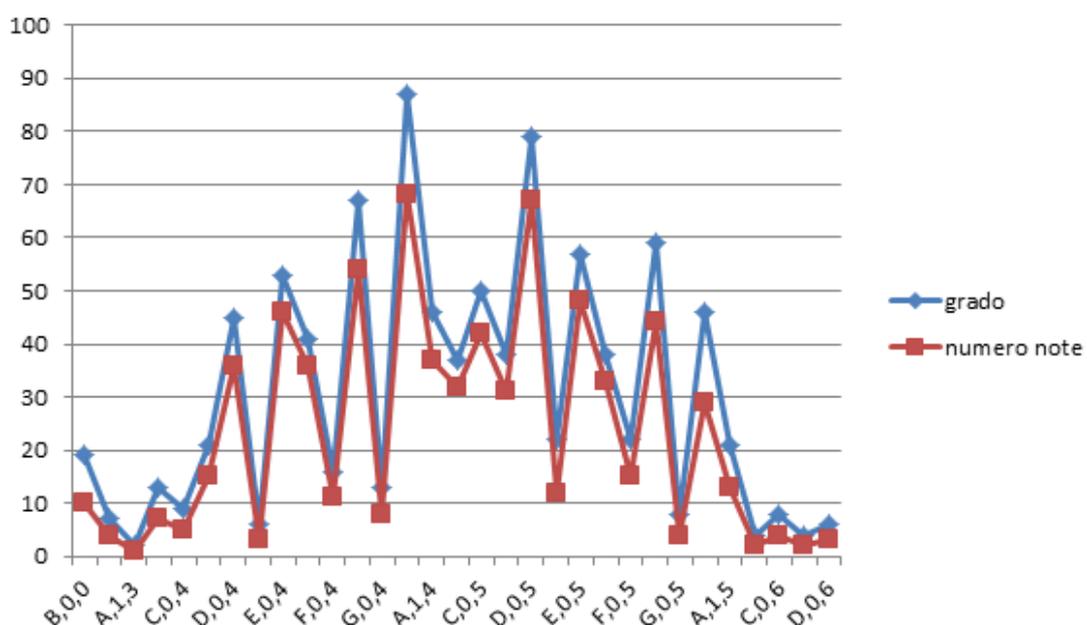


Figura 6.22. Correlazione tra numero di ripetizioni delle note e grado.

In un certo senso, la rete si sviluppa attorno a questa nota centrale (che è anche quella più ripetuta all'interno della melodia) e alle sue relazioni. Si può analizzare come avviene questa crescita attorno alla nota centrale. La rete della melodia analizzata possiede 105 nodi e 472 archi. La nota intrattiene numerosi legami con altre note nel suo intorno, espressi dall'in-Degree e dall' Out-Degree. L'intorno in questo caso è composto da 23 nodi, più la nota stessa (Figura 6.23).

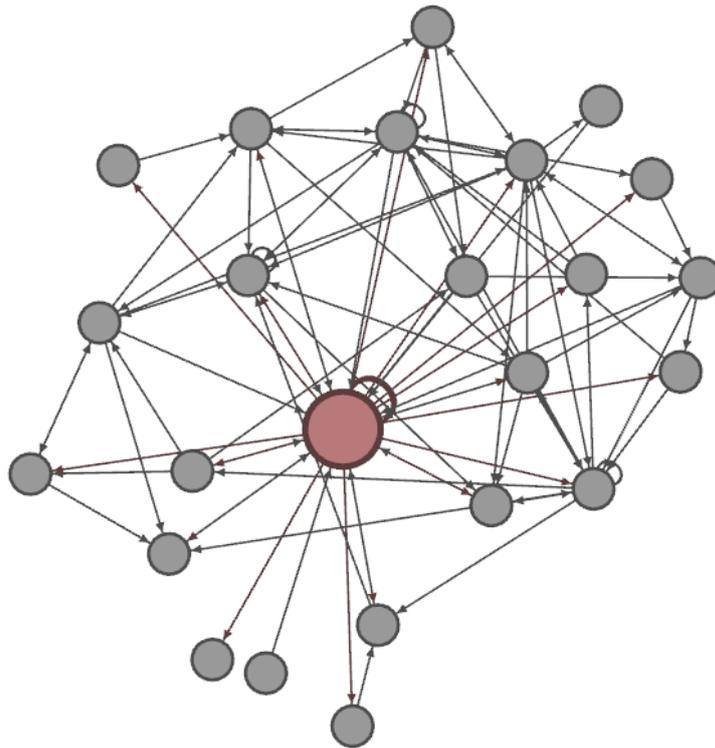


Figura 6.23. L'intorno della nota centrale(4,0,A,1/8), indicata in rosso.

Questa nota intrattiene con le note nel suo intorno una serie di relazioni che corrispondono a differenti motivi (Figura 6.24). Sono questi che caratterizzano i meccanismi di crescita della rete stessa. Nella Tabella 6.9, si riportano, a sinistra, i motivi matematici e, a destra, le occorrenze fisiche degli stessi.

<p>Motivi di 3</p>		
--------------------	--	--

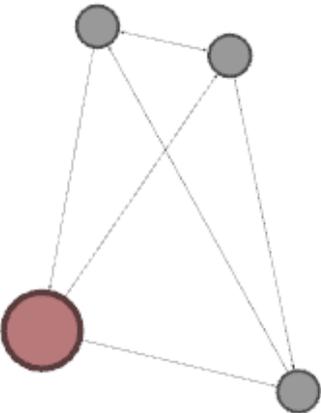
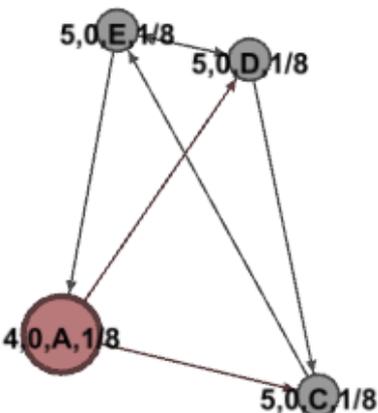
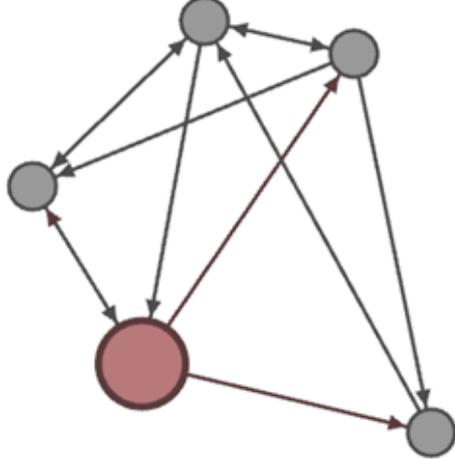
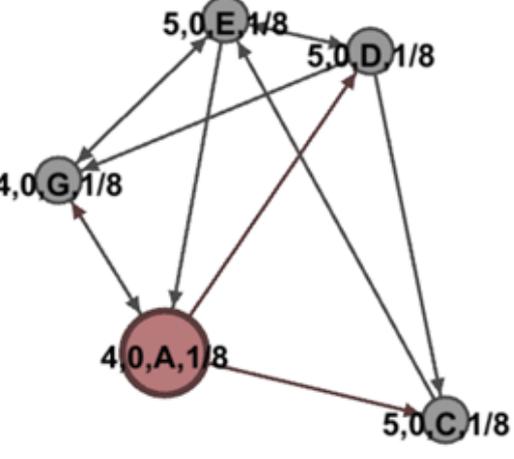
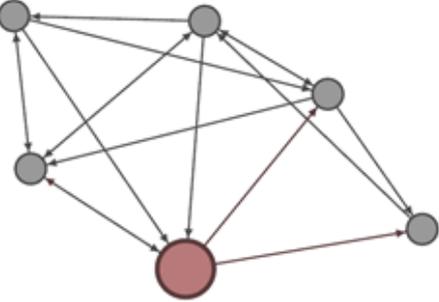
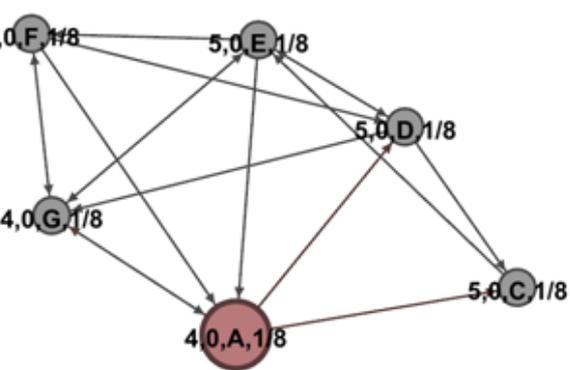
Motivi di 4		
Motivi di 5		
Motivi di 6		

Tabella 6.9. Motivi e crescita attorno alla nota centrale. A sinistra della Tabella sono riportati i motivi astratti, a destra i motivi che incorporano le occorrenze fisiche delle note musicali nella rete.

Se si continuano a indagare le dinamiche di crescita, ci si accorge che il secondo intorno della nota centrale aumenta enormemente la dimensione della rete. Infatti, il secondo intorno (composto dagli intorni delle note del primo intorno) presenta 80 nodi e 415 archi. Il diametro della rete del primo intorno è 2 (2 archi al

massimo tra un nodo e l'altro). Invece, il diametro del secondo intorno è 4 (al più 4 archi tra un nodo e l'altro). Il diametro della rete del terzo intorno è 6, con 102 nodi, e 8 per la rete finale.

Riassumendo, la crescita della rete avviene in vari passi. Il primo prevede un nucleo abbastanza rilevante di note, che fanno da corona alla nota centrale. Il secondo passo amplia notevolmente la rete. Con il terzo passo la crescita è molto rallentata. Alla fine, la rete si conclude con poche note periferiche, che poche connessioni hanno con il cuore della melodia. Le pause, considerate nella formalizzazione usata come note, hanno anche un ruolo rilevante, per numero di ripetizioni.

6.8. Conclusioni

Si rilevano per tutti i livelli di analisi i caratteristici raggruppamenti ritrovati per altri contesti di studio, con strutture di piccolo mondo che si organizzano nelle reti musicali e organizzazioni a invarianza di scala, che intercettano leggi di potenza troncate. Si dimostra, la presenza di *clique*, che, come nella vita sociale o in altre strutture organizzate, si coordinano a livello delle singole note, a livello di organizzazioni più complesse, *clique* che si uniscono ad altre *clique* per formare intere aree delle reti musicali probabilmente specializzate, o condivise con altre reti e altre collezioni. Si individuano anche modelli di crescita generativa che sono di natura compositiva, come avviene, come vedremo successivamente per le reti cerebrali e per le reti linguistiche.

Capitolo 7. Reti cerebrali in soggetti sani e con malattie neurodegenerative

7.1. Introduzione

Le tecniche di Brain Imaging consentono attualmente di lavorare su una base di dati innovativi per ampliare la comprensione dei meccanismi del cervello umano, sia in soggetti normali che portatori di patologie neurodegenerative o con tumori cerebrali. Attraverso lo studio di tali dati è possibile costruire modelli di funzionamento del cervello. Infatti, i sistemi di simulazione del cervello, sia per il funzionamento normale che patologico, sono utili per progettare sistemi artificiali, e analizzare a fondo le cause del deterioramento neurodegenerativo tipico di diverse malattie invalidanti quali l'Alzheimer o il Parkinson e i Parkinsonismi. Anche le dinamiche neuronali cerebrali, in soggetti sani e malati, possono essere strumento di investigazione per capire i fondamentali del funzionamento cerebrale. Le dinamiche cerebrali ci permettono inoltre di analizzare come si trasforma e viaggia l'informazione all'interno del cervello umano. In questo capitolo, si riporta un esperimento che applica la modellazione delle reti cerebrali, per mezzo della Connettistica (Bullmore & Sporns, 2009; Bullmore & Bassett, 2011) in soggetti sani e soggetti affetti da Parkinson (Quattrone et al., 2008; Nilsson et al., 2007). Infatti, per aumentare la precisione e la sensibilità della diagnosi di paralisi sopranucleare progressiva (PSP, sindrome di Steele-Richardson-Olszewski), abbiamo usato algoritmi di ricostruzione di fibre di aree di interesse per questa malattia e la teoria delle reti (Sporns, 2011) per analizzare le proprietà strutturali delle reti cerebrali in soggetti sani e soggetti con PSP. Questo ci ha permesso di definire statistiche di base (Barabási & Albert, 1999) delle reti cerebrali dei soggetti per identificare questo tipo di malattia degenerativa. I risultati mostrano che i parametri di piccolo-mondo e scale-free sono presenti sia in reti di soggetti affetti da questo tipo di Parkinsonismo che nelle reti dei soggetti sano, del gruppo di controllo.

7.2. Background di riferimento sul Parkinson e il Parkinsonismi

La Paralisi Sopranucleare Progressiva (PSP) (o malattia di Steele-Richardson-Olszewski), e l'Atrofia Multi-Sistemica (MSA) differiscono dal Morbo di Parkinson (PD,) in molte manifestazioni delle disfunzioni neurologiche, che determinano una degenerazione cerebellare piramidale, vegetativa, e cognitiva.

Entrambe le malattie si sviluppano a un ritmo sempre più marcato e sono meno sensibili ai farmaci rispetto al PD. La MSA presenta segni simili al Parkinson, provocati da un peggioramento del sistema piramidale nigro-striato, che coinvolge il cervelletto e le strutture cerebrali profonde. Nella MSA si rileva anche la degenerazione neuronale del neo-striato, area cerebrale che rimane invariata nel PD.

La Paralisi Sopranucleare Progressiva mostra anche sintomi di tipo parkinsoniano, soprattutto a causa della rigidità assiale che alcuni soggetti presentano e l'assenza di tremore. Si manifestano, inoltre, precoce instabilità posturale, con frequenti cadute, disturbi della motilità oculare, con l'incapacità iniziale a muovere gli occhi verso l'alto o verso il basso, e poi da sinistra a destra, comparsa di segni pseudo-bulbari. Farmaci anti Parkinson hanno poco effetto su soggetti con PSP. Per questa malattia la diagnosi si basa essenzialmente su tre criteri clinici, che la definiscono come la più possibile, la più probabile o la più definita. Una diagnosi accurata e precoce di PSP è importante, sia per scopi terapeutici - indirizzando la terapia più precisamente sui vari sintomi - sia in termini di prognosi, che è molto diversa dalla prognosi per malati di PD. Tuttavia, anche se avanzate tecniche diagnostiche sono state recentemente sviluppate per il PD (Oba et al., 2005; Quattrone et al., 2008; Cosottini et al., 2007), la PSP risente di una mancanza di criteri diagnostici universalmente accettati, che renda semplice distinguere PSP dal PD e da MSA. La pratica medica per questa specifica degenerazione cerebrale è caratterizzato da un alto tasso di diagnosi errate (Litvan et al., 1996). Infatti le differenze nei sintomi tra PSP e MSA e il PD sono estremamente labili (Hughes et al., 1992, 2001).

La tecnica che usa Immagini di Risonanza Magnetica Strutturale (MRI) è solitamente usata per rilevare i primi segni di PD, PSP e MSA. MSA è rilevata da una iper-intensità del bordo laterale del putamen e da un'atrofia del mesencefalo. Una iper-intensità a croce del ponte e del peduncolo cerebrale medio è un indicatore per tale malattia. La T2 * pesata assiale è invece utilizzata per misurare la disposizione dei gangli basali. Schocke e collaboratori (Schocke et al., 2002, 2004; Seppi et al., 2003) utilizzano il coefficiente di diffusione (RADC) per differenziare la regione del putamen nei pazienti con PD, MSA-P (una forma di Atrofia Multi Sistemica, con Parkinsonismo predominante) e soggetti di controllo. La tecnica delle Risonanza Magnetica Diffusiva, spiegata nel Capitolo 4, è usata per analizzare sia il putamen anteriore e posteriore per PSP e MSA-P (Seppi et al. 2006). Nicoletti e collaboratori hanno utilizzato (i) la risonanza magnetica (MRI) per determinare le misure del peduncolo cerebellare medio e (ii) la Risonanza Magnetica (MRI) Diffusiva per stabilire i coefficienti di diffusione apparente (rADC) del Peduncolo cerebellare medio (MCP), del Caudato, del Putamen e del Pallidum come strumento di differenziazione tra PD, MSA (Nicoletti et al., 2006 a e b; Nicoletti et al., 2008). Altri ricercatori hanno usato la Trattografia per mostrare le differenze nei peduncoli cerebellari medi e superiori (rispettivamente Middle Cerebellar Peduncle, MCP e Superior Cerebellar Peduncle, SCP) in pazienti con PD e MSA. Una MRI Volumetrica è stata utilizzata con lo stesso scopo attraverso tecniche di MRI (Oba et al. 2005). Tale tecnica ha permesso l'inizio di una serie di studi che

hanno portato alla creazione dell'Indice di Quattrone (Quattrone et al., 2008). L'indice è stato poi trasformato in un software automatico. Non è ancora disponibile al pubblico, ed è attualmente in fase di validazione sperimentale. C'è ancora tuttavia il potenziale per lo sviluppo di nuovi strumenti per sostenere la diagnosi medica per soggetti portatori di PSP. Tra le tecniche più utilizzate vi è la MRI Diffusiva. Questa tecnica consente di verificare l'orientamento della materia grigia e bianca, utilizzando il moto browniano delle molecole di acqua, nel tessuto interno, fornendo le informazioni essenziali circa le dimensioni e l'orientamento dei compartimenti del cervello, per gruppi di voxel presenti nei file ottenuti dalla RMI. Negli ultimi decenni, da quando è stato impiegato come biomarker per la rilevazione di tratti (Moseley et al., 1990), la MRI Diffusiva è stata applicata intensamente. Dopo questo passo, sono state identificate le fibre della sostanza bianca nel cervello umano (Conturo et al., 1999; Mori et al., 1999; Wedeen, 1996), sfruttando le sue proprietà di anisotropia (Basser et al., 1994; Douek et al., 1991). La MRI Diffusiva è diventata un metodo efficace per definire l'architettura connettività anatomica delle fibre nel cervello. Una dettagliata descrizione strutturale e funzionale del pattern di connettività del cervello umano (il concetto di 'Connectome') è stato recentemente proposto da Sporns et al. (2005). Da allora, diversi studi hanno impiegato metodi neuro-imaging e approcci grafici all'avanguardia, per esplorare il cervello umano in soggetti sani e malati, scoprendo reti di piccolo mondo (Small World Networks o SWN) in differenti regioni del cervello (Bullmore & Sporns, 2009; Stam, 2009; Sporns, 2011; Sporns & Kötter, 2004).

In particolare, nei pazienti con malattia di Alzheimer (AD) (Stam & Reijneveld, 2007; Stam et al., 2007; Supekar et al., 2008; de Haan et al., 2009; Liu et al., 2008), e schizofrenia (Rubinov & Sporns, 2010; Rubinov et al., 2009), questi approcci hanno rilevato disturbi nella tipica organizzazione del cervello, comprese quelle di modifica in topologia delle reti di small-world.

In questo capitolo sono analizzati 13 soggetti sani, 13 con diagnosi di PSP e di 13 soggetti di controllo, attraverso i passi metodologici, riportati nel Capitolo 4, che vanno dalla MRI Diffusiva alla Connettistica. Il nostro scopo è di investigare se sia possibile identificare nelle reti cerebrali che saranno create una specifica organizzazione di rete o topologia che identifichi in modo preciso la PSP. Riteniamo che l'uso dei parametri statistici della teoria delle reti per supportare la diagnosi, sia un ottimo strumento per aiutare a differenziare gli uni dagli altri PSP, PD e MSA, varianti del Parkinson. Il concetto di reti di piccolo mondo applicato alla vasta architettura neurale del cervello dipinge un quadro di un sistema complesso in cui l'informazione viaggia da nodi e bordi, in tutte le direzioni possibili in pochi passi. Il coefficiente di clustering (Watts & Strogatz, 1998) misura la compattezza di un gruppo di nodi all'interno della rete cervello. Alcuni ricercatori hanno presentato prove del carattere funzionale dei clustering nel cervello di soggetti umani (Sporns & Kötter, 2004; Milo et al., 2002; Milo et al., 2004 a, b) e di animali (Hilgetag et al., 2000).

Altre considerazioni importanti riguardano come le reti del cervello sono collegate le une alle altre, quale sia la loro organizzazione interna sia nei soggetti sani (Amaral et al., 2000; Barrat et al., 2004) sia nei

soggetti malati (Stam et al., 2007; Kubicki et al., 2003), e il “grado di distribuzione” di ogni nodo. Studi nell'organizzazione del cervello si focalizzano su come il cervello realizza la sua straordinaria architettura nello sviluppo e sulle modalità di crescita attraverso 'attaccamento preferenziale', ovvero come cresce la rete nel tempo. Si è rilevato che più di un nodo ha legami, tanto più ne otterrà di nuovi (Eguiluz et al., 2005).

Il metodo di lavoro che abbiamo adottato procede come segue:

- a. Nel cervello umano, la connettività assonale a lungo raggio forma una rete small-world, in soggetti sani, di controllo e in soggetti con diagnosi di PSP.
- b. C'è un rapporto cruciale tra i parametri delle reti di piccolo mondo, scale-free e la presenza o assenza di sintomi PSP.
- c. La teoria delle reti può essere utilizzata per indagare il malfunzionamento di regioni del cervello, contribuendo così a definire l'esatta topologia (o architettura) neurale di questa malattia così difficile da diagnosticare (motivi neurali Sporns et al., 2005).

Lo studio delle funzioni del cervello e delle strutture attraverso le reti rivela la natura elusiva di questo organo fondamentale. Tre criteri fondamentali - piccolo del mondo delle reti (Horowitz, 2003; Bullmore & Bassett, 2011; He et al., 2007; Bullmore & Sporns 2009; Sporns, 2011), di clustering (Sporns & Kötter, 2004; Milo et al., 2004a e b; Milo et al., 2002) e il grado di distribuzione (Bullmore & Sporns, 2009) determinano le tecniche diagnostiche proposte.

7.3. Materiali e metodi

In questo capitolo, immagini di Risonanza Magnetica di 26 adulti - 13 soggetti sani, con un'età media di 67 anni, 13 soggetti con diagnosi di PSP, con un'età media di 70 anni, sono state analizzate. I file T1 e TDI, di solito ottenuti per diffusione MRI, sono stati catturati alla risoluzione di 3 Tesla, su uno scanner GE, con una pesata in diffusione, singola sequenza di scatto EPI, e con i seguenti parametri di distribuzione: TR / TE / Δ / δ = 3000/100/47.6/35 ms e b-max = 12000 mm² / s. La variabile Q-space è stata campionata e i dati ricostruiti, secondo uno schema standard DSI. Il blocco di acquisizione è stato effettuato su 32 fette di una matrice 128 × 128, con una risoluzione spaziale di 2 × 2 × 3 mm³.

Come specificato nel Capitolo 4, i passi per ottenere reti da file DICOM sono:

- a. Identificazione del tensore di diffusione per individuare le principali traiettorie delle fibre;
- b. Tracciamento delle fibre.
- c. Determinazione del tensore di diffusione, attraverso cui si risale alla presenza e all'orientazione delle fibre.
- g. Tracciamento delle fibre a partire dalla fibre stesse per mezzo della Diffusion Spectrum Imaging

7.4. Risultati

Le variabili statistiche che abbiamo usato, come nel caso della musica, sono il numero di nodi e il numero di connessioni, il grado, la percentuale di clustering, l'efficienza, il cammino medio, il diametro e la densità, i parametri che identificano le reti di piccolo mondo (SWN) e infine i parametri che identifica l'altro modello di organizzazione delle reti Scale-Free (SF).

I valori dei parametri generali associati con la rete di ciascun soggetto sono riportate in Tabella 7.1.

Gruppo PSP	Nodi	Archi	Grado	Clustering	Efficienza	Cammino	Diametro	Densità
1.net	83	373	8,987952	0,54835	0,421476	2,865438	6	0,109609
2.net	83	428	10,31325	0,581537	0,458301	2,569894	6	0,125771
3.net	83	353	8,506024	0,564347	0,406832	2,9955	6	0,103732
4.net	83	384	9,253012	0,553441	0,428019	2,818987	6	0,112842
5.net	83	411	9,903614	0,572189	0,450328	2,622442	5	0,120776
6.net	83	379	9,13253	0,538931	0,424772	2,853825	6	0,111372
7.net	83	335	8,072289	0,530452	0,407293	2,957759	6	0,098443
8.net	83	331	7,975904	0,536019	0,395855	3,135433	7	0,097267
9.net	83	389	9,373494	0,557276	0,439823	2,690957	6	0,114311
10.net	83	382	9,204819	0,580165	0,432197	2,772246	6	0,112254
11.net	83	385	9,277108	0,56022	0,433539	2,759472	5	0,113135
12.net	83	409	9,855422	0,559062	0,446821	2,652054	5	0,120188
13.net	83	338	8,144578	0,559667	0,39537	3,143272	7	0,099324
Media		376,6923	9,076923	0,55705	0,426202	2,833637	5,923077	0,110694
Gruppo Controllo	Nodi	Archi	Grado	Clustering	Efficienza	Cammino	Diametro	Densità
1.net	83	396	9,542169	0,575184	0,44402	2,671796	6	0,116368
2.net	83	409	9,855422	0,566338	0,444108	2,709537	6	0,120188
3.net	83	383	9,228916	0,563234	0,437398	2,715634	5	0,112548
4.net	83	386	9,301205	0,570816	0,434621	2,742053	6	0,113429
5.net	83	369	8,891566	0,5376	0,399104	3,166788	6	0,108434
6.net	83	406	9,783133	0,554876	0,442972	2,688053	6	0,119306
7.net	83	402	9,686747	0,563631	0,433936	2,777181	6	0,118131
8.net	83	416	10,0241	0,5904	0,458385	2,560894	5	0,122245
9.net	83	398	9,590361	0,548903	0,444461	2,65786	6	0,116956
10.net	83	462	11,13253	0,572486	0,472598	2,491799	5	0,135763
11.net	83	451	10,86747	0,563489	0,469899	2,483379	5	0,13253
12.net	83	411	9,903614	0,565614	0,444975	2,680795	6	0,120776
13.net	83	386	9,301205	0,571485	0,433838	2,753665	6	0,113429
Media		405,7692	9,777572	0,564927	0,443101	2,699956	5,692308	0,119239

Tabella 7.1. Valori medi dei parametri associati con la rete di ogni soggetto, per soggetti sani (controllo) e soggetti con PSP.

I valori dei parametri per il modello Reti di Piccolo Mondo (SWN) sono riportati in Tabella 7.2.

Gruppo controllo				PSP			
Nome rete	Gamma	Lambda	Sigma	Nome rete	Gamma	Lambda	Sigma
1.net	3,757481	1,235117	3,042208	1.net	4,027397	1,292634	3,115651
2.net	3,663564	1,271327	2,881684	2.net	3,458769	1,220173	2,834654
3.net	3,927514	1,244619	3,155594	3.net	4,713772	1,330668	3,54241
4.net	4,125681	1,258897	3,277219	4.net	4,08095	1,293595	3,154735
5.net	4,180121	1,426612	2,930103	5.net	3,694858	1,23073	3,002167
6.net	3,761297	1,25822	2,989379	6.net	3,897556	1,299553	2,999153
7.net	3,919003	1,291601	3,034221	7.net	4,695569	1,287281	3,647665
8.net	3,458437	1,204592	2,871044	8.net	4,553262	1,351252	3,369662
9.net	3,815632	1,232924	3,094783	9.net	3,853619	1,239689	3,108536
10.net	3,303899	1,216425	2,716072	10.net	3,949089	1,267082	3,11668
11.net	3,526427	1,203782	2,929457	11.net	3,909597	1,265133	3,090266
12.net	3,687091	1,258021	2,930867	12.net	3,661654	1,235029	2,964833
13.net	3,924839	1,257808	3,120381	13.net	4,636227	1,367842	3,389447

Tabella 7.2. Parametri di Piccolo Mondo (SWN) per soggetti sani e per soggetti con PSP.

I valori dei parametri per il modello relativo alla legge di potenza troncata sono riportati in Tabella 7.3.

Gruppo Controllo					PSP				
Nome rete	a	kc	b	R2	Nome rete	a	kc	b	R2
1.net	0,948961	3,889867	0,76709	0,994024	1.net	1,176843	4,362313	0,511561	0,981741
2.net	0,741364	3,391218	1,061928	0,995217	2.net	0,948576	4,381112	0,692586	0,995106
3.net	0,736998	3,054025	1,151136	0,997617	3.net	0,942694	3,121966	0,956207	0,992132
4.net	0,75593	3,172369	1,101689	0,995694	4.net	0,716069	2,952449	1,217864	0,998261
5.net	0,95198	3,496304	0,847032	0,991071	5.net	0,799284	3,611384	0,95396	0,992893
6.net	0,704016	3,186826	1,162	0,996255	6.net	0,93609	3,564091	0,846382	0,990678
7.net	0,921169	3,755358	0,833086	0,986665	7.net	0,952401	2,858194	1,023375	0,991826
8.net	0,984795	4,380055	0,653878	0,995322	8.net	0,998059	3,176549	0,857689	0,990457
9.net	0,884087	3,532377	0,915224	0,987627	9.net	0,890465	3,222618	1,002423	0,992873
10.net	0,739629	4,046501	0,943534	0,997452	10.net	0,91423	3,605311	0,839751	0,996364
11.net	0,676687	3,734783	1,063674	0,990449	11.net	0,934575	3,524314	0,863011	0,995436
12.net	0,770893	3,538776	0,997718	0,99804	12.net	0,928384	4,154485	0,736406	0,987189
13.net	1,205302	4,460641	0,496622	0,987986	13.net	0,994751	3,28809	0,833454	0,988114

Tabella 7.3. I valori dei parametri identificano una potenza troncata, invece che una legge di potenza, tipica per l'andamento Scale Free. I valori sono per ogni rete, per il gruppo di controllo e il gruppo PSP (Bertacchini et al., sottomesso, b).

7.5. Analisi dei risultati

Se guardiamo in dettaglio la distribuzione dei valori della Tabella 7.1, vediamo che i soggetti sani hanno un numero di archi in genere superiore a 400. Per il soggetto 1, questo valore è leggermente inferiore a 400 (396) e per il soggetto 5 il valore scende a 369 (Figura 7.1a). Tale andamento riflette il grado medio - sempre sopra 0,91 - ad eccezione del soggetto 5. La densità della rete, salvo nel caso del soggetto 5, non

scende sotto 0,110. Questa tendenza è completamente opposta nel caso di pazienti PSP, per i quali il numero di collegamenti è sempre sotto 400, anche con l'improvvisa riduzione a 335, 331 e 338, come nel caso di soggetti 7, 8 e 13 rispettivamente. Questo fenomeno si ripete anche per i valori di densità e grado delle reti nel confronto fra gruppo di controllo e PSP.

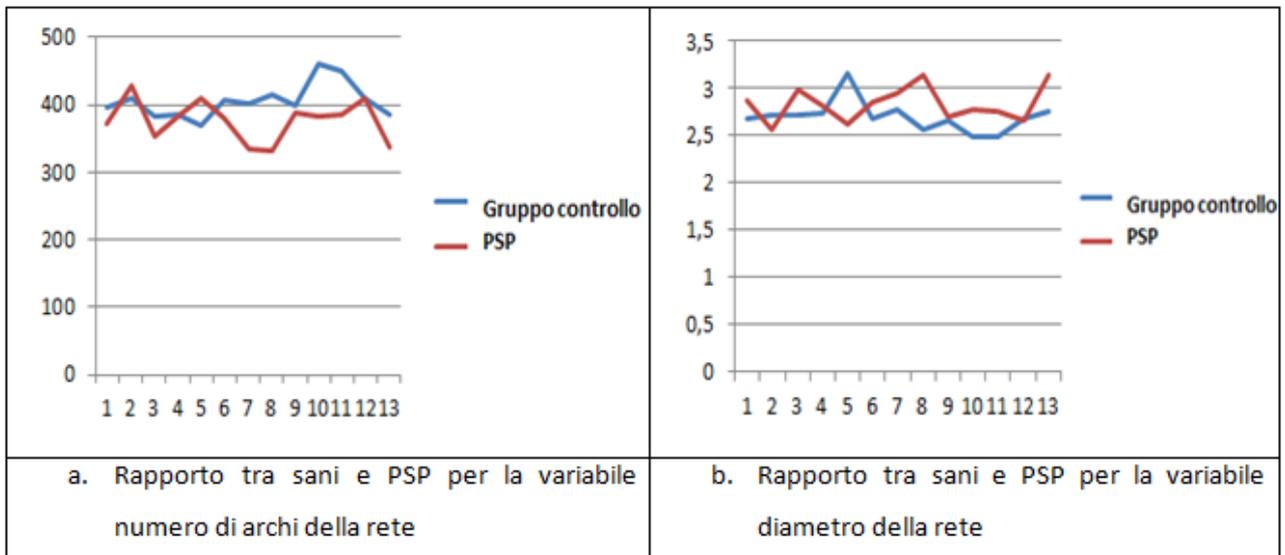


Figura 7.1. Grado di distribuzione e percorso medio nei 13 soggetti sani e nei 13 pazienti con PSP.

La tendenza si inverte quando si tratta di considerare il diametro, il cammino medio e l'efficienza della rete nei due gruppi. Ad esempio, la Figura 7.1b mostra i valori per la lunghezza media, dai quali si osserva che, in generale, il cammino medio è maggiore nei soggetti affetti da PSP rispetto a quelli sani, e che vi è una corrispondente diminuzione di efficienza. Nella Figura 7.2, infatti, si possono notare gli andamenti dei valori di clustering per gruppo di controllo vs. gruppo PSP e i valori di efficacia della rete, sempre confrontando gruppo di controllo vs. gruppo PSP.

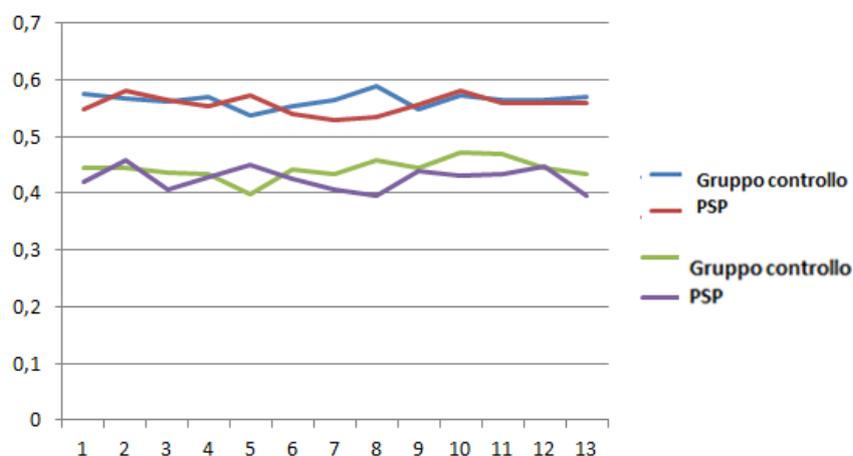


Figura 7.2. Confronto fra i valori di Clustering e efficacia della rete tra gruppo di controllo e gruppo PSP.

Mentre da un lato, come si può vedere, l'efficienza è in media superiore in soggetti sani, dall'altro non vi è alcuna differenza significativa nei valori di clustering fra i due gruppi. Tuttavia, tale differenza emerge se si considerano i parametri di SWN (Tabella 7.2).

In questo caso, si può osservare che tutte le reti presentano organizzazioni tipiche delle reti di piccolo-mondo (SWN), e che questa tendenza è più pronunciata in soggetti con PSP, come risulta dalla Figura 7.3. In questo gruppo di soggetti, se l'efficienza complessiva e il numero di connessioni diminuisce, si verifica un fenomeno opposto la creazione di reti locali di connessione diffuse in tutta la rete (Figura 7.3).

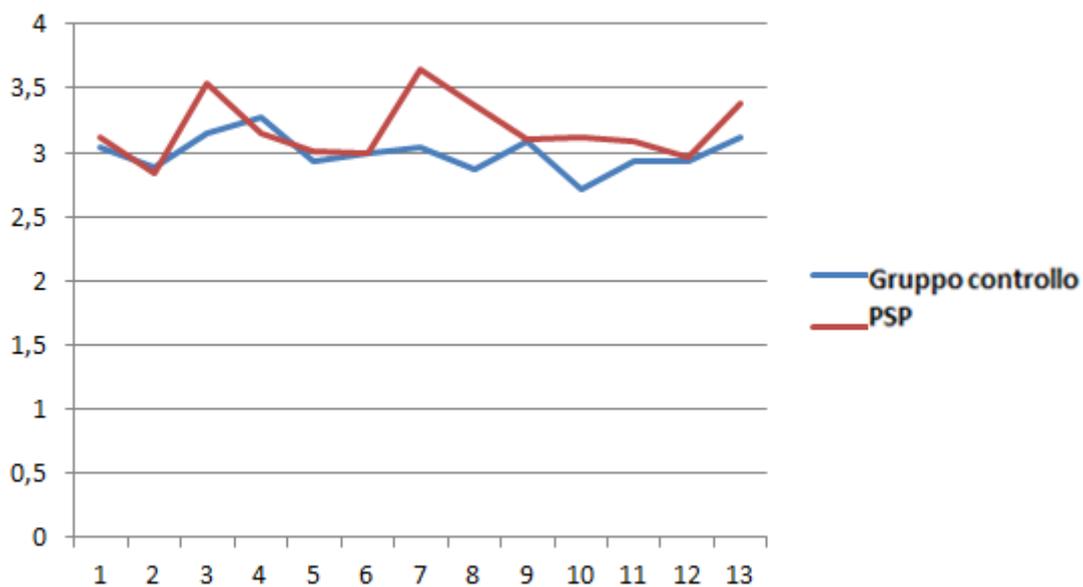


Figura 7.3, I valori di σ nei due gruppi. Le caratteristiche di SW si accentuano per i soggetti con PSP.

I valori dei parametri relativi alla legge di potenza troncata $P(k) = ae^{-\frac{k}{kc}} k^b$, alla legge di potenza $P(k) = ak^b$ e alla legge esponenziale $P(k) = ae^{-\frac{k}{kc}}$, calcolati per collezioni di reti dei due gruppi sono presentati nella Tabella 7.3. Indagando sulla distribuzione del grado, troviamo che, anche con pochi dati osservabili, le reti presentano un comportamento di legge di potenza troncata, confermando quanto già noto in letteratura.

Sembra che non ci siano variazioni significative tra la legge di potenza troncata e la legge di potenza esponenziale, ad eccezione del parametro kc , che è leggermente più grande in soggetti sani, rispetto ai soggetti con PSP (Figura 7.4). Questo indica che, per lo stesso grado, il numero di nodi è maggiore in soggetti sani, rispetto a quelli con PSP.

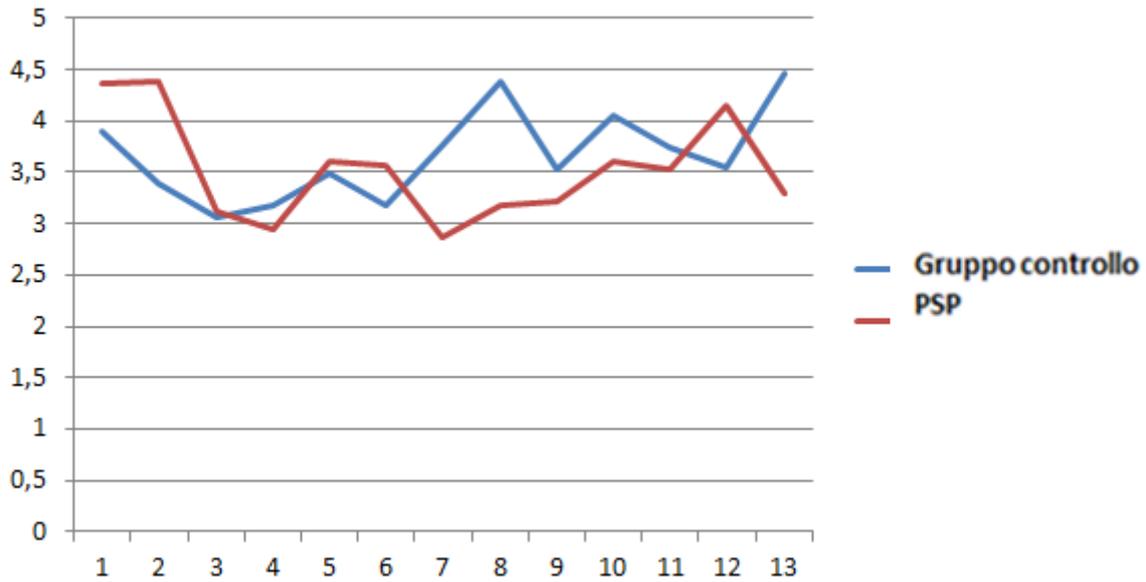


Figura 7.4. Valori di kc rispetto soggetti sani e soggetti con PSP.

7.5.1. Differenze statistiche fra i gruppi

Anche se nella letteratura citata è impossibile identificare un criterio che può sicuramente distinguere soggetti sani da soggetti malati, alcuni dati emergono, come si può notare confrontando i valori riportati nelle Tabelle 7.4, 7.5 e 7.6.

Gruppo	Archi	Grado	Clustering	Efficienza	Cammino	Diametro	Densità
PSP	376,6923	9,076923	0,55705	0,426202	2,833637	5,923077	0,110694
Controllo	405,7692	9,777572	0,564927	0,443101	2,699956	5,692308	0,119239

Tabella 7.4. Valori medi dei parametri generali di rete per i due gruppi.

Come è evidente dalla Tabella 7.4, il grado medio diminuisce significativamente nei pazienti con PSP rispetto ai soggetti sani; questo significa anche una diminuzione della densità di connessione nella rete. Un significativo aumento del diametro e della lunghezza del percorso medio con una diminuita efficienza sono tuttavia anche presenti, a testimonianza del fatto che il diametro aumenta, essendo le reti più sparse. Questi parametri sono misure significative della struttura a piccolo mondo delle reti di soggetti PSP. La Tabella 5 mostra che i valori di σ crescono in soggetti con PSP, indicando che la quantità di connessioni locali tende a crescere rispetto a quelle più distanti.

Reti Piccolo Mondo (SWN)	Gamma	Lambda	Sigma
PSP	4,087101	1,283128	3,179681
54	3,784975	1,255862	3,012501
67	3,773153	1,258457	2,997924

Tabella 7.5. Valori per I parametri di Piccolo Mondo (SWN).

Questo non comporta un aumento, bensì una diminuzione delle prestazioni complessive. Infatti, all'aumentare del diametro, così come il percorso medio, il numero totale di collegamenti non solo diminuisce, ma anche questi tendono a limitare l'accesso in molti luoghi della rete.

Per quanto riguarda i parametri di invarianza di scala delle reti dei due gruppi, il valore di kc per la legge di potenza è troncato (3,66), sia in soggetti sani che nei soggetti PSP (3,52). Una possibile interpretazione di questo fenomeno è che il numero di nodi è maggiore nei soggetti con PSP rispetto ai pazienti sani. Ciò significa che il grado decresce più lentamente in soggetti sani rispetto ai pazienti PSP. Questo processo è una conferma indiretta della diminuzione del grado complessivo, con una controparte nella legge di potenza. Il risultato è particolarmente significativo se si considera la frequenza di tutti i nodi per un dato livello di tutte le reti delle due categorie (Tabella 7.6). In questo caso, il valore di kc è molto maggiore nei pazienti con PSP, rispetto ai soggetti sani.

Nome gruppo	a	kc	b	R2	a	b	R2	a	kc	R2
Controllo	1,146838	4,509749	0,538977	0,988618	1,384229	-0,57077	0,636616	1,3990268	8,293854	0,925614
PSP	1,177249	4,174074	0,551681	0,989273	1,369155	-0,59208	0,646443	1,395524	7,682827	0,928372

Tabella 7.6. Confronto dei parametri di invarianza di scala per i due gruppi.

In Figura 7.8, queste distribuzioni sono rappresentate graficamente.

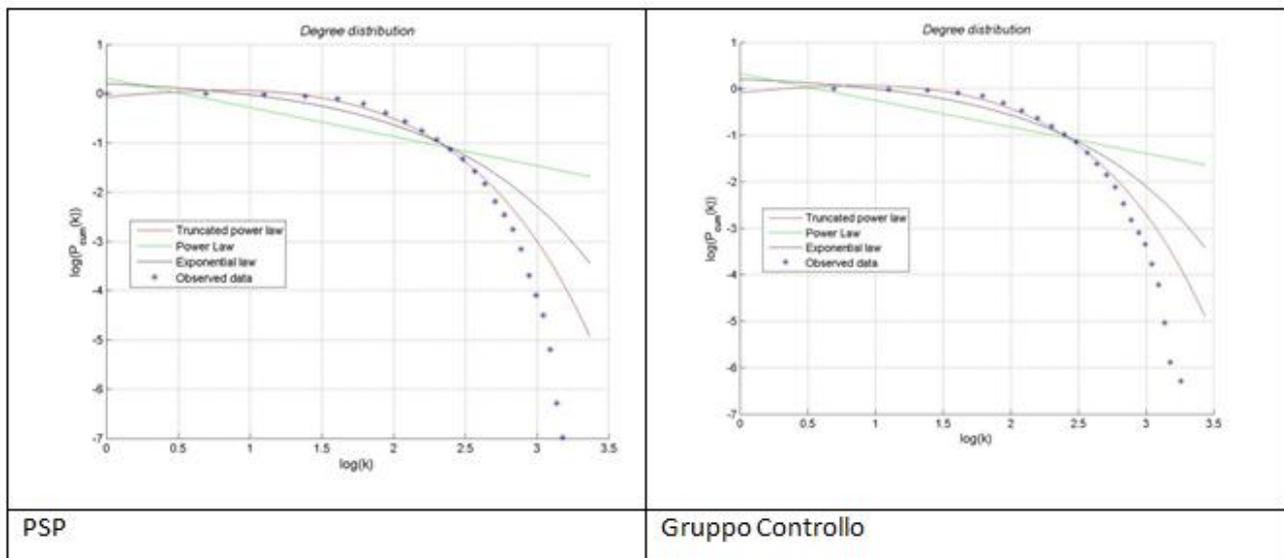


Figura 7.8. Le leggi di potenza troncate nel gruppo PSP e nel gruppo di controllo.

7.5 .2. Il grado (k) come strumento di analisi della degenerazione neurale

Il ponte, il mesencefalo e il peduncolo cerebellare superiore sono le ROI di riferimento per una diagnosi di PSP, in quanto sono proprio queste strutture che si degenerano quando la malattia compare. Tuttavia, non sempre una loro valutazione è sufficiente per diagnosticare in modo affidabile questa malattia. Esplorazioni più dettagliate sono necessarie. In questa sezione, presentiamo diverse modalità di discriminazione per determinare i volumi o le dimensioni delle ROI di riferimento, nelle reti cerebrali di soggetti affetti da PSP e soggetti sani del gruppo di controllo, utilizzando il parametro k . Si intende infatti sviluppare una tecnica più sofisticata come strumento di supporto per la diagnosi di PSP. L'analisi riguarda sia le caratteristiche globali che locali delle reti, e di conseguenza, per analogia, che gli effetti della neurodegenerazione determina nella struttura cerebrale. Come è evidente dalla Figura 7.9 (a, b), altre aree del cervello sono coinvolte nella PSP, se si confrontano i dati in soggetti sani e malati.

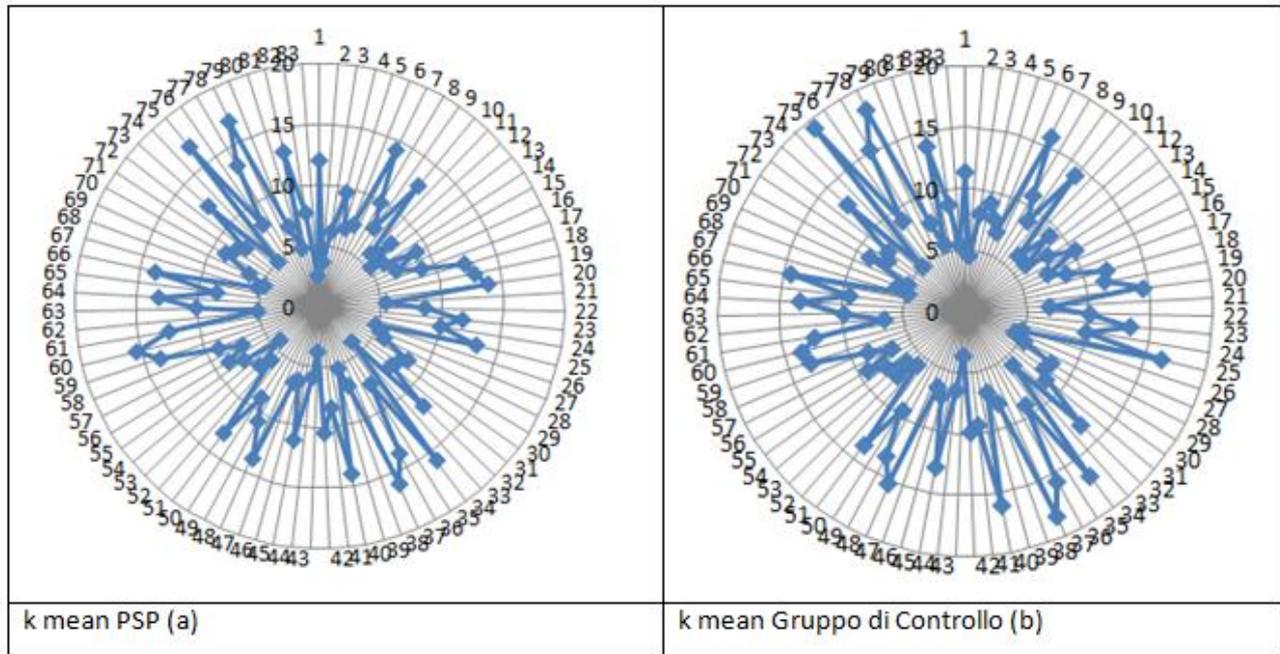


Figura 7.9. (a) k mean per soggetti con PSP; (b) k mean per soggetti sani, gruppo di controllo.

Sul diagramma radiale, sono rappresentate le 83 diverse aree del cervello. In contrasto con la tendenza generale di altre malattie neurodegenerative, che mostrano una progressiva diminuzione del numero di neuroni, la Tabella 7.6, che illustra altre aree del cervello potenzialmente non analizzate o non considerate nella diagnosi di PSP, mostra non solo un aumento considerevole delle aree coinvolte nella patologia considerata, ma addirittura, in alcune aree, si verifica un aumento di cellule neurali in soggetti con PSP, rispetto ai soggetti sani.

Numero nodo	Nome Cerebrale Area	k mean per soggetti sani (gruppo controllo)	k mean per soggetti con PSP	Deviazione standard Gruppo controllo/PSP
1	ctx-rh-lateralorbitofrontal	11,38461538	12	0,435142638
4	ctx-rh-medialorbitofrontal	9,076923077	9,538461538	0,326356976
6	ctx-rh-parsopercularis	6,923076923	7,230769231	0,217571318
10	ctx-rh-precentral	14,23076923	12,69230769	1,087856588
17	ctx-rh-supramarginal	8,692307692	8,846153846	0,108785659
18	ctx-rh-superiorparietal	11,92307692	12,23076923	0,217571319

24	ctx-rh-lingual	9,923076923	10	0,054392829
27	ctx-rh-entorhinal	4,384615385	5,307692308	0,652713952
28	ctx-rh-temporalpole	5,384615385	5,923076923	0,380749805
29	ctx-rh- inferiortemporal	8,153846154	8,461538462	0,217571318
42	ctx-lh- lateralorbitofrontal	9,846153846	10,46153846	0,435142634
43	ctx-lh-parsorbitalis	3,615384615	3,769230769	0,108785659
47	ctx-lh- parsopercularis	6,538461538	6,615384615	0,054392829
59	ctx-lh- superiorparietal	13,23076923	13,76923077	0,380749806
61	ctx-lh-precuneus	12,38461538	12,53846154	0,108785663
63	ctx-lh-pericalcarine	9,769230769	10,07692308	0,21757132
67	ctx-lh- parahippocampal	4,846153846	5,307692308	0,326356976
69	ctx-lh-temporalpole	5,153846154	6,307692308	0,81589244
71	ctx-lh- middletemporal	7,461538462	8,230769231	0,543928293

Tabella 7.6. Aree cerebrali che potrebbero essere coinvolte con un'analisi diagnostica più accurata per una malattia neurodegenerativa come la PSP. Per soggetti con questa malattia, contro intuitivamente, 19 aree mostrano addirittura una crescita neurale.

Nella Tabella 7.6, 19 punti della rete mostrano da lieve a notevoli progressi in aree cerebrali (considerato sempre attraverso il valore medio di k). Ciò mostra che la perdita neuronale in alcune aree del cervello determina cambiamenti in altre aree. Questi cambiamenti possono causare lievi riorganizzazioni dell'architettura del cervello. Tali riorganizzazioni funzionali non sono sempre evidenti o facilmente individuabili, dato l'esito di questa letale malattia.

Una diminuzione moderata, da 0,054392829 a 0,924678093 è mostrata nella Tabella 7.7, per 29 aree cerebrali.

Numero nodo	Nome Area Cerebrale	k medio per soggetti sani (gruppo controllo)	k medio per soggetti con PSP	Deviazione standard Gruppo controllo/PSP
30	ctx-rh-middletemporal	7,923076923	7,846153846	0,054392829
31	ctx-rh-bankssts	8,538461538	7,615384615	0,652713952
32	ctx-rh-superiortemporal	13,15384615	11,84615385	0,924678093
34	ctx-rh-insula	16,84615385	15,84615385	0,707106781
38	Right-Pallidum	7,923076923	6,923076923	0,707106781
41	Right-Amygdala	9,307692308	8,461538462	0,598321122
44	ctx-lh-frontalpole	6,461538462	5,846153846	0,435142635
46	ctx-lh-parstriangularis	7,153846154	6,384615385	0,543928293
50	ctx-lh-caudalmiddlefrontal	9,538461538	8,923076923	0,435142634
51	ctx-lh-precentral	13,69230769	13,15384615	0,380749806
53	ctx-lh-rostralanteriorcingulate	7,615384615	7	0,435142634
55	ctx-lh-posteriorcingulate	7,846153846	7,615384615	0,163178488
56	ctx-lh-isthmuscingulate	9,307692308	8,692307692	0,435142635
57	ctx-lh-postcentral	6,692307692	7,153846154	0,326356976
58	ctx-lh-supramarginal	8,615384615	9	0,271964147
64	ctx-lh-lateraloccipital	13,38461538	13,15384615	0,163178487
65	ctx-lh-lingual	9,461538462	8,461538462	0,707106781
66	ctx-lh-fusifiform	14,38461538	13,76923077	0,435142631
68	ctx-lh-entorhinal	5,846153846	4,846153846	0,707106781
70	ctx-lh-inferiortemporal	8,923076923	8,846153846	0,054392829
72	ctx-lh-bankssts	7,923076923	7,615384615	0,217571318
73	ctx-lh-superiortemporal	12,92307692	12,30769231	0,435142631
74	ctx-lh-transversetemporal	5,076923077	4,923076923	0,108785659
76	Left-Thalamus-Proper	8,923076923	8,153846154	0,543928293
78	Left-Putamen	18,30769231	17	0,9246781

79	Left-Pallidum	7,692307692	7	0,489535464
80	Left-Accumbens-area	5,692307692	5	0,489535464
81	Left-Hippocampus	13,84615385	13	0,598321125
82	Left-Amygdala	8,769230769	7,769230769	0,707106781

Tabella 7.7. In questa Tabella sono evidenziate, 29 aree cerebrali che vanno incontro a una degenerazione del tessuto, compresa tra 0,05 e quasi 1.

Nella Tabella 7.7, 29 aree cerebrali subiscono una degenerazione del tessuto, compresa tra 0,05 e quasi 1. Questo significa che la maggior parte del cervello perde cellule in diverse aree cerebrali andando ad intaccare numerosi circuiti cerebrali, che indubbiamente influenza negativamente il sistema cognitivo nel suo insieme. Come accade nei sistemi caotici, piccole variazioni nei parametri iniziali possono portare a cambiamenti catastrofici. Nel caso della PSP, si arriva ad una graduale disintegrazione delle capacità motorie e cognitive, e infine alla paralisi del corpo. Una diminuzione più rilevante delle aree cerebrali (in media da 1 a 2,5) tra soggetti sani e pazienti PSP è rilevabile per le aree cerebrali mostrate in Tabella 7.8. Queste involuzioni indicano un progressivo deterioramento della architettura del cervello. Insieme alle aree tradizionali impiegate per la diagnosi PSP (come il mesencefalo), molte altre aree sono solitamente poco considerate nella valutazione della PSP. I valori medi di k ci permettono di esplorare nuovi parametri per una corretta diagnosi della malattia, così letale e così sottilmente simile al Parkinson (PD).

Nome nodo	Nome Area Cerebrale	K mean per gruppo di controllo	K mean per gruppo PSP	Deviazione standard
2	ctx-rh-parsorbitalis	4,538461538	3,615384615	0,6527139
3	ctx-rh-frontalpole	8,076923	6,230769	1,305428
7	ctx-rh-rostralmiddlefrontal	15,84615	14,23077	1,142249
14	ctx-rh-posteriorcingulate	8,076923	6,230769	1,305428
19	ctx-rh-inferiorparietal	11,46154	12,92308	1,033464
22	ctx-rh-pericalcarine	10,15385	8,538462	1,142249
23	ctx-rh-lateraloccipital	13,46154	11,76923	1,196642
25	ctx-rh-fusiform	16,30769	13,15385	2,230106
33	ctx-rh-transversetemporal	5,846154	3,846154	1,414214
35	Right-Thalamus-Proper	9,076923077	7,615384615	1,033463757

36	Right-Caudate	15,76923	13,69231	1,468606
37	Right-Putamen	18,30769	16	1,631785
39	Right-Accumbens-area	6,923077	5,384615	1,087857
40	Right-Hippocampus	16,07692	14,07692	1,414214
45	ctx-lh-medialorbitofrontal	13	11,15385	1,305428
48	ctx-lh-rostralmiddlefrontal	15,38462	13,76923	1,142249
49	ctx-lh-superiorfrontal	13,46154	10,76923	1,903749
54	ctx-lh-caudalanteriorcingulate	6,307692	4,307692	1,414214
60	ctx-lh-inferiorparietal	13,69231	15,46154	1,251035
62	ctx-lh-cuneus	6,538462	5,076923	1,033464
75	ctx-lh-insula	19,23077	16,92308	1,631785
77	Left-Caudate	15,15385	13,38462	1,251035
83	Brain-Stem	5,384615	2,538462	2,012535

Tabella 7.8. Intervalli di degenerazione in molte aree del cervello non direttamente considerate nella diagnosi di PSP. I parametri medi k ci permettono di approfondire, attraverso questa analisi multifattoriale, come e quanto avviene nella degenerazione neuronale, anche in aree cerebrali, tradizionalmente non considerate rilevanti nella diagnosi di questa malattia.

Numero nodo	Nome area cerebrale	k mean gruppo controllo	k mean gruppo PSP
26	ctx-rh-parahippocampal	4,769230769	4,769230769
52	ctx-lh-paracentral	5,923076923	5,923076923

Tabella 7.9. Aree cerebrali con valori uguali di k per i due gruppi di soggetti.

La teoria delle reti, applicata ai metodi diagnostici per PSP, mostra che i parametri coinvolti nel riconoscimento di questa malattia sono molteplici e difficili da individuare. E' quindi necessario rivedere, in senso olistico, il processo diagnostico. Per raggiungere questo obiettivo di analisi olistica, si usa il parametro k , che ci fornisce il valore che ogni nodo ha nella rete, considerando il numero di connessioni in uscita e in entrata per lo stesso nodo. In poche parole, k indica il grado di connessione e rilevanza di ciascun nodo. Si può quindi utilizzare k come indicatore dello stato della rete. Se k ha un valore elevato per un nodo di rete specifico, significa che questa area del cervello funziona normalmente. Se il parametro è basso, significa il contrario. Confrontando i pazienti PSP con quelli di controllo, troviamo addirittura che in alcune aree

cerebrali che la situazione si è rovesciata, il valore di k cresce. Numerosi sono i punti di divergenza tra i due gruppi, in relazione al valore di k . Questo è anche esemplificato dalla Tabella 7.10, che presenta le organizzazioni per due aree cerebrali l'area numero 1 (Ctx-rh-lateralorbitofrontal., corteccia latero-orbito-frontale destra) e l'area numero 83 (Brain-Stem, Mesencefalo). Come è possibile notare, le differenze non sono molto rilevanti nell'area 1, diventano invece drammatiche nell'area 83. La Tabella 7.11 presenta invece una serie di motivi che si ricollegano ai circuiti funzionali cognitivi. I motivi rappresentano sistemi dinamici che si organizzano, in relazione a mutate condizioni ambientali e interne al soggetto, differenti funzioni cognitive e motorie, diverse funzioni sociali e emotive, nel corso della giornata.

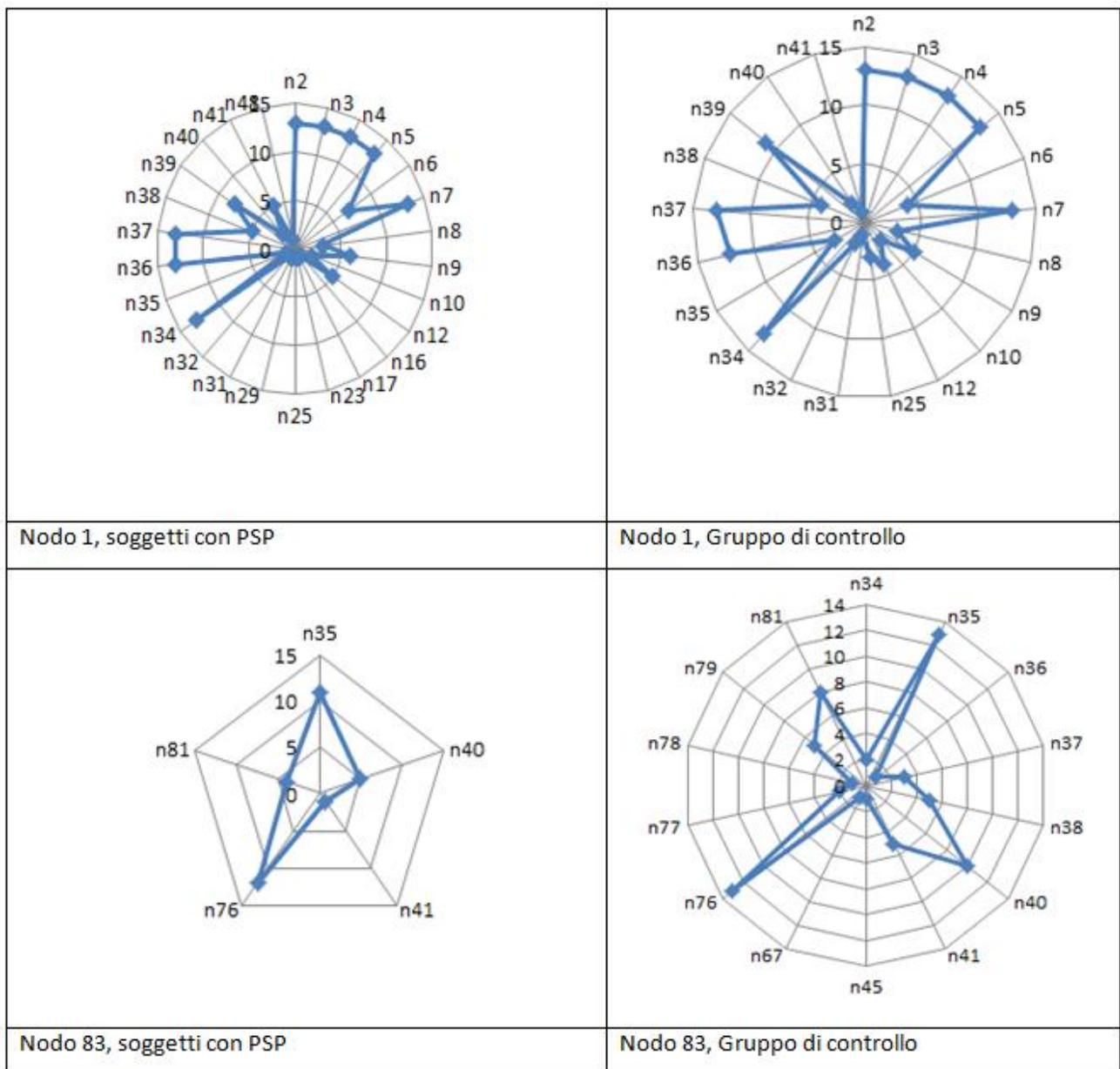


Tabella 7.10. Alcuni pattern di organizzazione nel cervello per la prima area del cervello e l'area del mesencefalo. Come è evidente, queste sotto-reti sono diverse, a seconda dell'organizzazione del cervello del soggetto e a seconda se sano o malato.

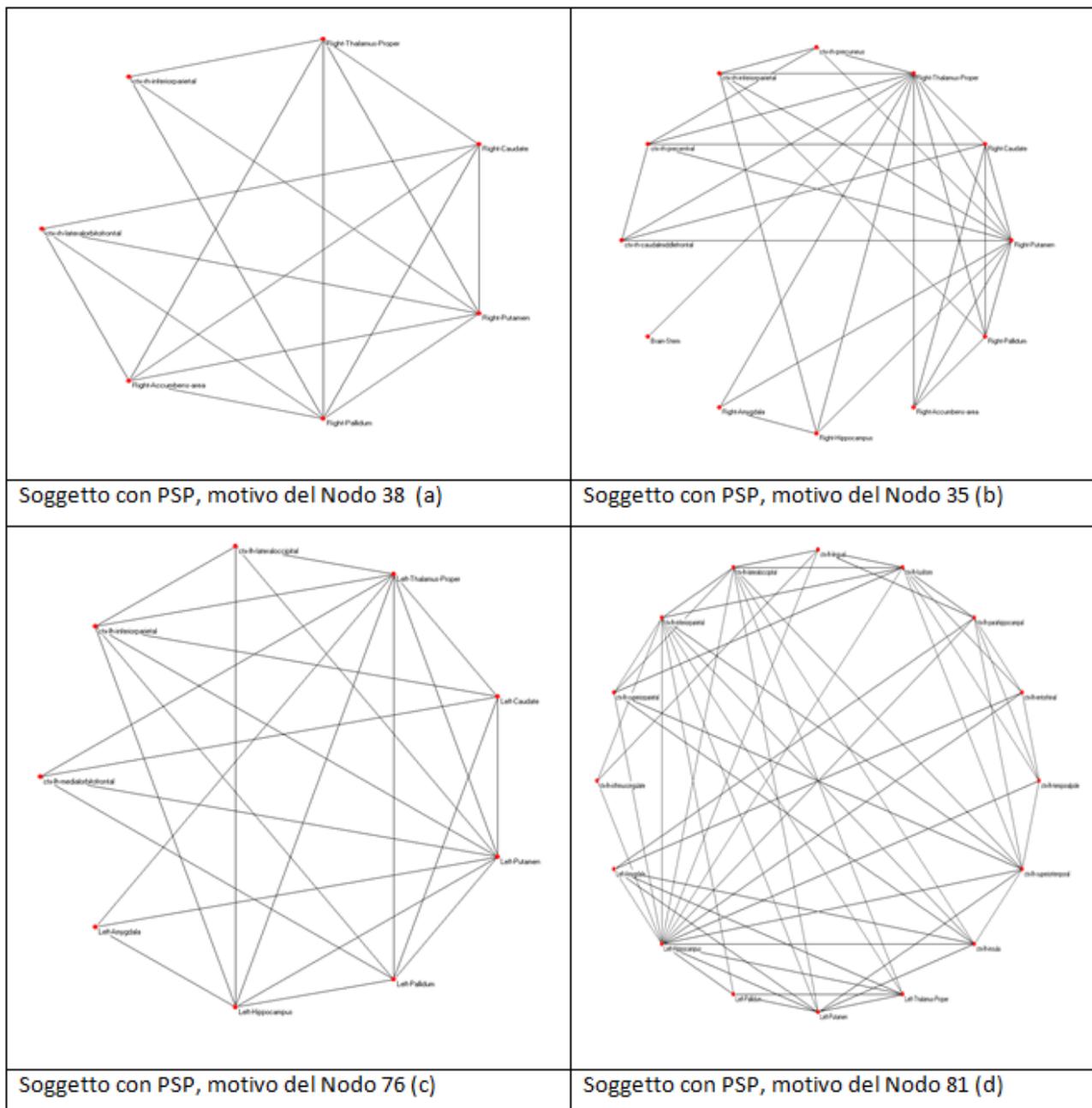


Tabella 7.11. Motivi organizzativi per soggetti con PSP. Le organizzazioni in motivi variano da soggetto a soggetto e nodo per nodo.

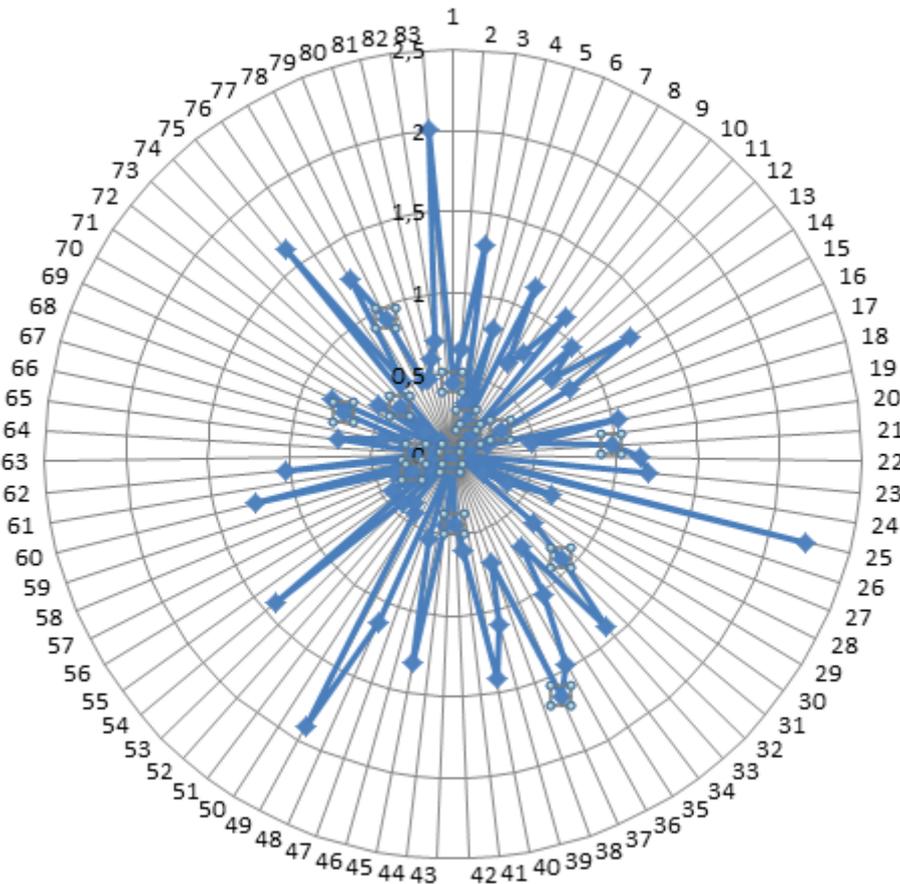


Figura 7.10. Valori di deviazione standard per tutte le 83 aree del cervello tra PSP e soggetti sani.

Le organizzazioni in motivi variano da soggetto a soggetto e nodo per nodo. Infine, nella Figura 7.10 è riportato il grafico della deviazione standard tra soggetti sani e PSP per tutte le 83 aree cerebrali considerate.

7.6. Conclusioni

Molti ricercatori hanno cercato di capire l'organizzazione del cervello umano e numerosi metodi sono stati avanzati e sviluppati per questo scopo. Attualmente, uno dei più utilizzati è la teoria della rete, che fornisce nuovi indizi per comprendere la complessa architettura del cervello umano e quanti milioni di neuroni sono collegati tra loro per ottenere la complessità del comportamento umano. In questo lavoro abbiamo usato la Connectomica e la teoria delle reti per analizzare in profondità le manifestazioni neurofisiologiche della Paralisi Sopranucleare Progressiva (PSP) e, infine, come un cervello umano si ristrutturava, se sottoposto a cambiamenti dovuti ad una malattia neuro-degenerativa. In particolare, abbiamo dimostrato che la connettività cerebrale, sia in soggetti con PSP e soggetti di controllo presenta strutture di piccolo mondo (SWN), con differenze in base al numero di collegamenti, il grado di ogni nodo, il grado di clustering, l'efficienza delle reti, lunghezza del percorso medio, il diametro e la densità delle reti tra i due gruppi. Tuttavia, la prova

che la malattia PSP determina profondi cambiamenti nelle reti del cervello e che quindi lascia il segno nella struttura del cervello è confermata dal fatto che i valori di σ crescono in soggetti con PSP, indicando una crescita di connessioni locali, mentre connessioni più distanti tendono a cadere. Questo spiega il calo delle prestazioni di una rete. L'aumento del diametro, infatti, aumenta la lunghezza dei percorsi nella rete, causando una diminuzione del numero di collegamenti in ciascun nodo e limitando l'accesso a molti nodi essenziali della rete del cervello. Per valori di parametri scale-free, la legge di potenza troncata, il valore è meno rilevante per i pazienti PSP rispetto al gruppo di controllo. Per contro, tuttavia, il grado di ogni nodo è maggiore nei pazienti PSP rispetto ai soggetti sani. A sua volta, questa architettura cerebrale anomala si traduce in una riduzione delle funzioni cognitive e comportamentali, tipiche di soggetti con PSP. Pertanto, abbiamo dimostrato che il cervello ha sia modelli di piccolo mondo che modelli di scale-free troncata, modelli che seguono la legge di potenza, anche se entrambi i sistemi possono avere molteplici manifestazioni specifiche della loro organizzazione, così come accade all'interno del cervello degli individui.

Per dimostrare che la teoria delle reti può essere utilizzata per diagnosticare PSP abbiamo usato il parametro (k). Dopo aver ridotto il cervello in una rete composta da 83 nodi, corrispondenti a 83 aree cerebrali funzionali e strutturali, k ci ha fornito i legami di ciascuna area con tutte le altre all'interno della rete. Seguendo queste connessioni, siamo stati in grado di analizzare con attenzione sia i collegamenti globali che quelli locali, in tutte le aree, in ciascuna materia e in tutte le materie, ma anche come analizzare come l'architettura delle degenera brain, subendo una diminuzione (anche se di tanto in tanto anche una ri-organizzazione) delle sue strutture. Nella nostra analisi, i valori più bassi / uguale / maggiore di k indicano degenerazione / range di normalità / riorganizzazione delle aree del cervello, rispettivamente. Importanti cambiamenti sono rilevati analizzando le aree del cervello corrispondenti ai nodi della rete, identificando tutti i collegamenti per ogni area, per ogni paziente. Ciò ha determinato una mappatura personalizzata di ogni paziente. L'analisi complessiva dimostra che vi è una diminuzione non solo in alcune aree tipicamente identificate con i metodi tradizionali, ma anche un complesso processo di neuro-degenerazione. Abbiamo rilevato cambiamenti a diversi livelli di scala. Ci sono leggere variazioni nel grado di settori cerebrali, sebbene anche piccole variazioni di misura risultano sempre essere interpretati come relativamente grandi nel trattare con un organo sensibile alle alterazioni come il cervello umano. Abbiamo identificato le aree con valori significativamente più bassi per i pazienti di PSP rispetto ai soggetti sani, tra cui la degenerazione delle tradizionali aree del cervello analizzati per diagnosticare la PSP. Abbiamo trovato che solo alcune aree hanno valori identici in pazienti PSP rispetto a soggetti sani e, infine, abbiamo trovato aree che, a differenza della degradazione cellulare tipica della malattia, aumentano i collegamenti rispetto ai soggetti di controllo.

In conclusione, possiamo dire che questo nuovo approccio nella diagnostica medica si articola su quattro aspetti principali:

1. Questo tipo di analisi, coinvolgendo l'MRI diffusivo, non è invasivo. Anche con l'impiego di 83 nodi, è possibile individuare le caratteristiche topologiche del cervello dei pazienti PSP, rispetto a quelle dei soggetti sani. Tendenze future potrebbe comportare la creazione di reti standard per confrontare tutti gli altri soggetti, per i diversi tipi di malattie neuro-degenerative.
2. La caratterizzazione del cervello è locale e globale. La classificazione locale ci permette di individuare tutte le sotto-reti danneggiate nella PSP. In questo modo si può vedere come funziona la comunicazione all'interno del cervello danneggiato, per analizzare come il cervello ristruttura le sue funzioni quando le strutture designate si sono deteriorate. Questo metodo ci permette anche di confrontare i pazienti PSP con un gruppo di soggetti sani, per capire meglio come avviene la comunicazione inter-neuronale o come si interrompe quando il cervello si deteriora.
3. come abbiamo dimostrato in questa tesi, la diagnosi per questa malattia potrebbe essere più specifica e più sensibile rispetto ai parametri utilizzati.
4. Il trattamento può essere personalizzato per ogni individuo, tenendo conto delle caratteristiche specifiche della specifica struttura cerebrale e l'entità del deterioramento. La tecnica ci permette di confrontare individui e gruppi di persone. Infine, il nostro approccio è robusto e semplice, in quanto richiede solo una risonanza magnetica (con tempi variabili a seconda del tipo di sistema di imaging utilizzato e la risoluzione del segnale) e poi per elaborare i dati, come sin qui descritto (Bertacchini et al., sottomesso, b).

Capitolo 8. Crescita di reti emozionali nel sistema WordNet

8.1. Introduzione

Come illustrato nel Capitolo 1, all'interno della Psicologia e delle Scienze Cognitive, si è sostituita la prospettiva "embodied" della cognizione alla visione tradizionale che vedeva la cognizione come computazione. Secondo tale approccio, lo studio dei processi cognitivi non può essere effettuato disgiuntamente dal corpo degli organismi. La cognizione è studiata nel corpo che la esibisce, perché corpo e processi cognitivi si influenzano reciprocamente nell'interazione con l'ambiente. La cognizione si fonda sui sistemi senso-motori (Barsalou, 1999; 2008). I simboli percettivi, in quest'ottica, diventano strumenti della produttività e creatività, caratteristiche tipiche del pensiero umano. Lo sviluppo e l'evoluzione del significato, all'interno di un sistema di comunicazione naturale come la lingua, si basa su tali schemi percettivo-comportamentali. Il bambino apprende i significati delle parole e la struttura sintattica della lingua, utilizzando varie abilità percettive e concettuali che ha imparato a coordinare, individuando le parole e gli oggetti dell'ambiente cui le parole si riferiscono. Per l'acquisizione del lessico, gli si richiede l'uso di capacità cognitivo-sociali, nonché la coordinazione di diversi tipi di informazione. E' sulla base di sistemi di conoscenza essenziali e specifici che il bambino segmenta il continuum degli eventi e inizia a trovare un insieme di concetti, usabili per determinare i significati delle parole. A sua volta, l'uso delle parole influenza la formazione di categorie e, quindi, l'organizzazione concettuale nel bambino (Werker et al., 1998). La comprensione e la produzione delle prime parole sono parte integrante di un processo molto complesso. Le prime vere "parole", contenenti una relazione fra la produzione fonica e un oggetto nella realtà, riproducono versi di animali (onomatopee). Il suono /bau/ è sicuramente immediato per la costruzione della relazione parola-oggetto, rispetto al suono /cane/, che si correla con un rapporto completamente arbitrario all'animale che il termine rappresenta (Golinkoff & Hirsh-Pasek, 1994). Le proto-parole compaiono tra i 9 ed i 14 mesi (Capute et al., 1986; Vihman & McCune, 1994), in stretta connessione con l'ambientale socio-culturale in cui il bambino vive. Questo proto-linguaggio è personale e originale,

comprensibile solo al nucleo familiare del bambino. Al contrario, le parole reali, sono comprese da chiunque possieda lo stesso codice linguistico, realizzando l'idea di codice comune ad un gruppo di parlanti, socialmente condiviso. La maturità linguistica si conquista con la manifestazione di intenzioni e scopi espressi linguisticamente, consentendo al bambino di costruire *categorie naturali prototipali* (Rosch, 1975), in grado di rappresentare gli oggetti, gli eventi e le azioni che mantengono una invarianza nella loro occorrenza. Queste categorie prototipali sono la base della conoscenza e della cognizione (Basile, 2001).

Ma come si forma il significato? Come si passa da un significato ad un altro? Come facciamo a gestire questa mole enorme di significati quando parliamo? Questi problemi sono di fondamentale importanza nella scienza contemporanea, soprattutto in questi ultimi dieci anni, dato il progressivo aumento di testi elettronici che necessitano di essere catalogati, indicizzati per parole chiave, sintetizzati, ritrovati attraverso motori di ricerca e/o interrogazioni a database, e, infine usati per diversi scopi conoscitivi.

Esiste una vasta letteratura sull'argomento che si riferisce a diversi settori (Linguistica, Informatica, Psicolinguistica, Psicologia Cognitiva, Intelligenza Artificiale, Linguistica computazionale, Sistemi per la processazione automatica del parlato e dei testi, Sistemi di traduzione automatici da una lingua all'altra e di disambiguazione dei termini all'interno di vasti database di documenti solo per citarne alcuni). I più importanti problemi sono riferibili ai lavori Miller e all'organizzazione gerarchica utilizzata per la costruzione di WordNet.

Secondo Miller (1995), il vocabolario di una lingua può essere definito *“come un insieme di coppie $W(f, s)$, dove una forma f è una stringa su un alfabeto finito, e un senso s è un elemento da un dato insieme di significati. I moduli possono essere enunciati composti da una serie di fonemi o iscrizioni, costituiti da una stringa di caratteri. Ogni modulo di una lingua con un senso si chiama una parola in quella lingua. Un dizionario è un elenco in ordine alfabetico di parole. Una parola che ha più di un senso è polisemica, due parole che condividono almeno un senso in comune sono detti sinonimi. L'uso di una parola è rappresentato dall'insieme C dei contesti linguistici in cui la parola può essere utilizzata. La sintassi del linguaggio C si divide in categorie sintattiche. Parole che occorrono nel sottoinsieme N sono sostantivi, parole che occorrono nel sottoinsieme V sono verbi, e così via. All'interno di ciascuna categoria di contesti sintattici vi sono ulteriori categorie di contesti semantici - l'insieme dei contesti in cui una particolare f può essere utilizzata per esprimere un particolare s . La morfologia della lingua viene definita in termini di una serie di rapporti M tra forme di parole. Per esempio, la morfologia della lingua Inglese è divisa nelle relazioni morfologiche flessionali, derivazionali, e composte. Infine, la semantica lessicale della lingua viene definita in termini di una serie di rapporti S tra sensi parola. Le relazioni semantiche in cui una parola entra determina la definizione di quella parola”*. Abbiamo riportato questa ampia citazione della modellazione del linguaggio, realizzata da Miller in WordNet per indicare che da questa prima incarnazione di una teoria

psicolinguistica in un sistema computazionale, sono derivati numerosi studi e ricerche che hanno coinvolto ricercatori appartenenti a diversi settori. Da notare in particolare:

- a. la critica di Johnson-Laird (1984) sulle reti semantiche, non idonee ad incarnare un modello cognitivo di rappresentazione mentale nell'uomo. Secondo questo psicologo, i fenomeni semantici che si producono in WordNet sono riferibili solo ai collegamenti nel sistema stesso (e alla sua strutturazione gerarchica) e non al mondo reale; i problemi sembrano più linguistici che cognitivi.
- b. La critica della teoria componenziale del significato (Katz & Fodor, 1963; Fodor, 1975), secondo la quale i processi cognitivi sono realizzati da computazioni sulle rappresentazioni mentali, considerando questo insieme di rappresentazioni come un linguaggio del pensiero (o *mentalese*), che esibisce le stesse proprietà del linguaggio naturale. Le componenti delle rappresentazioni sono strutturate in pezzi. Per cui per il termine "uomo", le componenti di significato sono: ([umano], [maschio], [adulto], [essere animato], ecc.). Tali componenti, detti *semi* o *tratti semantici* di un concetto, che sono la forma minima del significato, si combinano tra loro in base a delle regole, organizzando il significato. Dal confronto tra i tratti semantici si possono ricavare le relazioni semantiche tra i concetti primitivi e i quattro tipi di relazioni semantiche sinonimia, opposizione, gerarchia e meronimia, trattati nel paragrafo che segue e già individuati nel Capitolo 5 per lo sviluppo degli algoritmi di ricerca e la modellazione delle lingue naturali.

8.2. Il contesto di riferimento psicolinguistico

Lo studio dei segni linguistici, come le parole, le espressioni e le frasi appartiene alla Semantica (Greimas, 1966; Hjelmslev, 1961). Si indaga su come i segni linguistici acquistino la proprietà di trasmettere i significati. Dagli studi di Semantica, gli psicologi ottengono informazioni per la descrizione dei processi mentali di produzione/comprendimento di una lingua naturale negli umani, per arrivare a determinare la natura del linguaggio e la formazione della base di conoscenza, di solito contenuta nella memoria semantica di ogni soggetto (Schacter, 1994; Lockhart, 2000; Tulving, 1972). Un insieme di parole di una lingua, che condivide una certa proprietà semantica e che si riferisce ad uno stesso gruppo organizzato di significati legati tra loro, è detto campo semantico (Lyons, 1977; Simone, 1990). Nella linguistica tradizionale, tale concetto si usa per individuare un gruppo di parole nel lessico usate per parlare di fenomeni generali, come per esempio l'organizzazione dei nomi dei colori (problema che si sposa con la relativa organizzazione culturale dello spettro visivo), i nomi di parentela, o i termini legati ai nomi dei cibi e del cucinare, i nomi di animali e piante. Tali confini variano da cultura a cultura. Simone (1990) parla anche di famiglia semantica, un concetto più esteso di *family resemblance*, riferibile a Wittengstein (1953). Secondo Wittengstein (1953), poiché è molto difficile specificare un insieme di condizioni necessarie e sufficienti per le categorie naturali, una definizione probabilistica, basata sulla somiglianza di famiglia, è più

consona per la definizione di categorie concettuali (e/o campi semantici), depositati nella memoria semantica (Tulving, 1972). Alcuni membri di una famiglia di concetti sono "tipici", perché condividono numerose caratteristiche con molti altri membri della famiglia; altri sono atipici perché condividono solo poche caratteristiche con gli altri membri. In una lingua naturale molte parole appartenenti semanticamente ad una famiglia, sono sentite come fare parte dello stesso dominio, o come aventi significati affini. Perché tali parole siano sostituibili nello stesso contesto, esse devono avere lo stesso valore paradigmatico, cioè devono poter occupare lo stesso posto nell'enunciato. Dalla teoria della somiglianza di famiglia di Wittgenstein scaturisce la teoria del prototipo (Rosch, 1975). E' possibile derivare un prototipo da una struttura di somiglianza di famiglia, combinando tutte le caratteristiche più tipiche in un'unica descrizione che sarà la descrizione prototipica. La teoria della somiglianza di famiglia apre la strada anche alle teorie probabilistiche dei concetti (natura probabile più che necessaria di molte proprietà e confini sfuocati – fuzzy - delle categorie. L'ipotesi della somiglianza di famiglia presume:

1) gradualità di appartenenza categoriale;

2) gradualità determinata dal diverso possesso di attributi tale per cui gli esemplari che condividono con gli altri esemplari del concetto un maggior numero di attributi sono più centrali rispetto agli esemplari che condividono un numero minore di attributi;

3) centralità come caratteristica strutturale della rappresentazione delle categorie.

Inoltre, un altro elemento di fondamentale importanza nella scienza contemporanea, è sapere come è acquisito, stivato e generato il significato, soprattutto in ambito della linguistica computazionale applicata, che ha adottato gli approcci della Psicologia cognitiva e della Intelligenza Artificiale per modellare il comportamento linguistico, ritenendo che il funzionamento della mente possa essere paragonato a un computer che elabora le informazioni provenienti dall'esterno, e restituendo le informazioni in forma di rappresentazione della conoscenza, organizzata in reti semantiche. Per cui, una rete semantica è una forma di rappresentazione della conoscenza. Le reti semantiche sono state proposte come modello della memoria umana (Tulving & Patkau, 1962; Tulving & Craik, 2000; Tulving, 1985, 2001), e largamente utilizzate per la comprensione del linguaggio naturale. Originariamente sono nate come un modo di rappresentare il significato delle parole di una lingua naturale e quindi sono state utilizzate per rappresentare enunciati. Le parole nel cervello sono organizzate come una grande rete: esse ci forniscono la chiave per comprendere la struttura linguistica. Un processo degno di nota del nostro cervello, che ha fatto riflettere molti ricercatori (Loftus, 1974; 1975; Greely & Illes, 2007; Gazzaniga, 2005), è il falso ricordo. Non si tratta di una menzogna volontaria, della quale il soggetto ha piena coscienza, ma qualcosa che non è mai accaduto, è un evento originato nella nostra memoria (dovuto probabilmente ad associazioni fra contenuti, che si sono stabilizzate nel tempo). Per cui, il soggetto che ha creato le i collegamenti tra concetti o eventi che sono solo

immaginati, torna a pensare a tali associazioni, come se fossero eventi avvenuti realmente nella sua esperienza. Il campo della memorizzazione e della creazione di conoscenza riveste un particolare interesse nella neuroscienza cognitiva. I lavori di Quillian intorno alla fine degli anni sessanta (Quillian, 1967, 1968, 1969), segnano una tappa fondamentale della rappresentazione della conoscenza attraverso le mappe semantiche. Si basano sull'idea che un concetto possa essere rappresentato da un nodo (esattamente come nella teoria delle reti), e che il suo significato dipenda dall'insieme di relazioni che ogni nodo-concetto instaura con gli altri nodi-concetto. Sempre come nelle reti, le relazioni sono individuate da collegamenti o archi. Gli archi possono essere dotati di pesi (hanno degli attributi o proprietà che articolano la struttura del significato), che individuano il contributo di ogni proprietà nel determinare il significato del concetto (Collins & Quillian, 1969). I nodi-concetti connessi ad un dato nodo-concetto sono a loro volta connessi ad altri nodi e così via.

Collins & Quillian, riprendendo la tradizione della retorica classica, nella quale tali elementi erano sviluppati, usò diversi tipi di arco tra cui quelli che rappresentano la relazione "è un", "ha un".

La prima relazione specifica un concetto-nodo con un altro concetto, mentre la seconda implica la componibilità di un concetto-nodo con altri concetti-nodi. Entrambe queste relazioni permettono di specificare i concetti-nodo o di aggiungere sotto classi ai concetti-nodo. Esempio di "è un" e "ha un",

1. *L'albero è una pianta,*
2. *L'albero ha le foglie.*

Secondo il modello di Collins & Quillian, durante la ricerca di significati, gli umani analizzano in lungo e in largo le mappe semantiche. La ricerca del significato di un concetto parte da uno specifico concetto-nodo e si propaga sulle connessioni e, da questi ai nodi ad essi connessi e così via. Nella teoria dei grafi queste esplorazioni sono i cammini minimi tra nodi. Quando tale processo (detto propagazione dell'attivazione) raggiunge un nodo, quest'ultimo viene "attivato" e etichettato, in modo da registrare il nodo di partenza e quello che lo precede. Se un nodo viene attivato da due cammini, con un'attivazione proveniente da più nodi concetto, esiste una relazione di significato o *intersezione* tra i due nodi di partenza dei due cammini.

Collins & Loftus (1975) generalizzano il modello di Collins & Quillian, basandosi sui risultati dei loro studi sulla fenomenologia delle reti semantiche in soggetti umani. Basandosi e rafforzando la metafora circuitale della rete semantica, i due autori ritengono che quando un concetto è stimolato, l'attivazione si propaga attraverso i possibili cammini della rete che hanno origine nel concetto stesso, come un segnale elettrico. Inoltre, al pari di un segnale elettrico, la propagazione dell'attivazione è attenuata in modo inversamente proporzionale alla forza (peso) degli archi dei cammini. Numerose sono le dinamiche che si individuano in un campo o in una famiglia semantica, per decidere l'ordinamento del lessico che modella i processi

cognitivi, legati all'accesso e alla selezione di porzioni di tale sistema di conoscenze strutturate, nei processi di pensiero e più specificatamente nella produzione di linguaggio scritto e parlato in una data lingua storico-naturale.

Le loro tesi sulla natura delle mappe semantiche e sulle loro connessioni sono espresse sinteticamente di seguito:

- 1) Più a lungo un concetto è stimolato (per esempio attraverso la lettura o l'ascolto), maggiore è la durata dell'attivazione. Può essere processato solo un concetto alla volta, limitazione imposta dalla natura seriale dei processi centrali umani (Collins & Quillian 1972) e, possiamo dire oggi, dal particolare modello euristico utilizzato dalla Scienza Cognitiva di quel periodo.
- 2) L'attivazione decresce con il tempo, o a causa di altre attività cognitive, intraprese dal soggetto durante il processo.
- 3) L'attivazione che proviene da sorgenti differenti, si somma, proprio come nei sistemi fisici circuitali e, nel momento in cui l'attivazione su un nodo di intersezione supera un livello di soglia, inizia un processo di analisi del cammino (quello con origine in un nodo diverso dal nodo di partenza) che ha prodotto l'intersezione.
- 4) Esiste una memoria lessicale diversa dalla rete semantica. La rete semantica è tale che maggiore è il numero di proprietà che due nodi-concetti condividono, maggiore è il numero di archi che li congiungono. Inoltre, maggiore è la lunghezza delle linee che congiungono due nodi, minore è la vicinanza di significato e di relazioni tra i due concetti che i nodi rappresentano (Miller, 1969; Arnold, 1971).
- 5) I nomi dei concetti sono immagazzinati in una rete lessicale i cui archi rappresentano proprietà fonemiche dei nomi. La similarità fonemica svolge il ruolo svolto dalla similarità semantica nella rete semantica. Ogni nodo della rete lessicale è connesso ad uno o più nodi della rete semantica.
- 6) Una persona può controllare se ha attivato la rete semantica, quella lessicale o entrambe.

Un ruolo importante nello studio dell'organizzazione cognitiva del linguaggio è giocato da Miller (Miller, 1967, 1969, 1972), che, nel 1956, insieme a Noam Chomsky, Herbert Simon e Alan Newell, partecipando al Simposio sulla Teoria dell'Informazione, tenuto al MIT, dà origine a quell'insieme di idee e di ipotesi sul funzionamento della mente umana che confluiscono nelle Scienze Cognitive. La Psicologia Sperimentale umana, la linguistica teorica e la simulazione dei processi cognitivi sono tutti pezzi di un insieme più largo, che, in seguito, avrebbero dato vita a "*progressive elaborazioni e coordinamento dei loro fini condivisi*" (Miller, 1979).

George Miller e i suoi collaboratori hanno sviluppato, in quasi 30 anni di lavoro il sistema WordNet, (Miller et al., 1990; Miller, 1998), il più grande data base elettronico lessicale, interrogabile e liberamente accessibile, fondato su migliaia di esperimenti realizzati da soggetti umani. Il sistema è formato da due tipi di elementi: parole e concetti. Un tipico esperimento è fornire ai soggetti liste di parole, cui i soggetti devono associare i concetti, nell'organizzazione sinonimica, iperonimica, iponimica, e dei contrari. I termini utilizzati sono sostantivi, verbi, avverbi e aggettivi. WordNet è il tentativo di incapsulare in una rete semantica, la teoria psicolinguistica attraverso la quale è possibile definire e modellare il significato di una parola, le associazioni di significato tra parole, e le associazioni tra significati. Le parole possono essere collegate tra loro attraverso una serie di rapporti quali sinonimia, antonimia, iperonimia, iponimia. Parole correlate, e contrari. Parole che sono unite insieme da associazioni rappresentano un insieme di sinonimi, chiamati synset (set di sinonimie). Ogni concetto ha uno o più synset associati. I concetti sono collegati da relazioni quali iperonimia (acero e albero) e meronimia (uccello e becco). WordNet contiene solo quattro parti del discorso: verbi, sostantivi, avverbi e aggettivi. Per cui WordNet può essere considerato un thesaurus esteso e migliorato, con informazioni non di solito rintracciabili in altri thesauri.

Come riportato nel Capitolo 5, le funzioni di WordNet sono le seguenti:

- a. iponimia, anche detta inclusione (relazione detta "tipo di" o "IS PART OF"), rappresenta un ordinamento attraverso il quale il significato di un termine come tavolo è incluso nel campo semantico di mobile (Basile, 2001), che, al contrario, è invece un iperonimo. Iponimia e iperonimia ci forniscono le dimensioni estensionali e intensionali rispettivamente, dove la classe degli iperonimi (tutte le categorie generali dei nomi del lessico di una lingua) che comprendono tutte le categorie di nomi specifici di una classe generale. Due parole iponime di uno stesso iperonimo si dicono co-ponime ("topo", "gatto", sono iponimi di "animale"). L'iponimia è anche detta relazione "IS A", che si riferisce al significato generale dell'iperonimo corrispondente, condividendo alcuni tratti semantici, intesi quasi come elementi di una *genetica dei tratti semantici* (Simone, 1990). Chaffin (1992) formalizza questa appartenenza sostenendo che Y è un tipo di X, dove Y è l'iponimo e X l'iperonimo. Per cui si deduce che a un termine iperonimo corrispondano numerosi iponimi, che tra loro stabiliscono un rapporto di co-ponimia (Lyons, 1977). Si parla di *entailment* per riferirsi al discorso di inclusione. Per cui si ha che se X è la classe degli animali, e Y la classe dei gatti, allora Y include X, oppure si può dire che Y è un iponimo di un altro termine X se e solo se Y implica X (Y X) (Carnap, 1950). Secondo tale autore, lo status di (Y X) è costruito secondo il modello dei postulati di significato. Gli iperonimi possono a loro volta costituirsi insieme, e essere collegati a iperonimi di iperonimi, in uno schema di organizzazione gerarchica della conoscenza. Per cui esistono dinamiche orizzontali e dinamiche verticali (come succede nella musica). Per cui tra iponimi e iperonimi (a

differenti livelli gerarchici), si instaurano rapporti di transitività. Per esempio, le caratteristiche del termine “*pecora*” si riflettono su “*ovino*” e si trasferiscono sul termine “*animale*”

- b. Antonimia è la tipica organizzazione di significato che, per ogni dato nodo in una rete, individua il suo significato contrario. Questo meccanismo serve a far passare attraverso gradazioni di significato, con un’esplorazione completa del campo semantico per arrivare ai suoi confini e superarli. Sempre secondo Miller, anche questa è una relazione simmetrica. Questa relazione interessa in particolare i significati di aggettivi e avverbi.
- c. Meronimia è una relazione detta “*parte di, in relazione a un tutto*”, che distingue in parti sostanziali e parti membri di un oggetto. Esempio, il termine “*automobile*”, di cui fanno parte, “*ruota*”, “*motore*”, “*pistone*”, “*sedile*”, ecc. Questa dinamica è molto simile all’iponimia, ma il processo di transitività e di trasferimento del contenuto informativo da un termine all’altro alle volte può essere possibile alle volte no. In un testo scritto, la co-occorrenza del nome “*automobile*”, ci informa che prima o poi anche gli altri termini compariranno. Questo vuol dire che il testo ha una sua coesione interna e chi ascolta o legge o parla riesce a capire che anche se non si fa più riferimento ad “*automobile*”, le frasi che contengono “*ruota*”, “*finestrino*”, “*volante*”, appartengono allo stesso contesto.
- d. Sinonimia, sono insiemi di termini che possono essere intercambiabili all’interno di un contesto linguistico, senza che la frase subisca modifiche di significato. Di solito gli insiemi di termini che condividono lo stesso significato, creano *clique*, ovvero nodi dove ogni elemento è legato a tutti gli altri. Secondo Miller, questa è la relazione di base e simmetrica tra forme di parole di WordNet.

In questo capitolo analizziamo la rete WordNet prima come rete lessicale composta da circa 140.000 termini, almeno fino al momento di accesso alla rete nel mese di Luglio. Poi analizzando i meccanismi di crescita dei termini emozionali definiti nel Capitolo 5. Il metodo messo a punto consentirà di osservare le caratteristiche strutturali e funzionali e il comportamento di una grande varietà di particelle emergenti, caratterizzando il linguaggio come sistema complesso, oltre che come rete linguistica.

8.3. Risultati sperimentali per la rete WordNet.

L’intero database è stato acquisito in tempi differenti, acquisendo l’enorme quantità di termini che esso presenta. l’intero database WordNet è stato trasformato in una rete. In Figura 8.1, è presentata una immagine della rete risultante. E’ composta da 139.999 nodi e da 1021169 archi. Come è possibile notare, nessuna struttura si evidenzia, data l’enorme consistenza della rete.

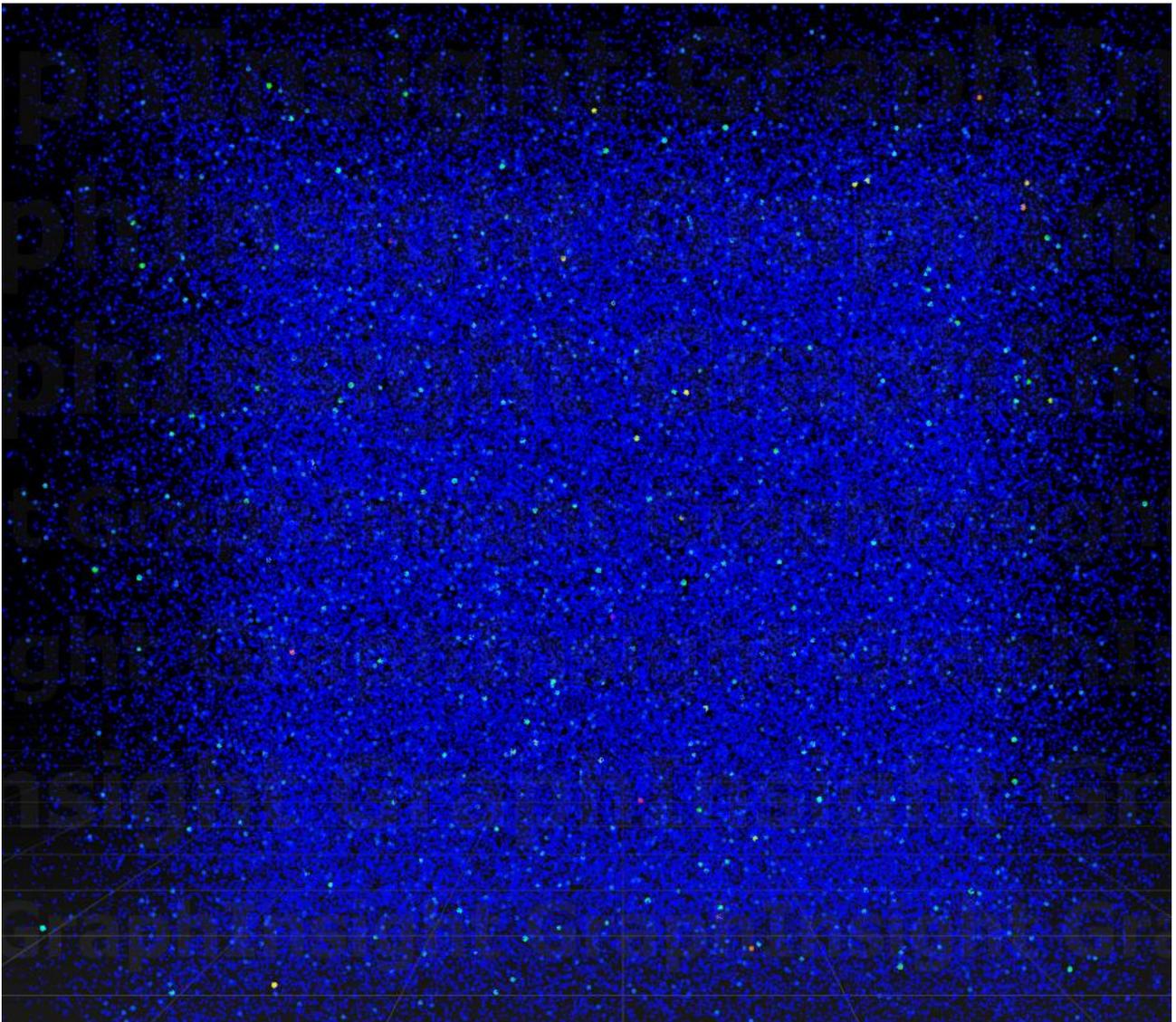


Figura 8.1. La rete di WordNet presenta 139.999 termini e 1021169 connessioni.

Nella Tabella 8.1 si riportano i 24 termini col grado maggiore.

Termine	Grado
change	2037
writer	1968
author	1968
herb	1855
city	1816
metropolis	1810
herbaceous_plant	1810

urban_center	1809
bird_genus	1567
bush	1561
shrub	1505
move	1452
alter	1422
mammal_genus	1402
modify	1390
person	1346
individual	1339
soul	1327
animal_order	1193
asterid_dicot_genus	1176
fish_genus	1150
rosid_dicot_genus	1136
class	1033
arthropod_genus	1012

Tabella 8.1. I 24 nodi con il grado maggiore della rete WordNet.

Nella Tabella 8.2 è riportato il rank order dei primi 20 nodi per *Authority e Hub*, calcolati tramite il software Pajek. Questi termini, che in parte coincidono con i termini riportati in Tabella 8.1, dimostrano che esiste una relazione tra il grado dei nodi e la loro importanza all'interno della rete.

Valore di Authority	Termini	Valore di Hub	Termini
0.0840	change	0.4497	city
0.0662	move	0.4494	urban_center
0.0615	alter	0.4494	metropolis
0.0588	modify	0.2900	change

0.0545	turn	0.2380	alter
0.0539	make	0.2332	modify
0.0480	take	0.1259	port
0.0476	set	0.1105	move
0.0470	cut	0.0631	turn
0.0465	go	0.0623	go
0.0464	break	0.0595	make
0.0445	get	0.0562	take
0.0410	pass	0.0521	Set
0.0395	draw	0.0515	set
0.0361	touch	0.0442	pass
0.0350	raise	0.0433	place
0.0348	play	0.0424	travel
0.0347	fix	0.0407	position
0.0347	check	0.0406	get
0.0346	work	0.0391	Be

Tabella 8.2.Valori dei termini per Authority e Hub.

Il grado medio di questa rete è 14,59; la sua densità è 0,000545, mentre il coefficiente di clustering medio è 0,565. Accanto a queste proprietà globali, è interessante analizzare anche le sottoreti relative a ciascun termine. Tali sottoreti sono dette *locali*. Dato un termine w_i , di indica con *rete locale di ordine 1*, associata ad i , la sottorete $G_i \subset R_G$, della rete globale che contiene il termine w_i e il suo intorno. Si chiama *rete locale di ordine n*, associata ad i , la sottorete $G_i^n \subset R_G$ che contiene tutti i termini la cui distanza da w_i è pari o minore di n .

Dalla Tabella 8.3 si rileva che la rete si organizza attorno al suo nodo centrale che è “change”.

Intorno	nodi
1	1060
2	16176
3	68807

4	120344
5	135691
8	139083
10	139095
15	139095

Tabella 8.3. Come la rete si organizza attorno al nodo centrale “change”. I calcoli sono stati fatti con il software Pajek.

Dalla Tabella 8.3 è possibile osservare che, se consideriamo G_{change}^{10} , questo insieme possiede 139095 termini, e, resta stabile, se si aumenta l’intorno. I restanti 904 nodi sono isolati rispetto a questo blocco principale, il cui diametro non può essere >20. Se si analizzano i valori della Tabella 5.2, si nota che 135.691 sono contenuti in G_{change}^5 . Questo significa che meno del 4% dei termini risiede ad una distanza superiore a 5, rispetto al termine “change” (il diametro di G_{change}^5 non può essere superiore a 10). Per cui, si può dire che il nucleo centrale del lessico è estremamente compatto.

8.4. Crescita dei termini emozionali, metodo di indagine

Dal punto di vista metodologico, l’esperimento condotto si sviluppa in differenti fasi:

- a. Individuazione dei termini da analizzare e scelta dell’algoritmo di indagine, fra gli 8 sviluppati. Per questa fase, si individua l’ambito all’interno del quale si conduce la sperimentazione. Per questo esperimento è stato scelto il vasto campo semantico del termine “emotion”, che dentro WordNet è un nodo terminale. Individuazione dell’algoritmo di crescita fra quelli sviluppati. Si ricorda che gli algoritmi sono i seguenti: , R_1 : Antonimia, R_2 : Iperonimia, R_3 : Iponimie, R_4 : Implicazione, R_5 : Similarità, R_6 : Sinonimie, R_7 : Related e R_8 : Overview. La scelta dipende molto dai diversi contesti in cui l’analisi è realizzata. Se si pensa, per esempio, ad un contesto letterario, potrebbe essere interessante individuare delle ricerche attraverso l’algoritmo R_5 : Similarità, o R_7 : Related, in funzione delle necessità specifiche dell’investigazione filologica. Altri settori potrebbero manifestare necessità completamente diverse.
- b. Analisi qualitativa delle reti, per ogni singolo livello di crescita. Si riportano le trasformazioni qualitative e quantitative nelle reti per le possibili relazioni sopra menzionate, evidenziando i pattern che tali relazioni creano e i percorsi di cambiamento qualitativo nelle reti. A titolo esemplificativo, si riporta tale analisi per due termini emozionali, “anger” e “disgust” e si danno

alcuni esempi per altri termini, per evidenziare fenomeni qualitativi quali la crescita rilevante della rete, la direzione di crescita, l'emergere delle particelle.

- c. Analisi statistica delle reti per ogni livello di analisi, per tutti i livelli.
- d. Individuazione delle strutture complesse, secondo la tassonomia riferita nel Capitolo 5. Ricordiamo a questo proposito che le *Particelle Emergenti* (PE) sono state classificate in 4 categorie principali: *grafi perfetti*, indicati con la sigla PE_di tipo_A, *grafi imperfetti a causa di rotture di simmetrie*, indicati con la sigla PE_di tipo_B, *grafi con simmetrie complesse*, indicati con la sigla PE_di tipo_C e *particelle complesse, risultanti da particelle più elementari*, indicati con PE_di tipo_D. Ogni particella, così classificata è seguita inoltre da un numero, che richiama l'ordine secondo cui la particella è stata individuata nella rete analizzata.

8.4.1. Evoluzione e analisi qualitativa delle reti

Nella prima fase di questo lavoro sperimentale si presenta la rete del lessico che deriva dal termine "Emotion" (Figura 8.2), e si discute il ruolo delle particelle emergenti in WordNet.

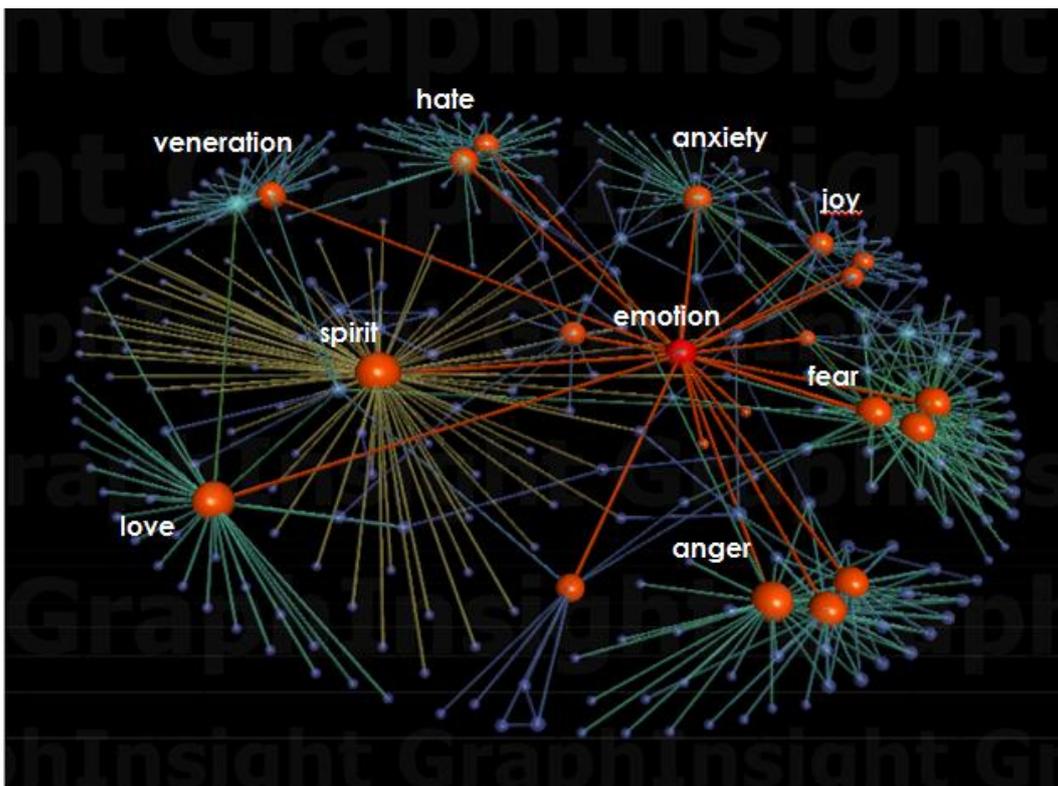


Figura 8.2. La rete del termine emozione al terzo livello degli iponimi. Ricordiamo che in WordNet esistono dei termini base, che danno origine ad una serie di termini sotto-ordinati (appunto iponimi) e classificati come R_2 , nel Capitolo 5.

Da questa rete, abbiamo estratto i termini emozionali, secondo la teoria di Plutchik (1980), che sono i seguenti :

"anger"	"disgust"	"expectation"	"fear"	"joy"	"sadness"	"acceptance"	"surprise"
---------	-----------	---------------	--------	-------	-----------	--------------	------------

Tabella 8.1. Termini indagati attraverso algoritmi di crescita. I termini fanno parte del campo semantico delle emozioni. Per cui derivano da un iperonimo per tutti i termini che è "emotion".

I differenti algoritmi di ricerca sviluppati portano a creare reti completamente diverse, individuando nella struttura profonda di WordNet, le sue relazioni. Per raccogliere tali dati sperimentali, è stato usato l'algoritmo di ricerca R_8 : *Overview*. Per ogni termine emozionale, la ricerca è stata realizzata per 6 livelli differenti, mantenendo costanti le variabili ALL (l'algoritmo cerca e inserisce nella rete tutte le categorie grammaticali per il termine oggetto di studio, variando solo la profondità di livello (per questo esperimento, dal livello 1 al livello 5). Per ogni livello di indagine, non abbiamo considerato le relazioni che si stabiliscono tra nodi terminali. Tutte le visualizzazioni sono state realizzate con il software *GraphInsight*.

"Anger"

Per il termine "anger", come per gli altri termini emozionali in questa parte sperimentale, sono stati realizzati 6 livelli di crescita. Il livello 1 e 2 sono visualizzati nella Figura 8.3.

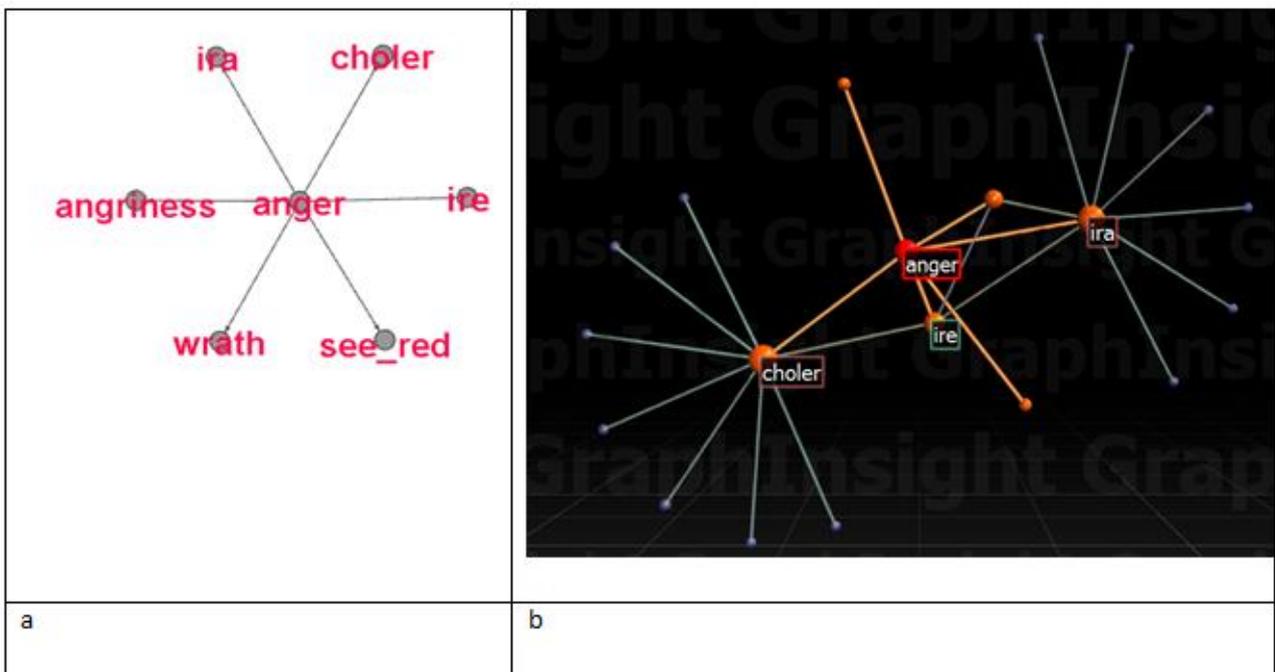


Figura 8.3. La rete del termine "anger" a livello 1 (a) e 2 (b).

A livello 1, la rete presenta 7 nodi e 6 archi. A livello 2, presenta 20 nodi e 33 archi. Accanto ad “*anger*”, i termini “*choler*”, “*ire*” e “*IRA*”, manifestano a loro volta rilevanti capacità di crescita, come si nota dalla Figura 8.4, che visualizza il livello 3.

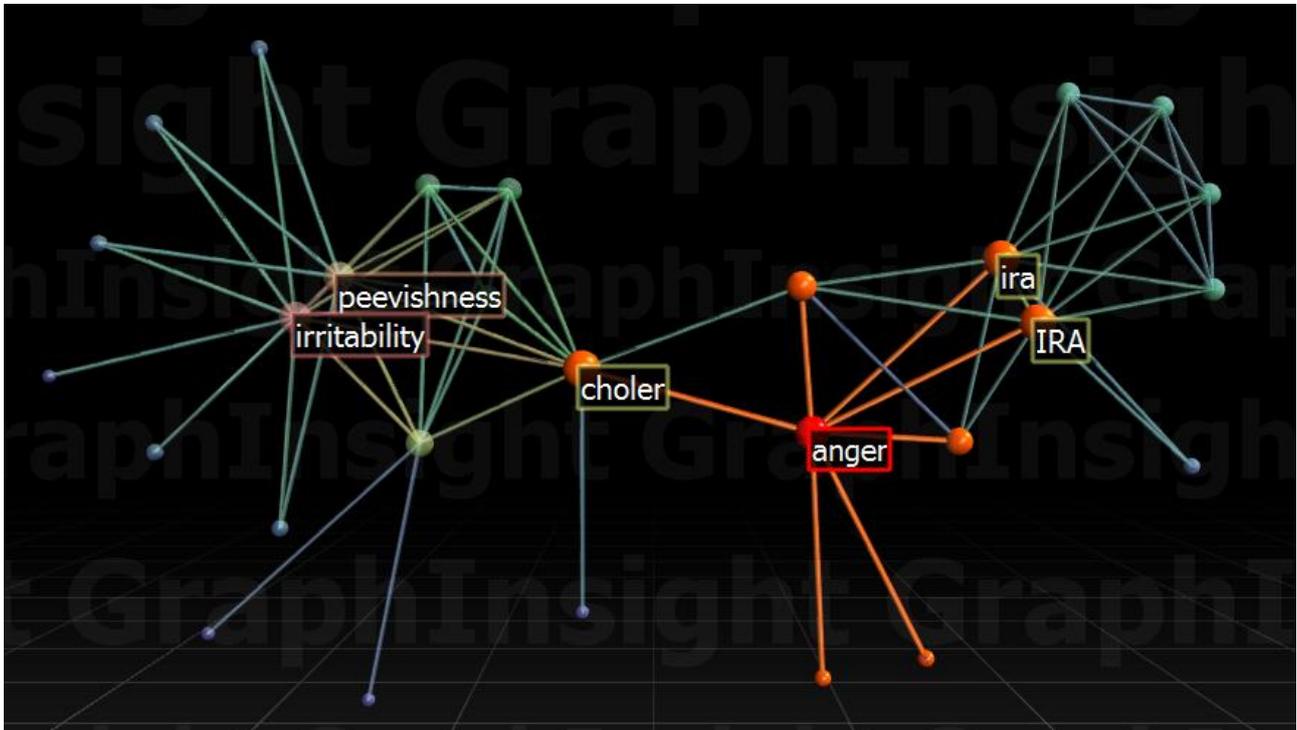


Figura 8.4. La rete del termine “*anger*” a livello 3.

Si rilevano due meccanismi di crescita differenti a partire dai due termini “*choler*” e “*IRA*”. Da una parte, si assiste all’emergenza di strutture complesse, che hanno come radice i termini “*peevishness*” e “*irritability*” (sulla sinistra di Figura 8.4); dall’altra, si assiste all’emergenza di una struttura complessa, basata sulla coppia “*ire*” e “*IRA*”. Nei paragrafi successivi, tali strutture sono analizzate in grande dettaglio, attraverso una tassonomia di classificazione dei grafi, sopra riportata

A livello 4, la rete si evolve sempre sulla sua parte sinistra (Figura 8.5), dove appare un nuovo nodo, con alto grado, relativo al termine “*temper*”.

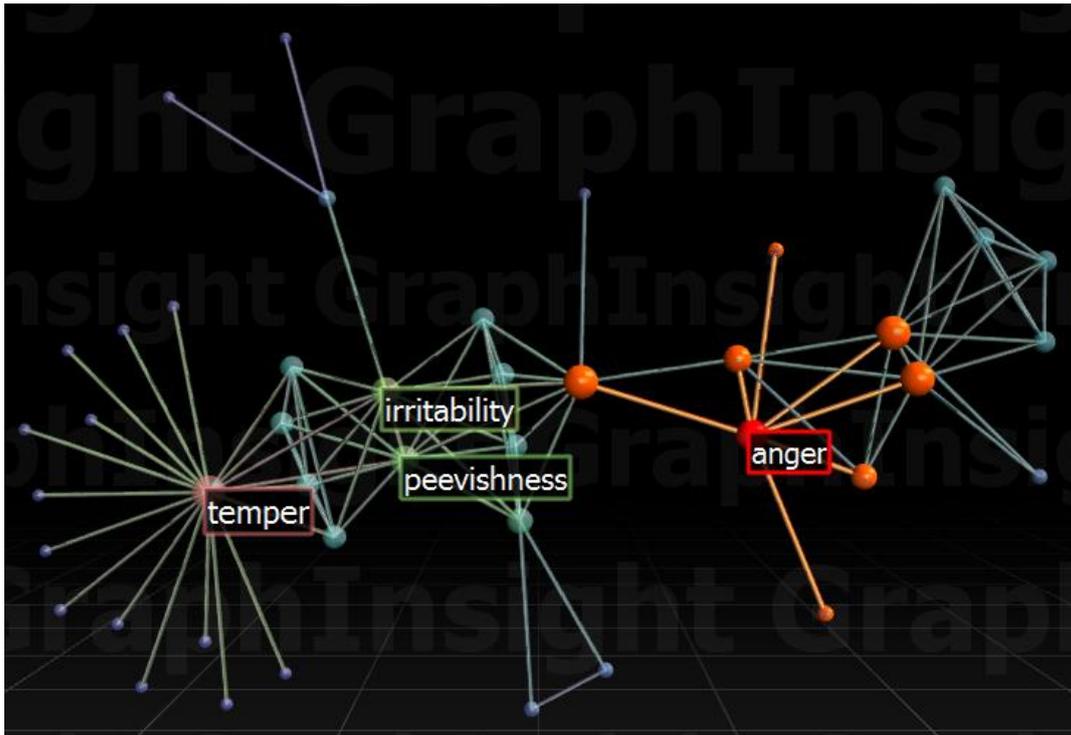


Figura 8.5. La rete del termine "anger" a livello 4.

La rete ha per questo livello, 43 nodi e 159 archi. Una crescita di tipo ben diverso dall'attaccamento preferenziale di Barabási & Albert si manifesta, avendo ogni nodo la possibilità di svilupparsi in molteplici dimensioni dello spazio semantico. Continuando ad analizzare l'evoluzione della rete, si ottiene, per il livello 5, la rete di Figura 8.6.

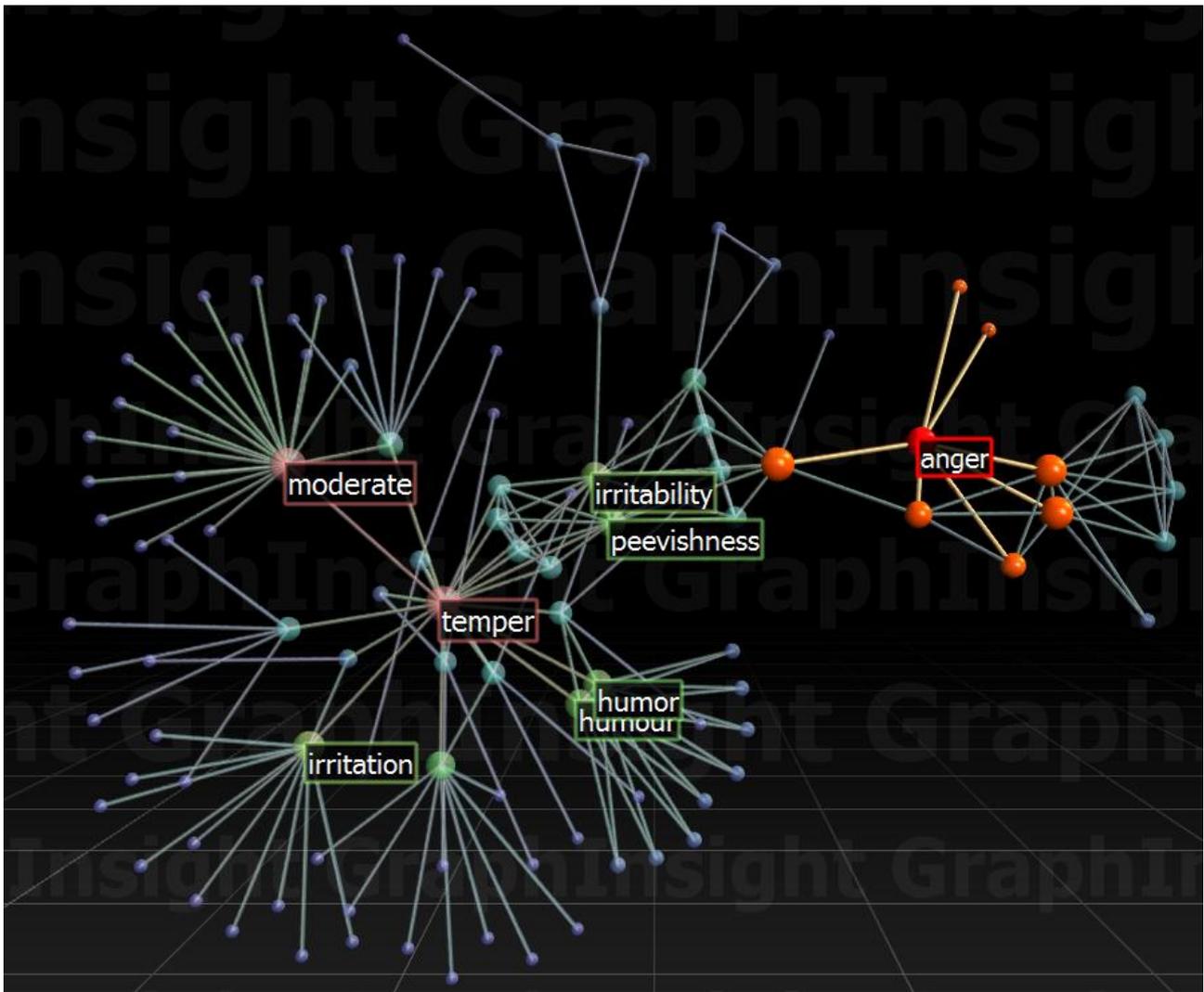


Figura 8.6. La rete del termine "anger" a livello 5.

Composta da 109 nodi e 267 archi, la rete si sviluppa ancora sulla parte sinistra dell'immagine, a partire dal termine "temper". Si noti che la coppia di termini "humor"- "humour" può essere la base della crescita di una particella emergente, che sarà analizzata nei paragrafi successivi. Riportiamo solo per il termine "amger" la rete trovata per il livello 6, anche se tale rete non viene considerata per il calcolo delle statiche dell'esperimento qui condotto. Il livello 6 è considerato solo a fini dimostrativi della esplosione di pattern e di organizzazioni nello spazio semantico.

Il livello 6 presenta una rete di termini rappresentata in Figura 8.7.

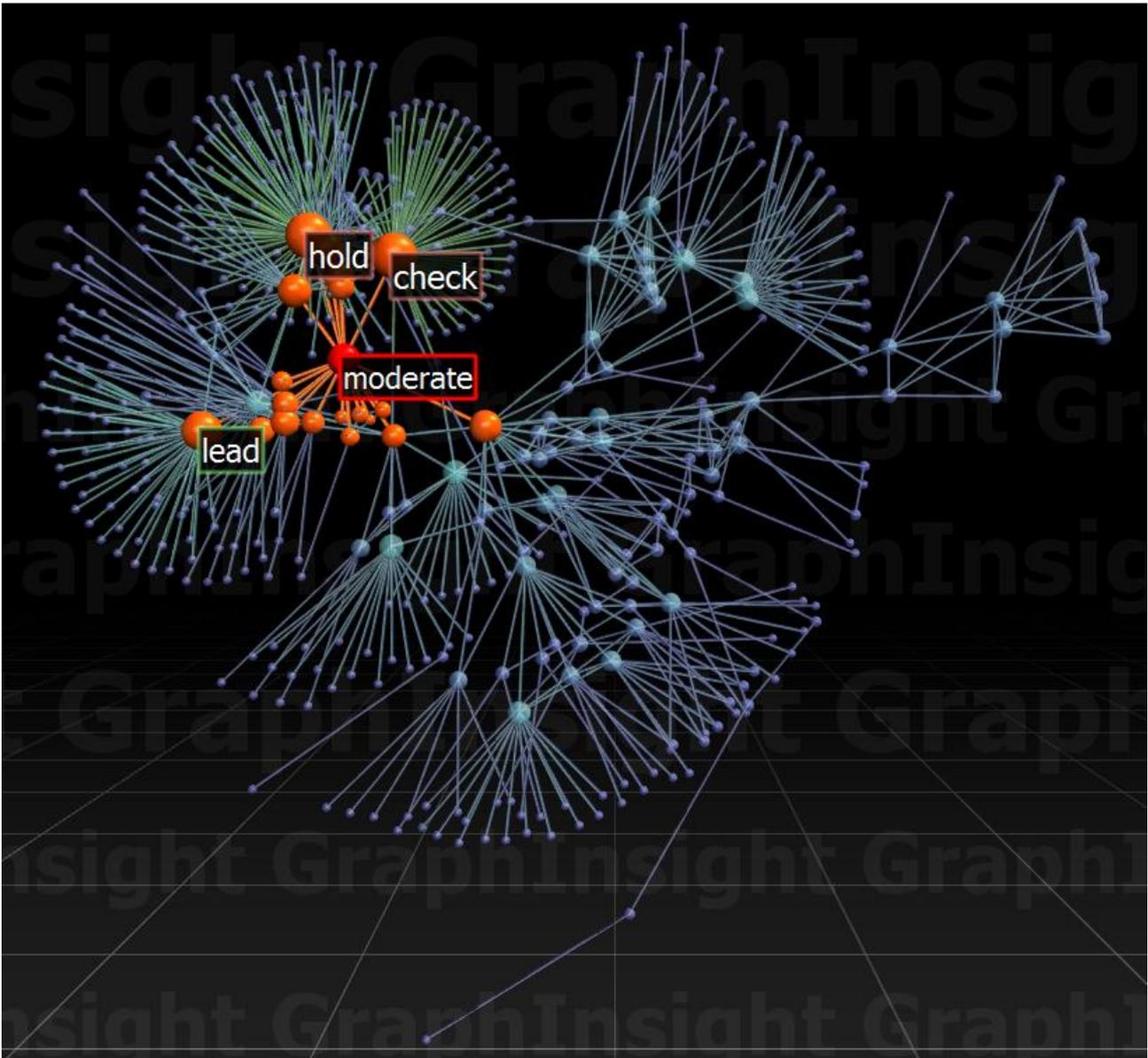


Figura 8.7. La rete del termine “anger” a livello 6.

Formata da 473 archi e 920 nodi, la rete del livello 6 è una struttura molto articolata, che sviluppa sempre a sinistra dell’immagine la direzione di crescita. Il termine “*temper*”, infatti, diventa uno degli snodi della crescita, al quale si aggiungono tre termini molto importanti come “*check*”, “*hold*” e “*lead*”. Questi termini rappresentano 3 verbi. Nel frattempo la coppia “*humor*”-“*humour*” si è articolata per formare una particella di 12 nodi, con interessanti simmetrie.

“Disgust”

Sempre usando come algoritmo di ricerca R_8 : Overview, analizzando nel dettaglio i processi di crescita per questo termine emozionale, si osserva che la rete a livello 1 possiede solo 7 termini (Figura 8.8a). A livello 2,

compare un termine con alto grado che è “*repel*”, mentre i termini del livello 1 iniziano a connettersi tra loro (Figura 8.8b).

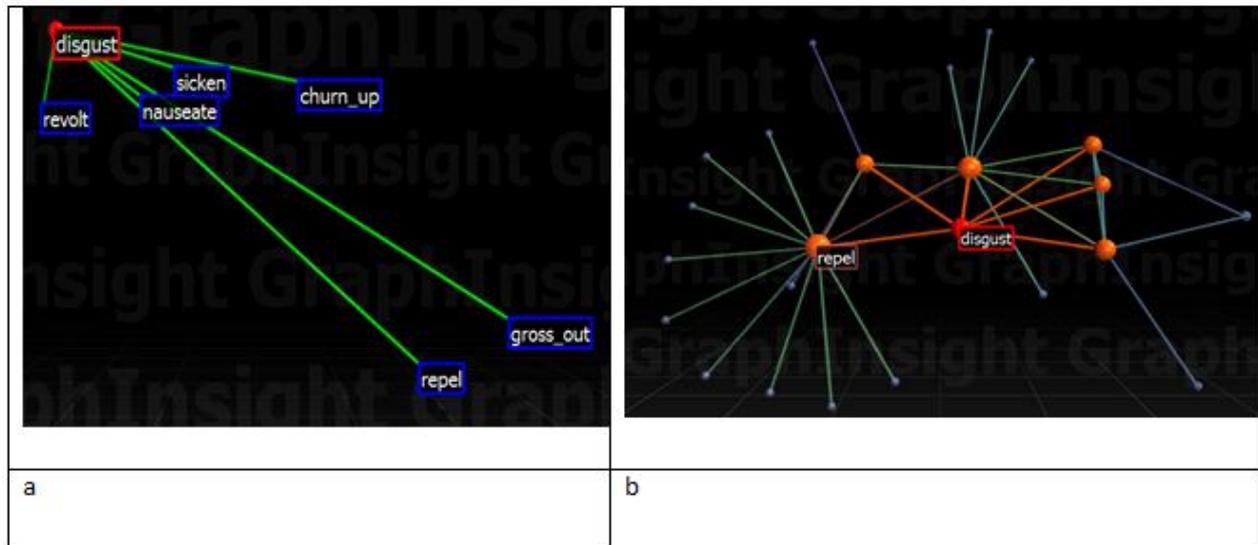


Figura 8.8. (a) La rete del termine “*disgust*” a livello 1. (b) La rete del termine “*disgust*” a livello 2. Compare il termine “*repel*”.

A livello 3, la crescita principale avviene in relazione al termine “*repel*” (Figura 8.9). Compare anche un secondo termine con alto grado: “*drive*”. I nodi nell’intorno di “*repel*” si strutturano per formare una nuova particella emergente, che sarà discussa nei paragrafi successivi.

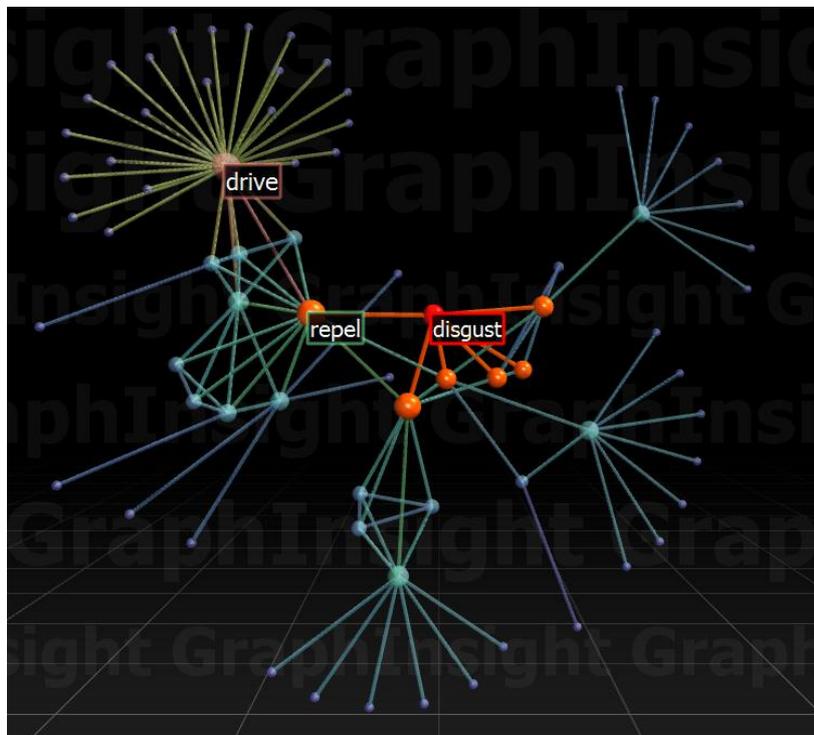


Figura 8.9. La rete del termine “*disgust*” a livello 3.

A livello 4 (Figura 8.10), la presenza di altri verbi, oltre al già menzionato “drive”, porta a un’espansione rilevante della rete, che continua anche a livello 5 (Figura 8.11). Nessun pattern a questo punto risulta distinguibile.

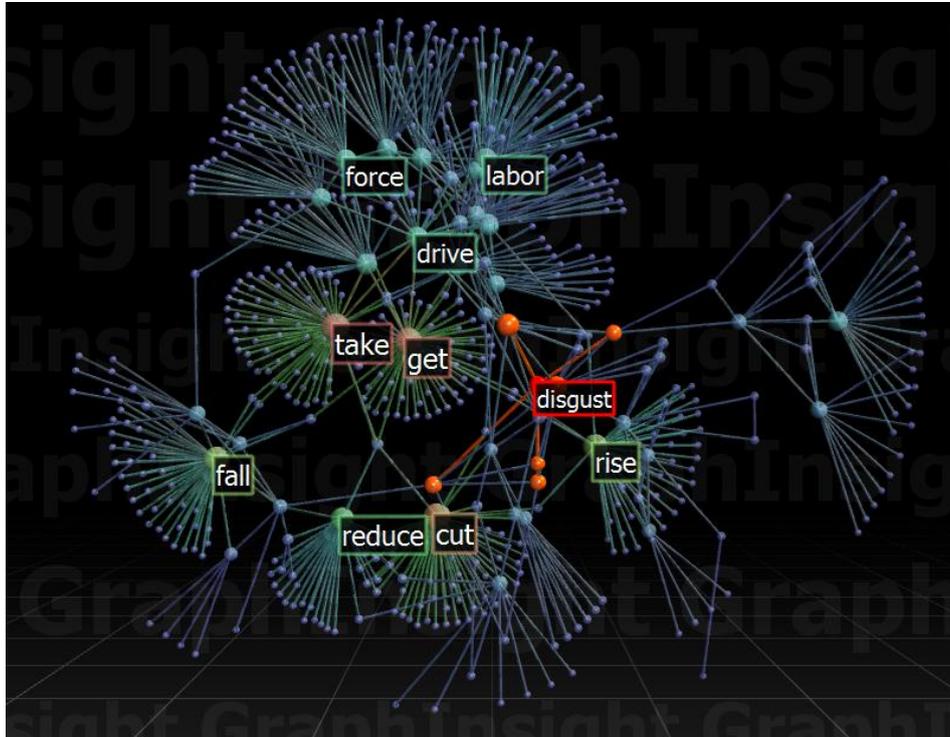


Figura 8.10. La rete del termine “disgust” a livello 4.

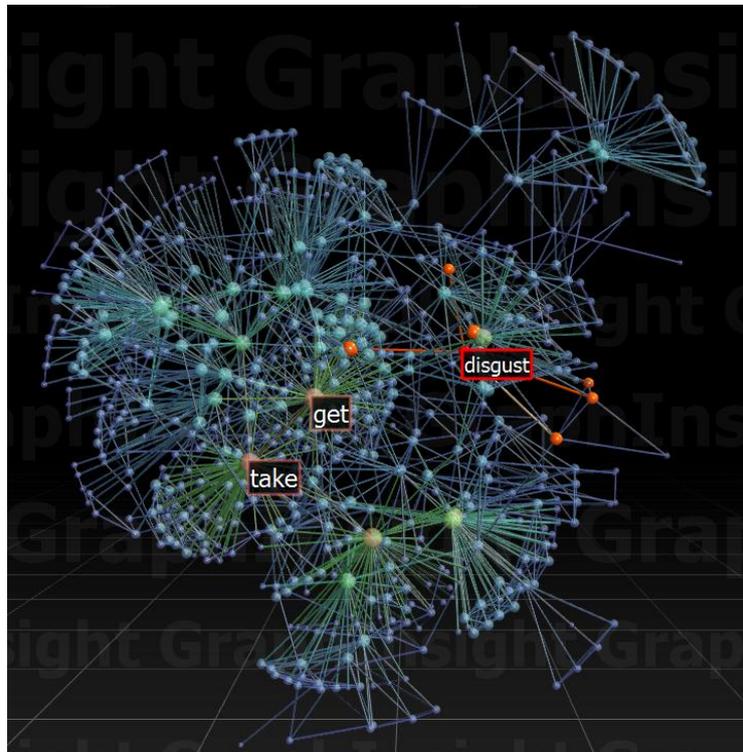


Figura 8.11. La rete del termine “disgust” a livello 5.

Queste reti manifestano l'emergenza di una moltitudine di particelle, differenti per forma e dimensioni, di cui si discuterà nei paragrafi successivi. Si rileva, ancora una volta, che i verbi permettono crescite veramente espansive, vista il loro alto grado e la natura delle loro connessioni.

Esempi di crescita a diversi livelli per altri termini emozionali

Nella Figura 8.12. si osserva la rete del termine "expectation" a livello 4, considerando anche gli accoppiamenti tra i termini. La rete risultante è particolarmente complessa. Si individuano nei verbi "see" e "take-in" i termini col grado più alto, che danno origini a crescite rilevanti, visto il grado di connessione che tali termini hanno. Si nota una rilevante formazione di particelle emergenti.

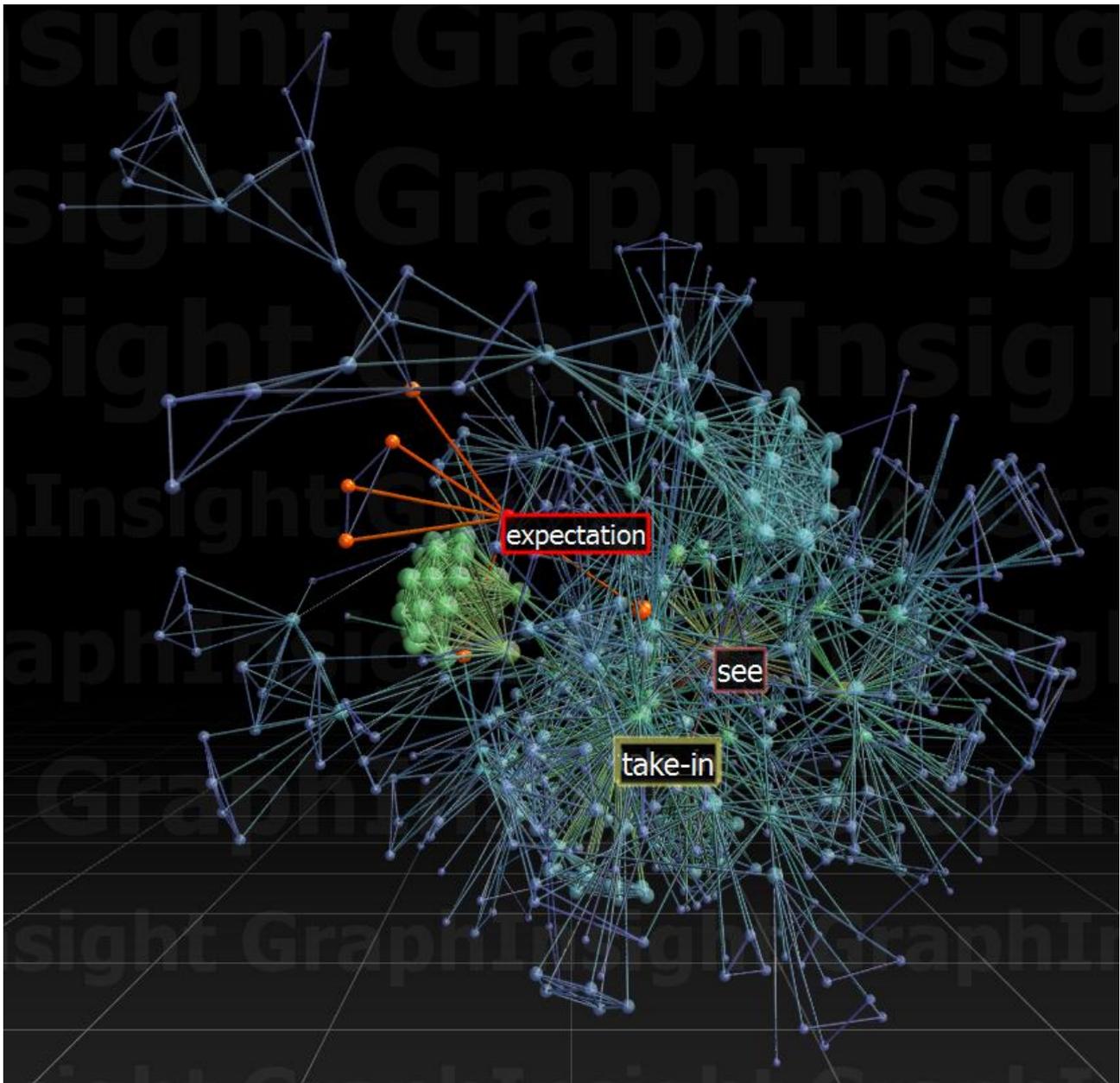


Figura 8.12 La rete del termine "expectation" a livello 4.

In Figura 8.13, la rete del termine "fear", a livello 3. Anche per questo termine, è stato considerato l'accoppiamento dei nodi. In questo caso, si osserva che i termini di grado maggiore sono "dread", "concern" e "care", sui si sviluppano due particelle di particolare complessità, discusse nei paragrafi successivi. L'immagine è riportata proprio per la sua capacità di far capire la formazione di queste strutture, evidenti nella loro complessità anche senza strumenti di analisi sofisticati.

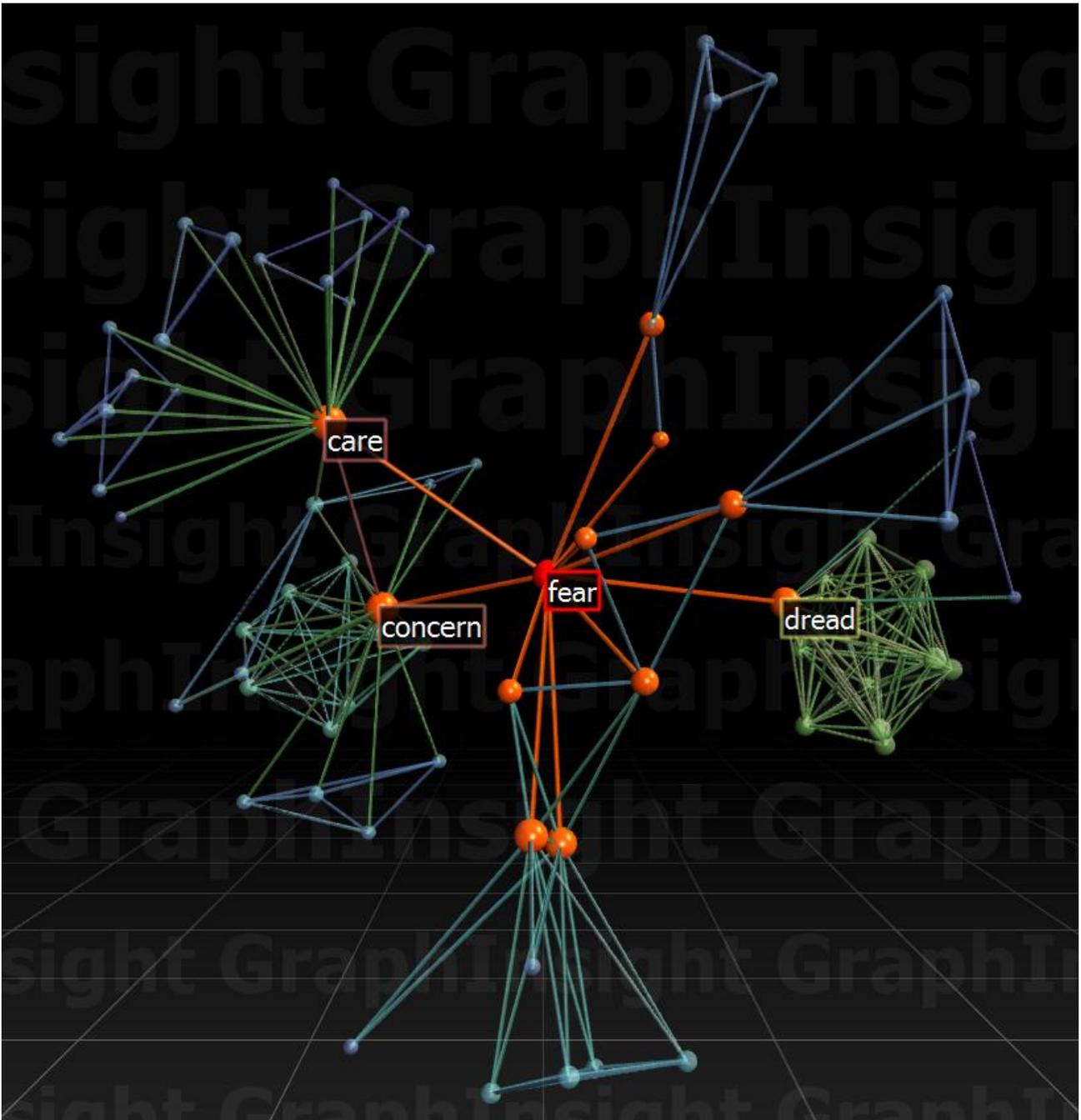


Figura 8.13 La rete del termine "fear" a livello 3.

Nella Figura 8.14, è possibile osservare la rete del termine “joy” a livello 5. I termini con grado massimo sono “love” (attorno al quale si sviluppa una particella emergente molto interessante, proprio per la sua doppia natura di nome e di verbo), insieme con altri due termini che sono “carry” e “catch”.

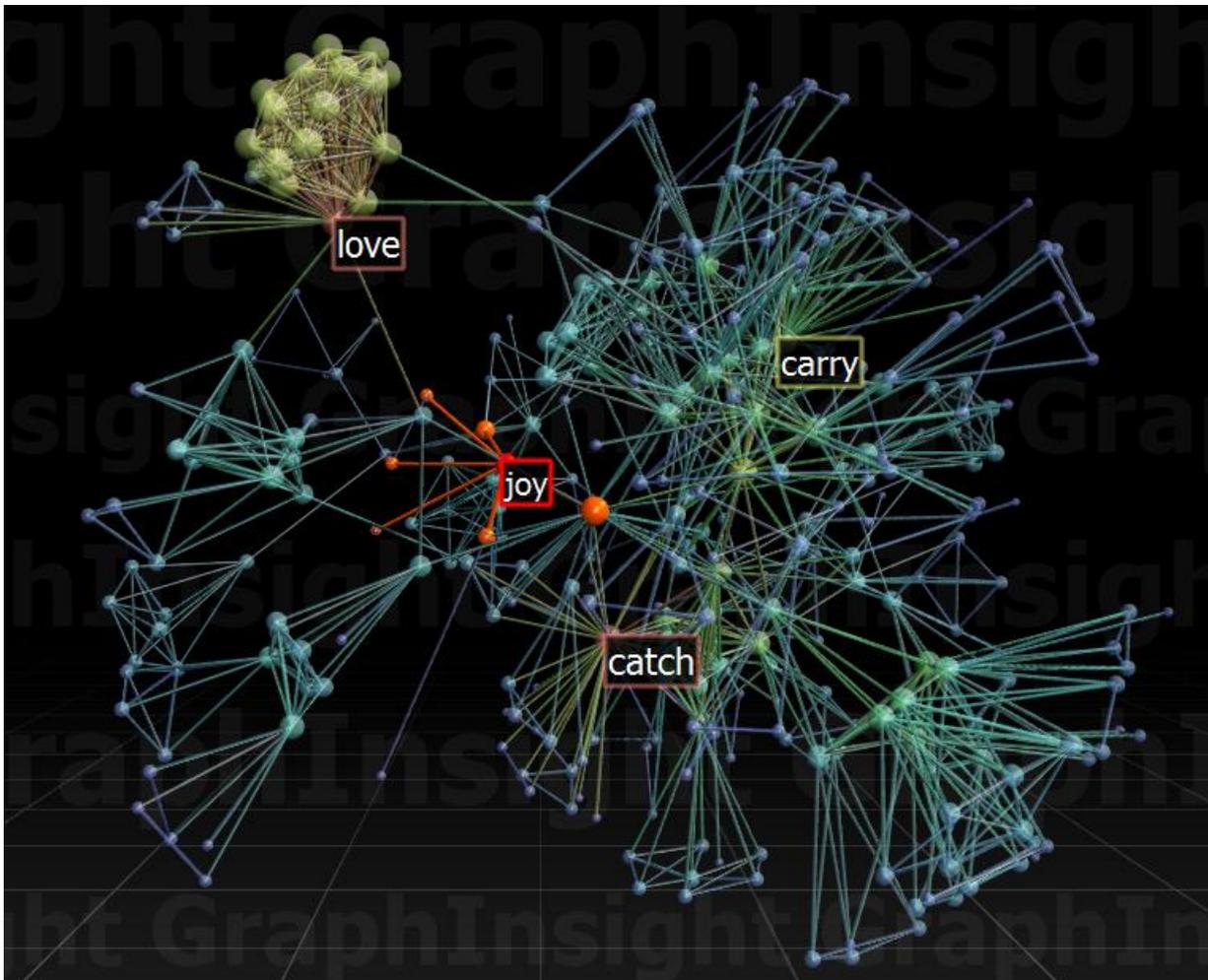


Figura 8.14 La rete del termine “joy” a livello 5.

Nella Figura 8.1, si riporta la rete per il termine “sadness”, a livello 6 di evoluzione. Anche in questo caso, la rete e la sua visualizzazione è fornita a scopo di visualizzazione, non considerandola per i fini del computo statistico di tale esperimento. Contrariamente alle altre reti, questa struttura cresce molto lentamente e ha un numero molto limitato di nodi e di archi, anche ai livelli di crescita 5 e 6. Alcuni nodi ad alto grado sono “sour” e “gravity”. Tale rete dimostra che le connessioni nello spazio semantico sono specifiche per ogni termine emozionale.

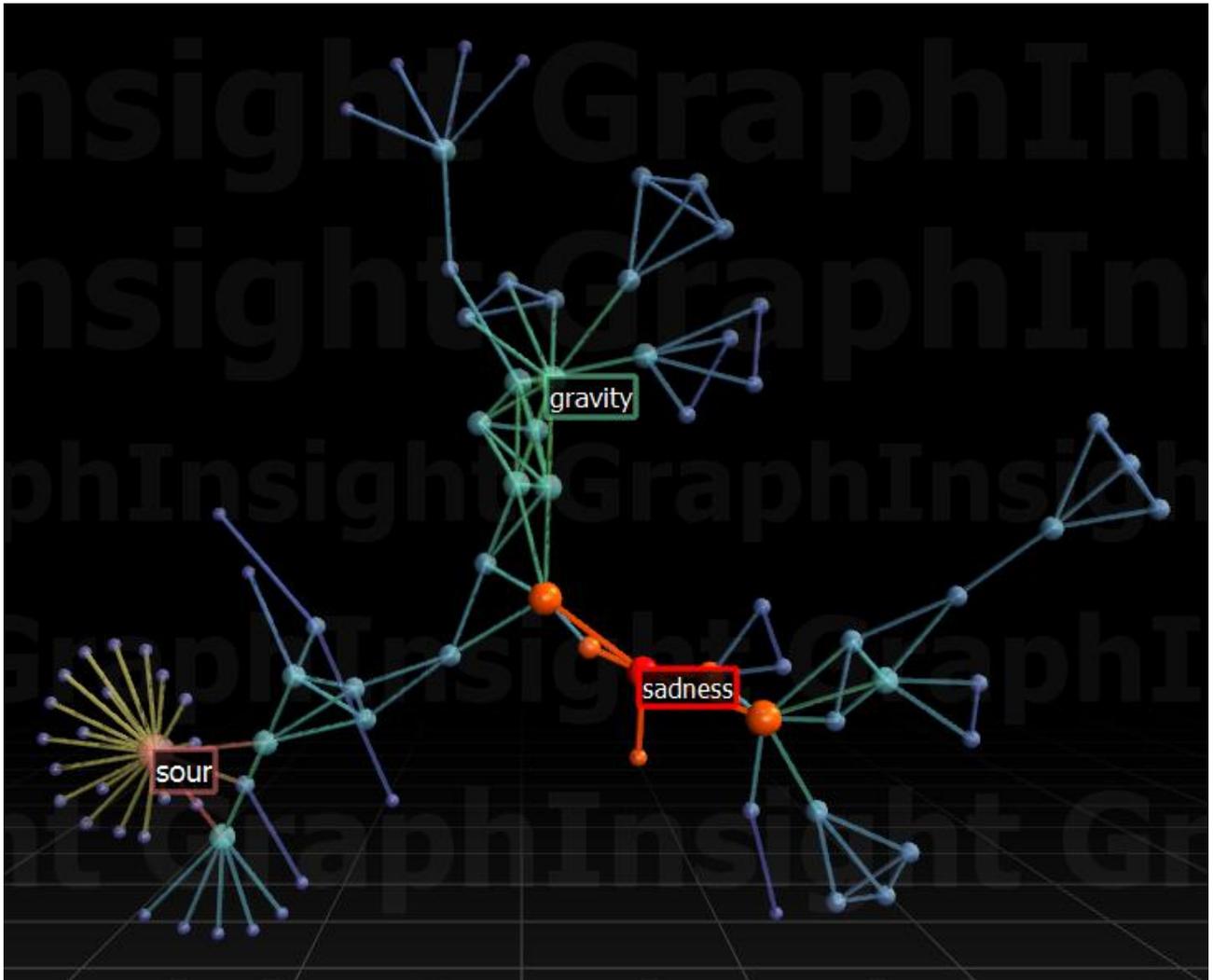


Figura 8.15. La rete del termine "sadness" a livello 5.

Si conclude questa prima fase di osservazione qualitativa, riportando la rete del termine "surprise", livello 6, di Figura 8.16. Anche per questo termine, il livello non è considerato nel computo statistico.

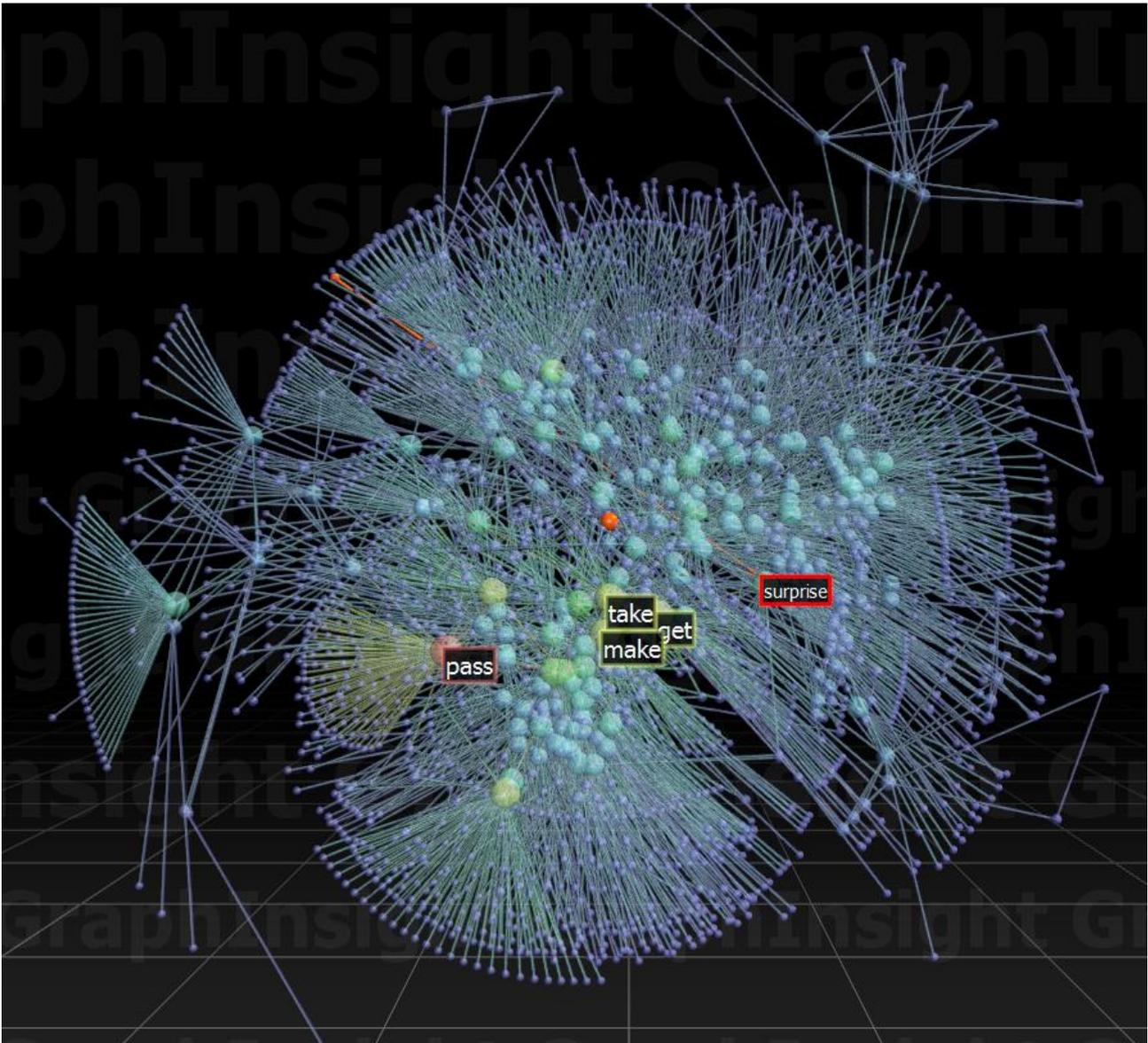


Figura 8.16 La rete del termine surprise a livello 6.

La rete è molto connessa visto l'alto numero di nodi (2179) e di archi (4225). I termini con grado più alto sono "pass", "get", "make" e "take". La rete possiede un grande numero di particelle emergenti, grazie all'alto numero di connessioni.

8.4.2. Statistiche di rete per i termini emozionali

Per ogni termine, per ogni rete risultante da ogni livello di crescita, sono state calcolate le seguenti grandezze statistiche:

1. Numero totale di nodi,
2. Numero totale di archi,
3. Grado medio,

4. Coefficiente di clustering,
5. Efficienza,
6. Lunghezza media,
7. Diametro della rete,
8. Densità.

In Tabella 8.2, Si riportano i valori statistici per il campione di reti esaminato, per tutti i termini emozionali, per tutti i livelli.

Nome rete	Nodi	Archi	Grado	Clustering	Efficienza	Cammino	Diametro	Densità
acceptance_all_overview_1.net	8	7	1,75	0	0,125	0,466667	1	0,125
acceptance_all_overview_2.net	14	28	4	0,272653	0,338828	1,745763	3	0,153846
acceptance_all_overview_3.net	18	40	4,444444	0,299471	0,373529	2,375	5	0,130719
acceptance_all_overview_4.net	32	62	3,875	0,268344	0,215356	3,19661	6	0,0625
acceptance_all_overview_5.net	102	175	3,431373	0,175021	0,100862	3,788842	7	0,016987
anger_all_overview_1.net	7	6	1,714286	0	0,142857	0,461538	1	0,142857
anger_all_overview_2.net	20	33	3,3	0,109964	0,198246	1,823529	3	0,086842
anger_all_overview_3.net	28	109	7,785714	0,621535	0,348214	2,544014	5	0,14418
anger_all_overview_4.net	43	159	7,395349	0,43094	0,272942	3,033634	6	0,08804
anger_all_overview_5.net	109	267	4,899083	0,299273	0,130112	3,725647	7	0,022681
disgust_all_overview_1.net	7	6	1,714286	0	0,142857	0,461538	1	0,142857
disgust_all_overview_2.net	24	48	4	0,204352	0,17814	1,681081	3	0,086957
disgust_all_overview_3.net	73	165	4,520548	0,20516	0,129433	2,964464	5	0,031393
disgust_all_overview_4.net	572	901	3,15035	0,15476	0,03296	4,476106	7	0,002759
disgust_all_overview_5.net	3531	7495	4,245256	0,170911	0,037052	4,695634	9	0,000601
expectation_all_overview_1.net	7	6	1,714286	0	0,142857	0,461538	1	0,142857
expectation_all_overview_2.net	23	36	3,130435	0,151006	0,165679	1,915254	3	0,071146
expectation_all_overview_3.net	98	158	3,22449	0,106879	0,085209	3,146844	5	0,016621
expectation_all_overview_4.net	585	1013	3,463248	0,145229	0,043825	4,266091	7	0,002965

expectation_all_overview_5.net	3028	7342	4,849406	0,204533	0,042623	4,812455	9	0,000801
fear_all_overview_1.net	11	10	1,818182	0	0,090909	0,47619	1	0,090909
fear_all_overview_2.net	72	98	2,722222	0,077192	0,069933	2,293083	3	0,019171
fear_all_overview_3.net	393	769	3,913486	0,160842	0,047517	3,658059	5	0,004992
fear_all_overview_4.net	2178	4778	4,387511	0,167451	0,037388	4,727272	7	0,001008
fear_all_overview_5.net	8228	23521	5,718	0,231	Non calcolato	5,405	13	0,000347
joy_all_overview_1.net	7	6	1,714286	0	0,142857	0,461538	1	0,142857
joy_all_overview_2.net	23	32	2,782609	0,116447	0,161067	1,949153	3	0,063241
joy_all_overview_3.net	86	157	3,651163	0,118421	0,102082	3,070554	5	0,021477
joy_all_overview_4.net	365	851	4,663014	0,21979	0,064722	4,279169	7	0,006405
joy_all_overview_5.net	2159	4734	4,385364	0,169963	0,035737	5,298362	9	0,001016
sadness_all_overview_1.net	6	5	1,666667	0	0,166667	0,454545	1	0,166667
sadness_all_overview_2.net	17	25	2,941176	0,104658	0,202206	1,769912	3	0,091912
sadness_all_overview_3.net	30	70	4,666667	0,430319	0,2409	2,8413	5	0,08046
sadness_all_overview_4.net	39	121	6,205128	0,521517	0,274375	3,662426	7	0,081646
sadness_all_overview_5.net	55	156	5,672727	0,402597	0,213963	4,460435	9	0,052525
surprise_all_overview_1.net	3	2	1,333333	0	0,333333	0,4	1	0,333333
surprise_all_overview_2.net	8	9	2,25	0	0,252976	1,310345	3	0,160714
surprise_all_overview_3.net	43	51	2,372093	0,0079	0,08624	2,282322	4	0,028239
surprise_all_overview_4.net	360	522	2,9	0,155648	0,038863	3,416155	5	0,004039
surprise_all_overview_5.net	2179	4225	3,877926	0,168336	0,038807	4,601851	7	0,00089

Tabella 8.2. La Tabella riporta tutti i valori statistici per il campione di reti esaminato, per tutti i termini emozionali, per tutti i livelli.

Analizzando i dati dell'esperimento, si nota che il rapporto fra nodi varia da termine a termine, con delle similitudini fra termini. Per esempio, un dato costante è che il rapporto fra nodi e archi, per alcuni termini, si mantiene contenuto fino a livello 4, per poi esplodere successivamente a livello 5 e 6 (Figura 8.17). La Figura 8.18, riporta la curva di evoluzione del grado, man mano che si realizzano le crescite delle reti

linguistiche. Si nota un aumento del grado man mano che la rete si evolve, con picchi sui livelli alti di interrogazione.

La Figura 8.19 illustra gli andamenti dei valori Clustering e di Efficienza per le reti linguistiche. Come è possibile notare, i valori si innalzano in corrispondenza di alti livelli di ricerca nella rete WordNet. Si nota infatti che, man mano che la rete cresce, la sua efficienza aumenta, per poi diminuire.

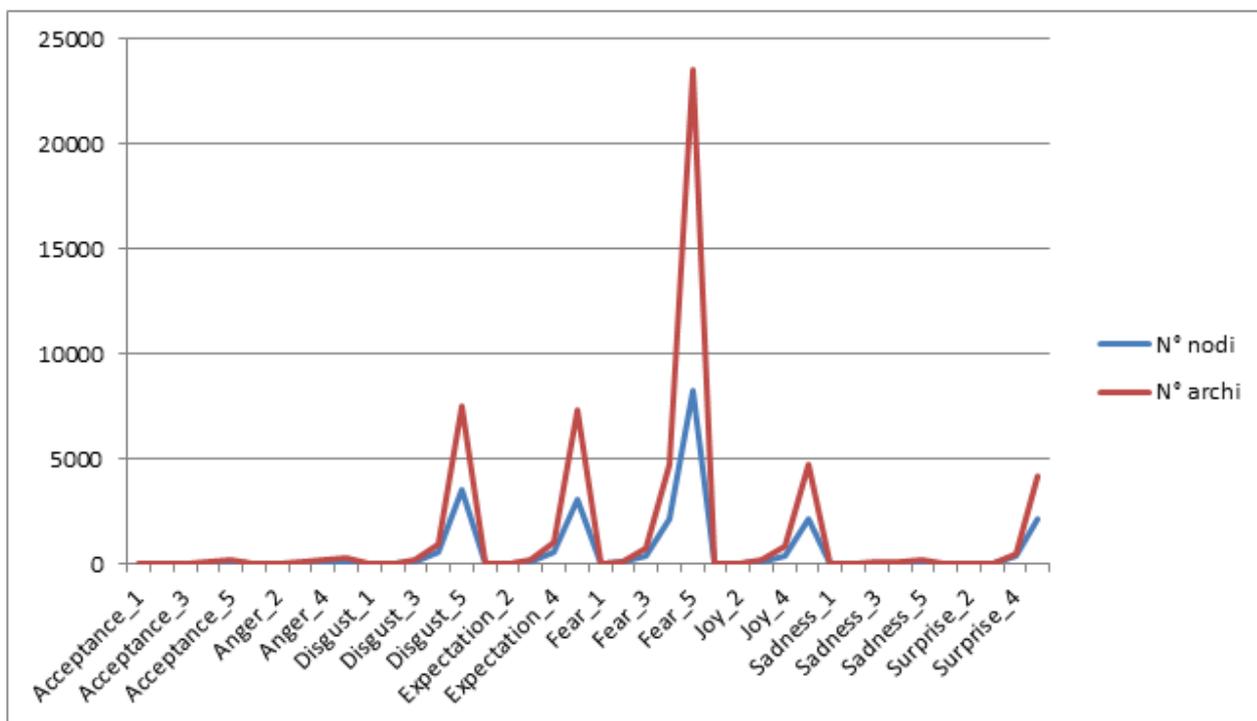


Figura 8.17. Andamento del rapporto fra numero di nodi e numero di archi per tutti i termini emozionali, per i diversi livelli esplorati. Si noti che per alcuni termini "disgust", "expectation", "fear", "joy" il numero complessivo di nodi e di archi è molto basso e che ci sono relazioni fra numero di nodi e numero di archi. Si noti che il termine "fear" è in assoluto quello che presenta il maggior numero di nodi e archi, per tutti i livelli considerati.

Anche il grado, nelle singole reti si distribuisce in relazione ad alcuni termini che hanno un rapporto molto alto tra nodi e archi, tra cui "disgust", "expectation", "fear", "joy" e "anger".

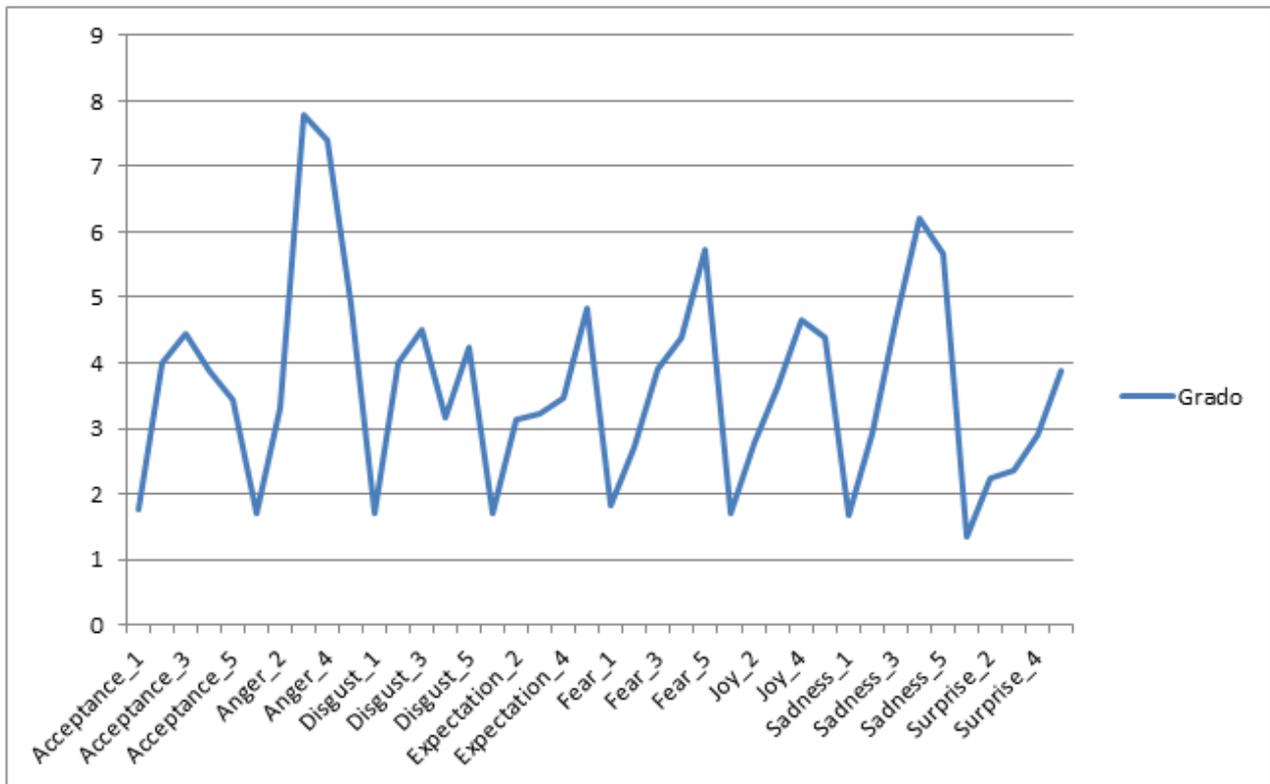


Figura 8.18. Andamento del grado per ogni livello di crescita delle reti linguistiche.

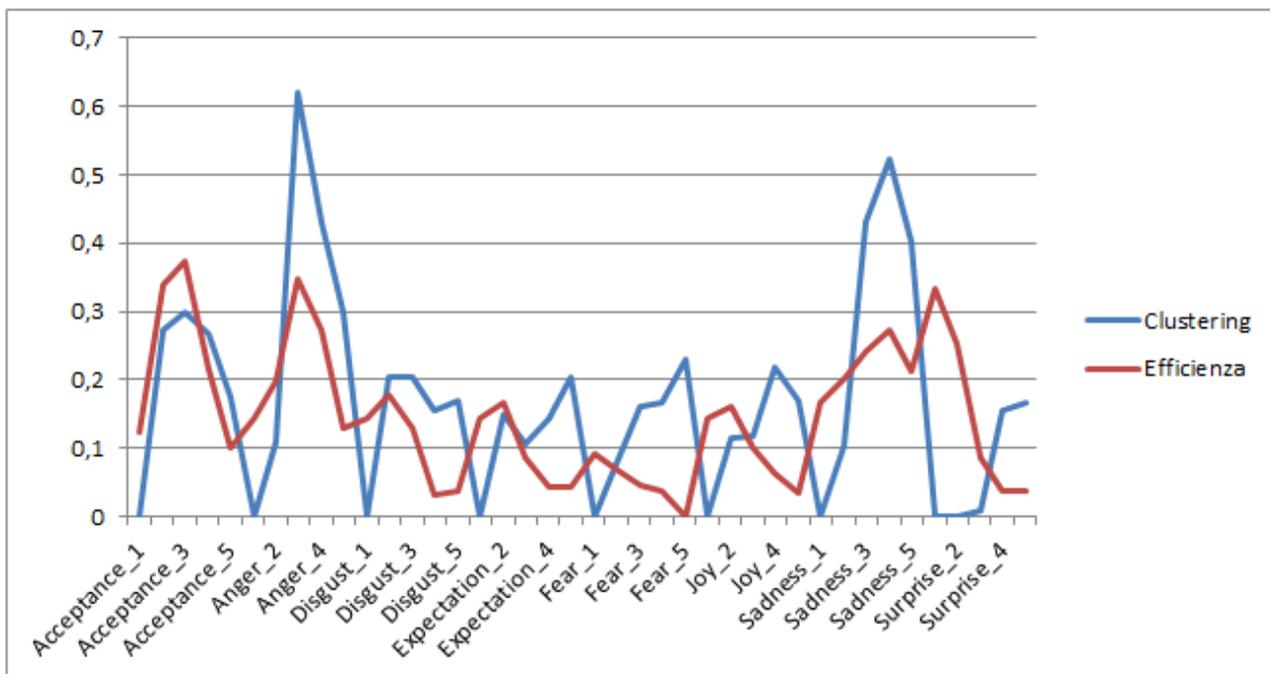


Figura 8.19. andamento delle curve di Clustering e di Efficienza di rete. Come è possibile notare, il clustering ha dei picchi in corrispondenza di "anger" e "sadness", in corrispondenza ad alti livelli di ricerca nella rete WordNet. L'efficienza delle rete sembra avere invece un andamento correlabile solo in parte con l'aumento dei livelli alti di ricerca.

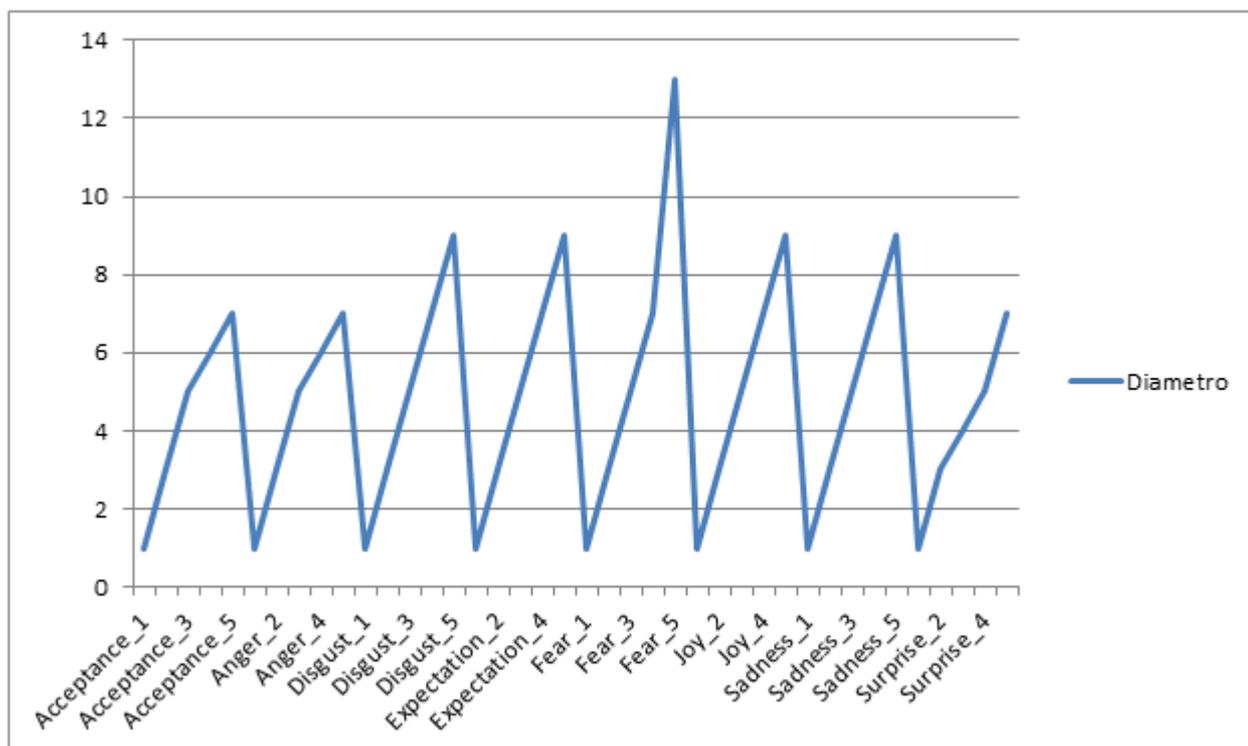


Figura 8.20. Diametro delle reti linguistiche durante l'evoluzione. Il diametro cresce in corrispondenza dell'aumento dei nodi e dei legami.

Un altro elemento importante è il diametro della rete, il cui andamento è riportato in Figura 8.19. Il diametro cresce in corrispondenza dell'aumento della dimensione delle reti linguistiche, raggiungendo il picco più alto di 13 per il termine "fear".

8.5. Particelle emergenti

Uno dei fenomeni più celebrati e onnipresenti dei sistemi complessi è l'emergenza di strutture auto-organizzanti (Langton, 1995, Bilotta & Pantano, 2010). Nei sistemi dinamici discreti come gli Automi Cellulari, queste particelle sono note come *glider* (Wolfram, 1984) e sono responsabili dello stivaggio e del trasporto di informazioni (Langton, 1986, 1990). Analogamente, la crescita delle reti linguistiche produce strutture di complessità crescente, indicando che, diversamente da quanto riportato in letteratura, esistono ben più di tre modalità di crescita delle reti.

In questa parte dell'esperimento, dopo aver discusso i dati statistici delle reti sui termini emozionali, si presenta un catalogo non esaustivo di *Particelle Emergenti* (PE), che rendono conto delle numerose manifestazioni organizzative, rilevate in ambito linguistico. Com'è stato già detto, le 4 categorie principali sono:

- a. *grafi perfetti*, indicati con la sigla PE_di tipo_A,
- b. *grafi imperfetti a causa di rotture di simmetrie*, indicati con la sigla PE_di tipo_B,

- c. *grafi con simmetrie complesse*, indicati con la sigla PE_di tipo_C,
- d. *e particelle complesse, risultanti dall'unione di particelle più elementari o dall'unione di particelle appartenenti alle categorie a., b., e c.*, indicati con PE_di tipo_D.

Ogni particella, è preceduta dal nome emozionale di partenza e seguita da un numero, che indica l'ordine di comparsa della stessa, nella crescita della rete linguistica. Un apparato iconografico è associato ad ogni particella: 2 immagini che presentano gli elementi costitutivi della particella stessa in due dimensioni, una con le etichette linguistiche e una senza, 1 immagine tridimensionale. Tutte le particelle trovate sono riportate nell'Appendice A. Nella trattazione che segue, saranno forniti solo alcuni esempi illustrativi.

8.5.1. Grafi Perfetti

Introdotti nel Capitolo 2, i *grafi perfetti* sono composti da n nodi connessi tra loro, in modo che ogni nodo sia connesso a tutti gli altri. Queste particelle possiedono sia simmetrie di riflessione che di traslazione. Lo studio specifico di queste simmetrie va al di là degli obiettivi di questo capitolo. Per maggiori dettagli, si rimanda al testo classico di Coxeter (1974).

Per motivi di chiarezza espositiva, si afferma solo che i *poligoni* hanno due dimensioni, i *poliedri* sono oggetti tridimensionali (il cui volume è racchiuso in un numero finito di poligoni piani), i *politopi* sono oggetti a più dimensioni (4, 5, n dimensioni), detti *policori*. Il concetto di politopo estende il concetto di *poliedro* a più dimensioni. Le particelle emergenti che sono state trovate appartengono alcune alla classe dei *poliedri*, altre ai *politopi*.

In Figura 8.21 sono visualizzati due grafi perfetti per il termine "*anger*": *angerPE_A1* e *angerPE_A2*. La prima particella è un solido platonico che si presenta come una piramide a base quadrata. Se un'altra piramide uguale è fissata alla base della particella, il risultato è a sua volta un solido platonico, detto *ottaedro*. La particella *angerPE_A1* si può rappresentare come un oggetto in uno spazio 4-dimensionale, noto come *4-simplesso*.

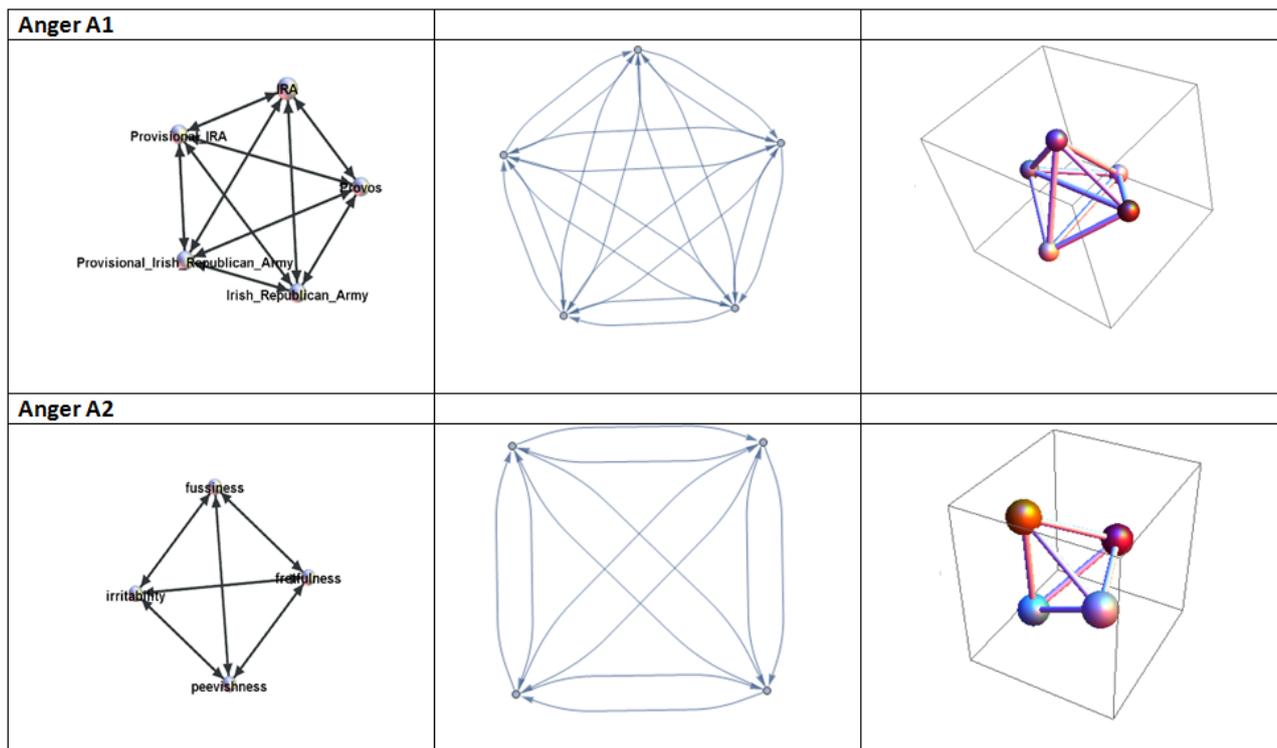


Figura 8.21. Due particelle complesse che appartengono al termine "anger". E' possibile notare che la particella classificata con angerPE_A1 è formata da 5 nodi, angerPE_A2 è formata invece da 4 nodi.

angerPE_A2 presenta le stesse simmetrie del tetraedro (piramide triangolare), che è uno dei *poliedri regolari convessi*: è un *3-simplex*.

Generalizzando questi concetti, si può notare che angerPE_A3 è un *7-simplex* mentre angerPE_A4 è un *6-simplex*, joyPE_A1 è un *7-simplex*. Tutti questi iper-solidi sono *politopi regolari* e la loro naturale collocazione è in spazi a 6, e 7 dimensioni rispettivamente (Figura 8.22).

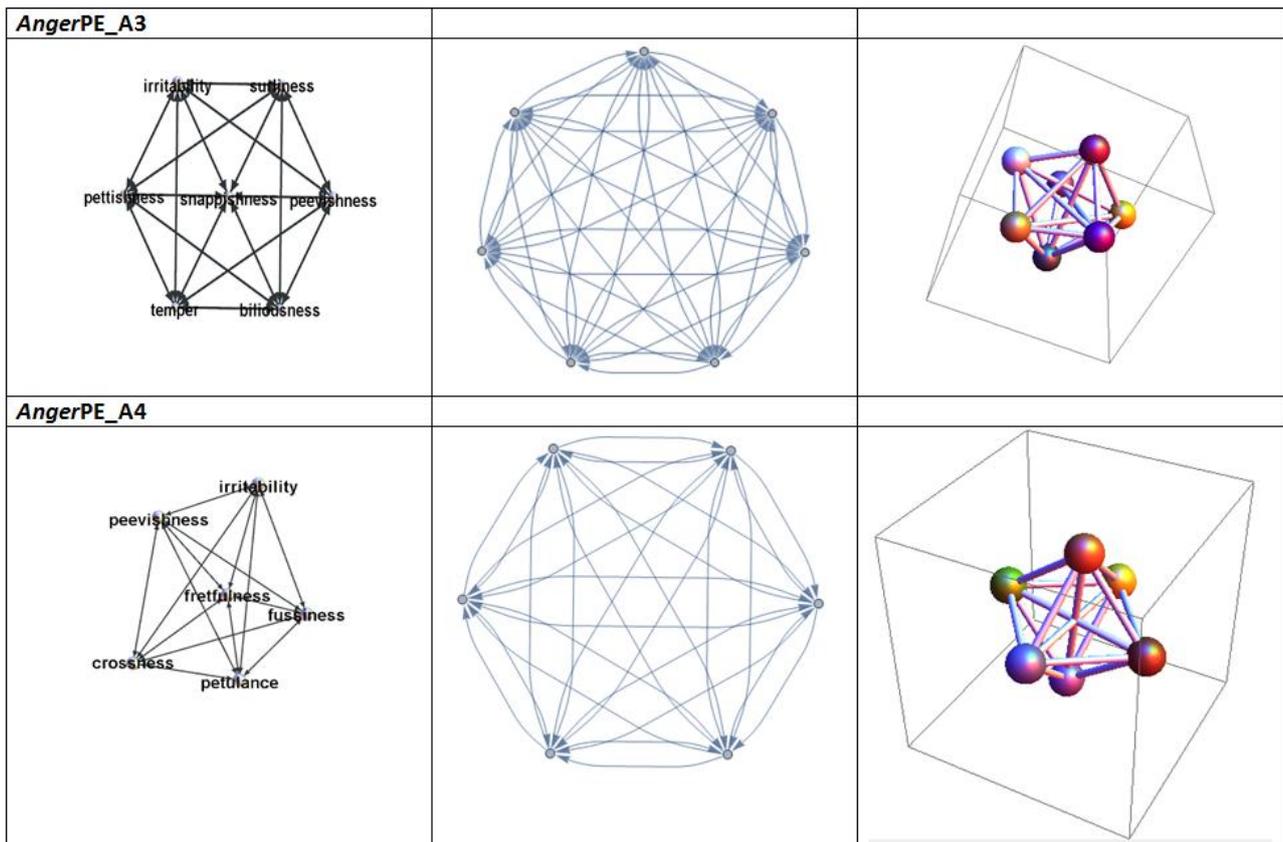


Figura 8.22. Due particelle complesse, prodotte dal termine “anger”. Le particelle sono politopi regolari e vivono in spazi a 6, e 7 dimensioni rispettivamente.

8.5.2. Rotture di simmetrie

Questa seconda categoria di particelle è simile alla precedente, solo che alcuni archi sono stati rimossi: l’eliminazione di archi comporta il venir meno di alcune simmetrie. Questo fenomeno è anche noto come *rottura di simmetria*. Questo fenomeno, molto complesso nella sua natura organizzativa, prevede anche che due rotture di simmetrie, opportunamente collocate, possano ripristinare una simmetria (Stewart & Golubitsky, 1995; Stuart, 2007).

In Figura 8.23 sono riportati due esempi di rottura di simmetria. Sul 5-simplesso *angerPE_B1*, e sul tetraedro *disgustPE_B1* 3, per esempio, si sono verificate 4 e 3 rotture di simmetrie rispettivamente. Sul 5-simplesso *disgustPE_B2*, invece, si sono verificate molte rotture di simmetria, pur ripristinando altre simmetrie di rotazione (120°) e di riflessione. Anche *disgustPE_B3* si sono verificate varie rotture di simmetria rispetto al 4-simplesso, ma altre simmetrie (di riflessione) vengono ripristinate. Anche *joyPE_B1* presenta rotture di simmetria sul 5-simplesso. Anche in questo caso, alcune simmetrie di riflessione vengono ripristinate (Figura 8.23).

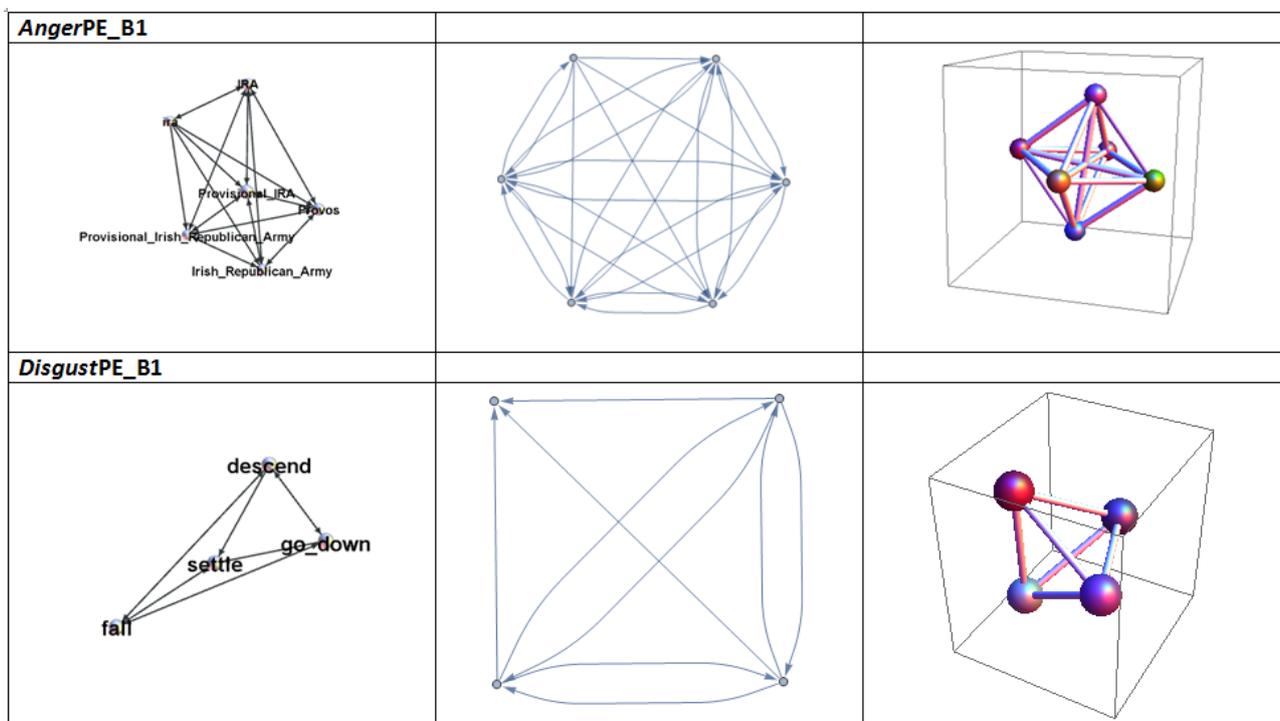


Figura 8.23. Particelle in cui si è verificata una rottura di simmetria.

8.5.3. Altre Simmetrie

Questa categoria di particelle evidenzia la presenza di simmetrie non convenzionali di grande interesse rispetto ai casi precedenti. Molte di queste particelle presentano una coppia di elementi unita da un doppio legame che punta ad altri nodi collocati in un piano ortogonale all'asse della coppia. Ad esempio, *angerPE_C1* presenta una struttura pentagonale ortogonale all'asse coppia. La simmetria di maggior evidenza della struttura è la rotazione di 72° , come nel pentagono tradizionale. Analogamente *angerPE_C3* presenta un ottagono nel piano ortogonale all'asse della coppia, con una simmetria di rotazione di 45° . Esempi di tali particelle sono riportate in Figura 8.24.

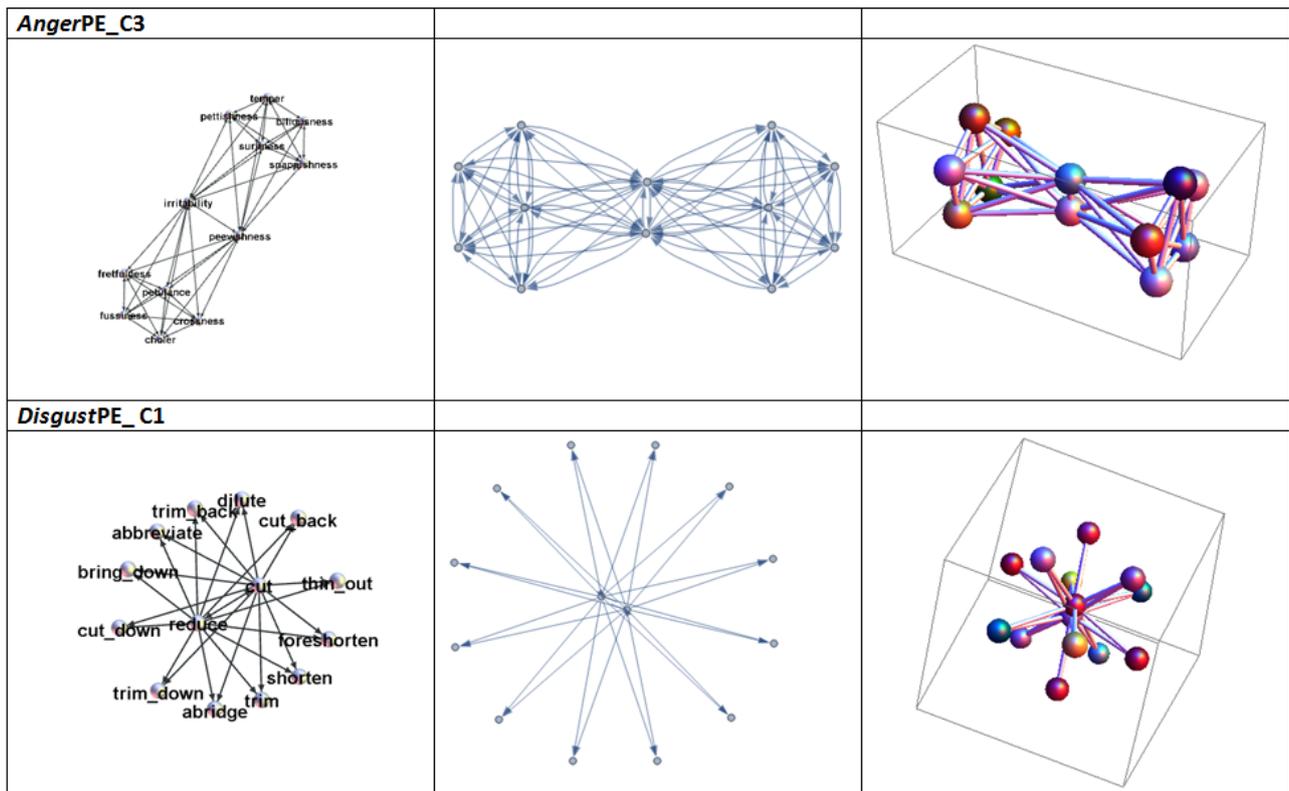


Figura 8.24. Altri tipi di simmetrie si strutturano in modo creativo nel regno della linguistica e del significato. In particolare questa raccolta di immagini presenta ben 5 strutture simmetriche.

Molto interessante infine è la simmetria di *angerPE_C2*, dove due coppie di pentagoni su piani paralleli all'asse della coppia, si riflettono in modo speculare. Si noti che ciascuna coppia di pentagoni forma un 4-simplesso, mentre, se incorporano anche la coppia centrale, formano un 6-simplesso. In questa particella, si può notare la straordinaria capacità compositiva delle forme, dove forme elementari si uniscono per produrre forme più complesse, mantenendo però alcune simmetrie (di riflessione in questo caso per la forma risultante).

8.5.4. Composizione di Particelle

In quest'ultima sezione, un gruppo di particelle complesse emergenti da particelle più semplici che si combinano fra loro in una incredibile varietà sono state raccolte, evidenziando l'emergenza di simmetrie, la rottura di simmetrie e altri tipi di fenomeni, ecc. Questa rassegna di fenomeni fornisce informazioni sulle straordinarie capacità compositive della struttura organizzativa e semantica della lingua, in un rincorrersi di motivi e pattern, ed evidenziando la natura tipica di una grammatica generativa, che richiama fortemente il linguaggio della chimica e la formazione delle macromolecole biologiche. Tutte le particelle che da qui in poi sono descritte si ritrovano in appendice. Per concludere questo paragrafo infatti, si riportano solo gli esempi di Figura 8.25, che visualizzano le particelle *AngerPE_D1*, *AngerPE_D2*.

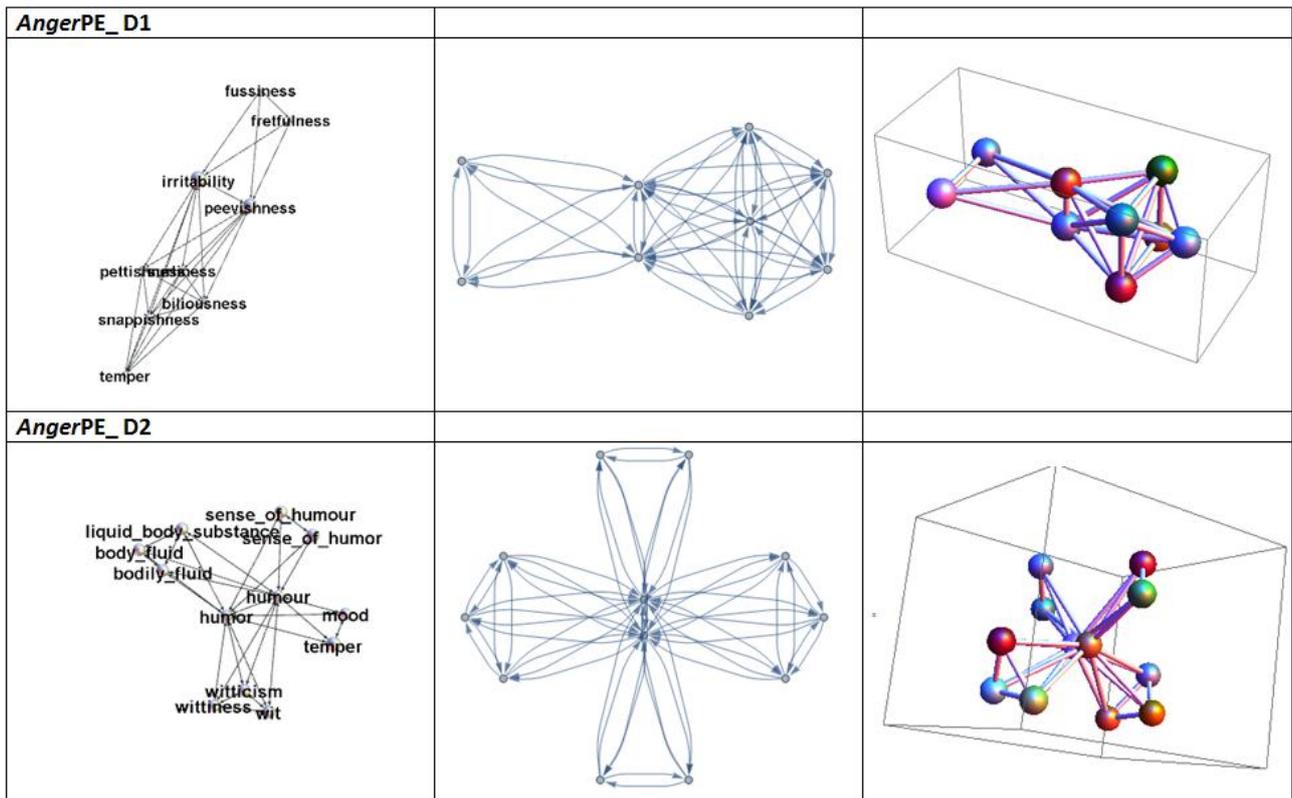


Figura 8.25. Altri tipi di simmetrie si rintracciano nelle seguenti particelle AngerPE_D1, AngerPE_D2.

Una variante di questa struttura si trova in *disgustPE_D3*, dove, rispetto all'asse della coppia troviamo un 3-simplesso ed un triangolo. Anche in questo caso, una doppia rottura di simmetria riproduce una simmetria.

Più complessa e di difficile visualizzazione (e rappresentazione) la particella *disgustPE_D4*. Ora parallelamente ad un piano contenente un triangolo si sviluppa da una parte un quadrato in un piano ed un triangolo dall'altra parte, con una rottura di simmetria. La particella *disgustPE_D5* evidenzia l'emergenza di una struttura complessa, partendo da un grafo completo di 3 particelle, di dipanano diversi legami con altre particelle, disposti con vari motivi nello spazio.

Veramente interessante la particella *expectationPE_D1*: attorno all'asse di una coppia di nodi si uniscono due grafi completi, costituiti da un 9-simplesso e da una particella a 3 nodi. Due nodi si collegano in opposizione l'uno rispetto all'altro alla coppia centrale e ad una delle due particelle rispettivamente.

expectationPE_D2 riserva un'ulteriore sorpresa: alla coppia centrale questa volta si collega addirittura un tetraedro da una parte e un 10-simplesso dall'altra, meglio rappresentato su una sfera.

Quasi analoga, la *fearPE_D1* dove un 11-simplesso è unito ad una coppia tramite un nodo.

fearPE_D2 si struttura invece con una catena di particelle, che presentano qualche rottura di simmetria al loro interno, ed una coppia di nodi in opposizione. Anche *fearPE_D3* presenta composizioni di particelle. In questo caso, 7-simplessi, 3-simplessi ed un 4-simplesso con alcune rotture di simmetria sono uniti attraverso un unico nodo. Questi processi compositivi sono chiaramente evidenti anche in *fearPE_D4*, dove simplessi di varie dimensioni, con varie rotture di simmetrie, si uniscono tra loro.

Il resto delle particelle emergenti presentano una vasta gamma di fenomeni, riconducibili a quelli precedentemente descritti e si ritrovano in Appendice.

8.6. Conclusioni

In questo capitolo, dopo aver introdotto alcuni concetti generali sulla semantica e l'approccio psicolinguistico all'organizzazione della lingua in reti semantiche, è stato presentato un approccio originale e innovativo per lo studio della semantica lessicale. Dopo aver esposto e discusso i dati statistici riguardanti alcune caratteristiche della rete WordNet, un esperimento sull'evoluzione delle reti, attraverso uno specifico algoritmo di ricerca, che considera tutte le categorie di termini grammaticali presenti nella rete WordNet, si è fornita una rassegna qualitativa dei pattern di crescita delle reti linguistiche, per 8 termini emozionali. Sono in seguito stati presentati tutti i dati statistici per tutte le reti realizzate, considerando almeno 5 livelli di esplorazione della rete WordNet. Ma la vera novità di questo capitolo risiede nell'aver messo in evidenza che la struttura delle lingue naturali, come altre strutture biologiche e fisiche, contiene una grammatica generativa di crescita. Tale grammatica, saltando tutte le regole finora conosciute e tutti i canoni di buona formazione delle strutture, in modo creativo e esplosivo, associa geometrie con geometrie per comporre particelle emergenti di diverso tipo. Tali strutture sono state classificate in 4 tipi fondamentali, dai grafi completi alle crescite e/o rotture di simmetrie, per livelli sempre più complessi e in molte dimensioni. Ma ancora molto resta da fare per capire la natura profonda di tali organizzazioni.

Si dimostra, in ogni caso, che anche le lingue naturali, come gli altri linguaggi, nonostante manifestino una struttura di superficie diversa dalla musica e dalle dinamiche cerebrali, possiedono una struttura organizzativa per pattern organizzati.

Conclusioni Generali

Questa tesi nasce con l'ipotesi che i linguaggi artificiali e/o biologici condividano le stesse strutture di base. Per dimostrare questo assunto, il linguaggio della musica, le connessioni cerebrali in soggetti umani e un esempio di lingua naturale, desunta da un data base internazionale (WordNet) sono modellizzati matematicamente in modo da poter trasformare ogni elemento, appartenente ad ambiti diversi, in reti. Particolari algoritmi di modellazione sono stati ad hoc sviluppati e incorporati in pacchetti software. In tal modo, file musicali MIDI (gli eventi sonori di una composizione) sono stati trasformati in elementi discreti, rappresentabili matematicamente. Immagini di Risonanza Magnetica (file DICOM), sono stati elaborati attraverso i metodi della Connettistica e della Trattografia, è stato scelto un modello di segmentazione del cervello in aree di interesse (27, 83 o 1000 aree di riferimento, in corrispondenza a reti rispettivamente di 27, 83 e 1000 nodi) e sono state ottenute reti cerebrali. Per la trasformazione in reti dei file DICOM sono stati sfruttati software e algoritmi già esistenti. Campioni di lessico della lingua Inglese sono stati acquisiti dal database WordNet e trasformati in reti per essere analizzati. Anche in questo caso, sono stati creati appositi software per creare i vocabolari, i sistemi computazionali di interrogazione del database, a diversi livelli di complessità. Dopo aver modellato i tre linguaggi, si è passati alla realizzazione della parte sperimentale. Le variabili statistiche che abbiamo usato, per tutti e tre gli ambiti identificano il numero di nodi e il numero di connessioni, il grado, la percentuale di clustering, l'efficienza, il cammino medio, il diametro e la densità, i parametri che caratterizzano le reti di piccolo mondo (SWN) e e l'invarianza di scala (SF).

Per la musica, sono stati trasformati in rete 241 file musicali, organizzati in 13 collezioni, tra le quali, le famose e celebrate Variazioni Goldberg e le due collezioni di preludi e fughe del Clavicembalo Ben Temperato, appartenenti alla quasi totalità del repertorio di musica barocca di J. S. Bach. In particolare, poiché più tracce possono corrispondere ad una singola melodia, sono stati trasformati in reti 698 file musicali.

Le reti derivanti dai file musicali sono state analizzate a livello della singola composizione, a livello delle Collezioni, a livello dell'intero repertorio della musica Bachiana. Per tutti i livelli di analisi realizzati, sono state rilevate strutture di piccolo mondo e invarianza di scala, anche se le leggi di potenza che dovrebbero realizzare l'invarianza di scala si manifestano come leggi di potenza troncate. Questo fenomeno ci suggerisce che, le occorrenze pratiche di dati reali (soprattutto se si pensa che tali dati provengono da file MIDI, che rappresentano eventi fisici reali) producono distorsioni rispetto al modello ideale, creando una divaricazione e manifestandosi con "interessanti variazioni", rispetto alla legge generale.

Diverse sono le conclusioni che si possono trarre, a livello strutturale e a livello funzionale. Sicuramente si può dire che tutte le composizioni possiedono una nota centrale, con più alto grado, attorno alla quale si organizza il resto della struttura. Questo processo si manifesta, resta invariato a livello delle 13 collezioni analizzate e dell'intero corpus di composizioni musicali. Si rileva anche che le note più ripetute all'interno di una melodia sono quelle che presentano il grado più alto e quelle con centralità maggiore. La forte correlazione si mantiene anche considerando la durata temporale della nota, quindi gli intervalli tra le note, dimostrando la presenza di strutture ritmiche condivise, nonostante le variazioni fra le varie partiture. Quindi una organizzazione interna che sussiste a livello profondo, malgrado le possibili variazioni a livello delle differenze della superficie melodica. Sono stati rilevati inoltre una serie di organizzazioni straordinarie, che intercettano le dinamiche musicali armoniche e melodiche, con strutture organizzative a diversi livelli di complessità come le *clique* (raggruppamenti di note con una precisa funzione ritmica o melodica), i *motivi*, con 2, 3, 4, 5, ... *n*, numero di note legate insieme come ritornelli, e i *cicli* (ripetizioni più estese che raccolgono più motivi). Seguendo tali strutture si può rintracciare l'ordito narrativo della musica di Bach, con i suoi andamenti ascendenti e discendenti, ma anche numerose simmetrie musicali come le trasposizioni di ottave, le inversioni melodiche, la ripetizione di motivi, schemi ritmici e melodici, nei quali si ravvisano la strutturazione di simmetrie, in simmetrie di secondo livello con l'emergenza dello straordinario ordito musicale della musica barocca. Tali geometrie forniscono *una corposità di espressione che risulta dalle proprietà strutturali del brano*. Si rileva, quindi, da parte del musicista, la strutturazione di un significato intrinseco alla stessa organizzazione musicale, una grammatica interna fatta di simmetrie organizzative del flusso melodico, che sono poi raccolte e interpretate, con lo stesso "linguaggio cognitivo" (o grammatica se si vuole) dall'ascoltatore. Si evidenzia inoltre, la presenza di modelli di crescita generativi, come avviene per le dinamiche cerebrali e la lingua naturale. Questa struttura, di natura intrinsecamente sintattica nel senso largo, che comprende sicuramente l'accezione di *organizzativo* e *temporale*, fondata su strutture universali innate e biologicamente definite, è la base condivisa tra lingua naturale, dinamiche cerebrali e musica.

I file di risonanza magnetica strutturale (DICOM) di 26 adulti (13 sani e 13 con diagnosi di malattia neurodegenerativa), trattati con i modelli della Connettomica e dalla Trattografia, e trasformati in reti

hanno evidenziato fenomeni interessanti che sono molto simili alle geometrie musicali. Rilevanti sono i risultati statistici per le conclusioni che si possono trarre sulla degenerazione che la malattia provoca a livello dell'architettura cerebrale. Il numero di collegamenti fra nodi, la densità della rete e il grado dei nodi sono considerevolmente più bassi nei soggetti con Parkinson rispetto ai soggetti sani, mentre sono più alti i valori per il diametro, il cammino medio, l'efficienza delle reti, i fenomeni di clusterizzazione. La tendenza si inverte quando si tratta di considerare il diametro, il cammino medio e l'efficienza della rete. Si osserva che il cammino medio è maggiore nei soggetti affetti da PSP rispetto a quelli sani, e vi è una corrispondente diminuzione di efficienza. Le reti presentano organizzazioni tipiche delle reti di piccolo-mondo (SWN), e questa tendenza è più pronunciata in soggetti con PSP. In questo gruppo di soggetti, mentre l'efficienza complessiva e il numero di connessioni delle reti diminuisce, si creano, per compensare gli effetti della neurodegenerazione, reti locali di connessione. Questo fenomeno ci induce a ipotizzare che, nelle malattie neurodegenerative, il decremento di neuroni nelle aree che poi determinano la semeiotica medica della malattia (l'insieme dei segni di interesse clinico che sono indispensabili per la diagnosi), si accompagna a una qualche forma di riorganizzazione che può portare, paradossalmente, ad un aumento di altre aree. Ben 19 aree sono interessate a questa riorganizzazione. Utilizzando il grado k come *termometro* della neurodegenerazione, si evidenzia che, rispetto alle tradizionali deformazioni del mesencefalo e dei peduncoli cerebellari medio e superiore, ben 29 aree sono coinvolte in una differenza nel grado che va da 0,05 a 1. Questo significa che la maggior parte del cervello perde cellule in diverse zone cerebrali, intaccando la funzionalità di numerosi circuiti cerebrali, con pesanti ripercussioni sul funzionamento del sistema cognitivo nel suo insieme. Una più considerevole diminuzione nel grado di altre aree cerebrali (in media da 1 a 2,5) è rilevabile in soggetti con Parkinson, rispetto a soggetti sani: molte di queste aree sono solitamente poco considerate nella valutazione di questa malattia. Queste involuzioni indicano un progressivo deterioramento della architettura del cervello. Fin qui i risultati dal punto di vista strutturale dimostrano che la malattia modifica profondamente l'architettura cerebrale, andando incontro ad una riorganizzazione con perdite in alcune aree cerebrali, aumento dei cammini medi, del diametro, ma anche parametri di clusterizzazione più rilevanti e parametri di invarianza di scala intercettati, anche nel caso delle reti cerebrali da una legge di potenza troncata.

Dal punto di vista funzionale, caratterizzato attraverso le sottoreti dei motivi, sono state individuate dinamiche circuitali che mostrano una sintassi strutturale-funzionale che opera attraverso le geometrie di cammini neuronali nella rete cerebrale.

Per quanto riguarda la modellazione e la trasformazione in reti di campioni di lingua Inglese, questa ha seguito un processo più diversificato. In primo luogo è stato reperito il database WordNet. In seguito sono stati elaborati diversi algoritmi di interrogazione, basati sui meccanismi di organizzazione lessicale del sistema. Le diverse interrogazioni hanno permesso una lettura, seppure non esaustiva, dei fenomeni

emergenti nella lingua: strutture di grande complessità che creano pattern che contengono geometrie tradizionali (come i solidi platonici), ma anche geometrie a molte dimensioni, i politopi, attraverso strutture note come *simplessi*. Sono state rilevate strutture a 3, 4, 5, 6, 7, ... n – *simplessi*, che dimostrano l'enorme abilità della lingua (nelle organizzazioni cognitive dei soggetti umani) nel creare organizzazioni. Il software Mathematica è stato utilizzato per visualizzare solo in 3 dimensioni le strutture, rilevate per i campi semantici delle parole emozionali, scelti come base per l'analisi delle reti linguistiche. E' stato verificato che anche in altri domini si evidenziano gli stessi meccanismi di organizzazione. Partendo dal termine "emotion", che dentro WordNet è un nodo terminale, sono state indagate le trasformazioni qualitative dei campi semantici delle principali emozioni, considerando solo i raggruppamenti che si attivano con l'algoritmo di ricerca R_8 (*Overview*, che comprende una restituzione di dati lessicali appartenenti a tutte le categorie grammaticali, nomi, verbi, aggettivi, avverbi). Numerose strutture complesse sono emerse, evidenziando pattern di grande complessità e bellezza, che organizzano le relazioni di significato nel lessico e creano percorsi di cambiamento qualitativo nelle reti. Il senso dei cambiamenti, delle corrispondenze, delle opposizioni, degli intrecci tra gli ambiti e i campi semantici sono stati caratterizzati, analizzando le strutture e le corrispondenze con i pattern di organizzazione rilevati. Anche le lingue naturali si organizzano con una grammatica geometrica di pattern, a molti livelli. Le geometrie rivelano le multidimensionalità del pensiero creativo umano.

I dati rilevati dimostrano che, in tutti e tre gli ambiti considerati, i linguaggi manifestano, nonostante le differenze superficiali attraverso le quali sono veicolati, strutture grammaticali e semantiche, *organizzativo-temporali* nel caso della musica, *dinamico-strutturale-funzionali*, nel caso delle reti cerebrali, *semantico-lessicali* nel caso della lingua naturale molto simili. E' una grammatica geometrica, piuttosto che sintattica, una grammatica di pattern, che cresce a salti, aggiungendo ora strutture semplici, ora strutture estremamente complesse, o gruppi di strutture. Queste dinamiche di crescita manifestano sviluppi imprevedibili, fenomeni che rendono ancora non disponibile una piena formalizzazione dei processi emergenti. Una manifestazione abbastanza complessa che è stato rilevata è la presenza di piccole differenze statistiche dei dati reali che sono stati analizzati con i modelli ideali di reti, sia per quanto riguarda le reti di piccolo mondo che quelle a invarianza di scala. Per tutti e tre gli ambiti non perfette corrispondenze con i modelli di reti di piccolo mondo (ovvero non tutti le reti, per i tre ambiti analizzati, manifestano strutture di piccolo mondo) sono state messe in evidenza. Ciò prova che non tutti i dati seguono l'invarianza di scala attraverso una legge di potenza. Si può dire che la legge di potenza si manifesta solo per un certo numero di reti dei tre campioni analizzati. E' una legge di potenza troncata a intercettare il comportamento delle altre reti. Si ritiene che tali variazioni dipendano dalla fisicità dei dati trattati che, nella realtà, si discostano dai modelli ideali. Si ritiene, inoltre, che piccoli cambiamenti nei dati fisici possano portare a grandi cambiamenti nelle dinamiche dei fenomeni. Un altro elemento che sembra influenzare la strutturazione delle reti è la natura diversa degli elementi, rispetto alla omogeneità degli

elementi delle reti nei modelli di Watts e Strogatz e di Albert e Barabási. Si nota, infine che, nonostante i contesti siano diversi, i meccanismi di organizzazione sembrano apparentemente dello stesso tipo. Per cui, concludendo si evidenzia una grande congruenza dei risultati sperimentali con l'ipotesi che era stata posta alla base di questo lavoro.

Bibliografia

Adami, C. (1998). *Introduction to Artificial Life*. New York: Springer Verlag.

Albert, R., Jeong, H., & Barabási, A. L. (1999). Diameter of the World Wide Web. *Nature*, 401, 130-131.

Albert, R. & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47-97.

Aitken, K. J. & Trevarthen, C. (2001). Infant Intersubjectivity: Research, Theory, and Clinical Applications. *Journal of Child Psychology and Psychiatry*, 42, 1, 3–48.

Amaral., L, Scala, A., Barthelemy, M., Stanley, H. (2000). Classes of small-world networks. *Proceedings of the National Academy of the United States of America*, 97, 11149–11152.

Anolli, L. (2002). *Le emozioni*. Milano: Edizioni Unicopli.

Appert-Rolland, P. Degond, S. Motsch, (2011). Two-way multi-lane traffic model for pedestrians in corridors, *Networks and Heterogeneous Media*. 6, 3, 351-381.

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neuro-linguistics. *Behavioral and Brain Sciences*, 28, 105-67.

Argyle, M. (1988). *Bodily Communication*. (2nd Ed.), New York: Methuen.

Arnold, M. B. (1960a). *Emotion and personality. Volume I: Psychological aspects*.

New York: Columbia University Press.

Arnold, M. B. (1960b). *Emotion and personality. Volume II: Neurological and physiological aspects*. New York: Columbia University Press.

- Arnold, J. B. (1971). A multidimensional scaling study of semantic distance. *Journal of Experimental Psychology*, 90,2, 349-372.
- Ashton, A., (2003). *Harmonograph: A visual guide to the mathematics of music*. New York: Walker & Company.
- Ball, B., Karrer, B., Newman, M. E. J. (2011). An efficient and principled method for detecting communities in networks. *Phys. Rev. E*, 84, 036103.
- Barabási, A. -L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286 (5439), 509-512.
- Barabási, A. -L., Albert, R., Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A*, 272, 173-187.
- Barabási, A. -L., Albert, R., Jeong, H. & Bianconi, G. (2000). Response to power-law distribution of the world wide web. *Science*, 287, 2115a.
- Barabási, A. -L. , Albert, R. & Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world wide web, *Physica A*, 281, 69-7.
- Barabási, A. -L. & Bianconi, G. (2001). Bose-Einstein Condensation in Complex Networks. *Physical Review Letters*, 86, 24, 5632-5635.
- Barabási, A. -L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A. & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A*, 311, 590--614.
- Barabási, -L. A. (2004). Link. *La scienza delle reti*. Milano: Einaudi.
- Barabási, A. -L.(2005a).The origins and heavy tails in human dynamics. *Nature*, 435, 207--211.
- Barabási, A. -L.(2005b). Network theory---the emergence of the creative enterprise, *Science*, 308, 639--641.
- Baroni, M. & Jacobini, C. (1978). *Proposal for a Grammar of Melody*. Montreal: Les Presses de l'Université de Montreal.
- Barrat, A., Barthelemy, M., Pastor-Satorras, R. & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of the United States of America*, 101, 3747--3752.
- Barsalou, L.W. (1999). Perceptual symbols systems. *Behavioral & Brain Sciences*, 22, 577-660.
- Barsalou, L.W. (2008). Grounded Cognition. *Annual Review of Psychology*, 59, 617-645.

- Basile G. (2001). *Le parole nella mente. Relazioni semantiche e struttura del lessico*. Milano: Franco Angeli.
- Basili, R., Serafini, A. & Stellato, A. (2004). Classification of musical genre: A machine learning approach. *Proceedings of the International Conference on Music Information Retrieval*, 505–8.
- Basser, P.J., Mattiello, J. & LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys Journal*, 66, 259–67.
- Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A. (2000). In vivo fiber tractography using DT-MRata. *Magn Reson Med*, 44, 4, 625–32.
- Bassett, D. S, Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., & MeyerLindenberg, A. (2008). Hierarchical organization of human cortical networks in health and schizophrenia, *Journal of Neuroscience*, 28, 9239–48.
- Beck, B. B. (1980). *Animal Tool Behaviour*, New York: Garland Publishing.
- Bedau, M. A. (2003). Artificial life: organization, adaptation and complexity from the bottom up. *Trends in Cognitive Sciences*, 7, 505-512.
- Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F. & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage*, 34,144-55.
- Bengio, Y. (1996). *Neural networks for speech and sequence recognition*. London: International Thomson Computer Press.
- Bertacchini, F., Bilotta, E., Gabriele, L., Mazzeo, V., Pantano, P., Rizzuti, C. & Vena, S. (2007). ImaginationTOOLS(TM): Made to play music. *Edutainment 2007, LNCS*, 369-380. New York: SpringerVerlag.
- Bertacchini, F., Bilotta, E. & Pantano, P., (2009). *Il caos è semplice e tutti possono capirlo*. Firenze: Muzzio.
- Bertacchini, F., Bilotta, E., Pantano, P. & Tavernise, A. (2012a). Motivating the learning of science topics in secondary school: A constructivist edutainment setting for studying Chaos. *Computers & Education*, vol. 59, p. 1377-1386, ISSN: 0360-1315, doi: 10.1016/j.compedu.2012.05.001.
- Bertacchini, F., Bilotta, E., Gabriele, L., Pantano, P. & Tavernise, A. (2012b). Art inside Chaos: New Routes to Creativity and Learning. *Leonardo*, ISSN: 1530-9282.
- Bertacchini, F., Bilotta, E., Laria, G. & Pantano, P., (2012c) The Genesis of Chua’s Circuit. Connecting Science, Art and Creativity. In Adamatzky, A. & Chen, G. *Chaos, CNN, Memristors and Beyond. A Festschrift for Leon Chua*. Singapore: World Scientific.

Bertacchini, F., Bilotta, E., Gabriele, L., Pantano, P. & Tavernise, A. (2013). Towards the use of Chua's circuit in education, art, and interdisciplinary research: some implementation and opportunities. *Leonardo, NuovaSerie*, Accepted for publication in 2013.

Bertacchini, F., Bilotta, E., Pantano, P. & Rosa, F. (sottomesso, a). *The networks of music language*.

Bertacchini, F., Bilotta, E., Pantano, P. (sottomesso, b). *Disentangling Progressive Supranuclear Palsy (PSP) using Connectomics and Networks Theory*.

Bertacchini, F., Bilotta, E., Pantano, P. (sottomesso, c). *The networks of language. An investigation of semantic cognition*.

Bharucha, J. J., & Todd, P. M. (1989). Modeling the perception of tonal structure with neural nets. *Computer Music Journal*, 13,4, 44–53.

Bilotta, E. (1999). *Dalla Scienza Cognitiva agli Agenti Intelligenti*. Cosenza: Bios, Edizioni Memoria.

Bilotta, E., Pantano, P. & Talarico V. (2000a). Music Generation through Cellular Automata: How to Give Life to Strange Creatures. *Generative Art GA2000*, Politecnico di Milano Università di Milano, Italia.

Bilotta, E., Pantano P. & Talarico V. (2000b). Synthetic harmonies: an approach to musical semiosis by means of cellular automata. In Bedau, M. A., McCaskill, J. S., Packard, N. H. & Rasmussen, S. (Eds.), *Artificial Life VII* , 537-546. Cambridge, MA: The MIT Press.

Bilotta, E. & Pantano., P. (2001). Artificial Life Music Tells Complexity. *Proc. of ALMMA I-Artificial Life Models for Musical Applications I, 6th European Conference on Artificial Life*. Prague, 17-28.

Bilotta, E., Miranda, E. R., Pantano, P. & Todd, P. M. (2002a). Artificial Life Models for Musical Applications: Workshop Report, *Artificial Life*, 8.1, 83-86.

Bilotta, E., Lafusa A. & Pantano, P. (2002b), Is self-replication an embedded characteristic of artificial/living matter? *Artificial Life VIII*, Cambridge, MA: The MIT Press.

Bilotta, E. & Pantano, P. (2002a). Self-reproducers use contrapuntal means, *Proc. of ALMMA II-Artificial Life Models for Musical Applications II : Searching for musical creativity, Workshop of the 8th International Conference on the Simulation and Synthesis of Living Systems (ALife VIII)*, University of New South Wales, Sydney, Australia., 3-8. Cambridge, MA: The MIT Press.

Bilotta, E. & Pantano, P. (2002b). An advanced tool for authoring multimedia presentation, in *Progress in Industrial Mathematics at ECMI 2000*, Anile, A. M., Capasso, V. & Greco, A. (Eds.), *Mathematics in Industry*. 1, 126-139, New York: Springer-Verlag,

- Bilotta, E. & Pantano, P. (2002c). Synthetic Harmonies: An Approach to Musical Semiosis by Means of Cellular Automata. *Leonardo*, 35, 35-42, Cambridge, MA: The MIT Press.
- Bilotta, E., Gervasi, S. & Pantano, P. (2005). Reading Complexity in Chua's Circuit by Music-Part I: a new way of understanding chaos. *Int. J. of Bifurcation and Chaos*, 15-2, 253-382.
- Bilotta, E. & Pantano, P. (2008). *A Gallery of Chua Attractors*. Singapore: World Scientific Publishing Co Pte Ltd.
- Bilotta, E. & Pantano, P. (2010). *Cellular Automata and Complex Systems: Methods for Modeling Biological Phenomena*, N.Y.: Hershey, IGI Global.
- Bilotta, E., Pantano, P. & Vena, S. (2011). Artificial Micro-Worlds. Part I: A new approach for studying life-like phenomena, *Int. J. of Bifurcation and Chaos*, 21-1, 373-398
- Bilotta, E. & Pantano, P. (2011a). Artificial Micro-Worlds. Part II: Cellular Automata growth dynamics, *Int. J. of Bifurcation and Chaos*, 21-2, 619-645.
- Bilotta, E. & Pantano, P. (2011b). Artificial Micro-Worlds. Part III. A taxonomy of self-reproducing 2D CA species, *Int. J. of Bifurcation and Chaos*, 21-3, 1233-1263.
- Bilotta, E. & Pantano, P. (2011c). Artificial Micro-Worlds. Part IV. Models of complex self-reproducers, *Int. J. of Bifurcation and Chaos*, 21-4, 1501-1521.
- Booij, G. E. (2007). *The Grammar of Words: An Introduction to Linguistic Morphology*. (2nd Ed.), Oxford: Oxford University Press.
- Brooks, R. A. (1990). Elephants Don't Play Chess. *Robotics and Autonomous Systems*, 6, 3-15.
- Bruner, J. S. (1976). Dalla comunicazione al linguaggio: Una prospettiva psicologica. In *Cognition*, vol. 3, 255-287. In Camaioni, L. (a cura di) *Sviluppo del linguaggio e interazione sociale*. Il Mulino: Bologna.
- Bruner, J. S. (1983). *Child's talk: Learning to use language*. New York: Norton.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V. & Rizzolatti G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Brain Res Cogn. Brain Res*, 24, 355-363.
- Bullmore, E. & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Rev Neuroscis*, 10, 186-198.

- Bullmore, E.T. & Bassett, D.S. (2011). Brain graphs: graphical models of the human brain connectome. *Annu Rev Clin Psychol*, 7, 113–140.
- Caldarelli, G., (2007). *Scale-Free Networks*. Oxford: Oxford University Press.
- Callaghan, P.T. (1991). *Principles of nuclear magnetic resonance microscopy*. Oxford: Clarendon Press.
- Camaioni, L. & Di Blasio, P. (2007). *Psicologia dello sviluppo*. Bologna: Il Mulino.
- Campolongo, G. & Vena, S. (2005). Analysing and Creating Music Through Small-World Networks. In *Proceedings of V International Conference Understanding and Creating Music*, Seconda Università di Napoli, Caserta, 52-55.
- Campolongo, G. & Vena, S., (2006). The Science of networks and music: A new approach on musical analysis and creation. *LNCS 3907*, 642-651.
- Canestrari, R. (1984). *Psicologia generale e dello sviluppo*. Bologna: Clueb.
- Cannon, W. B. (1927). The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *American Journal of Psychology*, 39, 106-126.
- Capute, A.J., Palmer, F.B., Shapiro, B.K., Wachtel, R.C., Schmidt, S. & Ross, A. (1986). Clinical linguistic and auditory milestone scale: prediction of cognition in infancy. *Developmental Medicine and Child Neurology*, 28, 762–771.
- Carnap R. (1952). Empiricism, Semantics and Ontology. In Linsky, L., (Ed.), *Semantics and the Philosophy of Language*. Urbana: University of Illinois Press, 208–228.
- Carr, H. Y. & Purcell, E. M. (1954). Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* 94, 630–638.
- Casti, J. (1994). *Complexification*. New York: HarperCollins.
- Casti, J. (1996). *Would-Be Worlds*. New York: Wiley.
- Cavanagh, L. (2009). A Brief History of the Establishment of International Standard Pitch A D 440 Hz, *WAM, Webzine about Audio and Music*, 4.
- Chaffin, R. (1992). The Concept of a Semantic Relation. In Lehrer, A. & Kittay, E. F. (Eds.), *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*. Hillsdale: NJ, 253–288.
- Chiarelli, B. (1985). *L'evoluzione del cervello umano, in Homo. Viaggio alle origini della storia*. Venezia: Marsilio.

- Chomsky, N. (1957). *Syntactic Structures*. The Hague, Mouton.
- Chomsky, N. (1959). A Review of Skinner's Verbal Behavior. *Language*, 35, 26-58.
- Chomsky, N. (1965). *Aspects of the Theory of the Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1968). *Language and Mind*. Hartcourt, N.Y.: Brace Jovanovich, Inc.,
- Clark J. & Holton D.A., (1995). *A First look at Graph Theory*. Singapore: World Scientific.
- Clauset, A. & Shalizi, C. R., Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM REVIEW*, 51,4, 661–703
- Collins, A.M. & Quillian, M.R., (1969). Tempo di recupero dalla memoria semantica. *Journal of Verbal Learning e Verbal Behavior*, 8, 240-247.
- Collins, A. M. & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In Gregg, L. W. (Ed.) *Cognition in learning and memory*. New York: Wiley.
- Collins, A.M. & Loftus E., (1975). Spread Activation Theory. *Psychological Review*, 82, 407-428.
- Corballis, M. (2009). Language as gesture, Human Movement. *Science*, 28, 556-65.
- Colton, S. & Wiggins G. A. (2012). Computational Creativity: The Final Frontier? *ECAI 2012*, 21-26.
- Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., et al. (1999). Tracking neuronal fiber pathways in the living human brain. *PNAS*, 96, 10422–7.
- Cosottini, M., Ceravolo, R., Faggioni, L., et al. (2007). Assessment of midbrain atrophy in patients with progressive supranuclear palsy with routine magnetic resonance imaging. *Acta Neurol Scand*, 116, 37-42.
- Coxeter, H. S. M. (1974). *Regular Complex Polytopes*. Cambridge: Cambridge University Press.
- Cupellini E., Rizzuti C., Bilotta E., Pantano P., Wozniowski M. & Cooperstock J. R. (2008). Exploring Musical Mappings and Generating Accompaniment with Chaotic Systems, *ICMC 2008*, 24-29 August , Belfast, ISBN 0 – 9713192-6x, 467-474.
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. (1st ed.), London: John Murray, retrieved 2012-10-23 .
- Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: John Murray.
- Darwin, C. (1872). *The expression of the emotions in man and animals*. London: John Murray.

- Del Bo, M., Cippone, P. & De Filippis, A. (1988). *La sordità infantile grave. Nuove prospettive mediche e nuovi metodi di rieducazione*. Roma: Armando.
- de Haan, W., Pijnenburg, Y.A.L., Strijers, R.L.M., van der Made, Y., van der Flier, W.M., et al. (2009). Functional neural network analysis in frontotemporal dementia and Alzheimer's disease using EEG and graph theory. *BMC Neurosci*, 10, 101.
- De Mauro, T. (2008). *Lezioni di linguistica teorica*. Bari: Laterza.
- de Nooy, W., Mrvar, A. & Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. Cambridge: Cambridge University Press.
- De Santis, D. (2010). Lo sviluppo del linguaggio nel bambino sordo e udente: due modalità comunicative a confronto. *Studi di Glottodidattica*, 1, 75-91.
- De Saussure, F. (1916). *Cours de linguistique générale*, (a cura di) Bally, C., Sechehaye, A. & Reidlinger, A., Losanna-Parigi, Payot.
- Douek, P., Turner, R., Pekar, J., Patronas, N. & Le Bihan, D. (1991). MR color mapping of myelin fiber orientation. *J Comput Assist Tomogr*, 15, 923-9.
- Ekman, P., Friesen, W.V. & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings*. New York :Pergamon.
- Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York: Times Books.
- Emde, R. N., Biringen, Z., Clyman, R. B. & Oppenheim, D. (1991). The moral self in infancy: Affective core and procedural knowledge. *Development Review*, 11, 251-270.
- Erdős P. & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, 290-297.
- Erdős, P. & Rényi, A. (1960). The Evolution of Random Graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 5, 17-61.
- Eysenck, H. J. (1964). Pavlov's Writings. *BMJ*, 2 (5401): 111.
- Eysenck, M. W. & Keane, M. T. (2006). *Manuale di psicologia cognitiva*. Napoli: Idelson-Gnocchi.
- Fenson, L., Dale, P., Reznick, J. S., Bates, E., Thal, D. & Pethick, S. (1994). Variability in early communicative development. Monographs for the Society for Research. *Child Development*, 59, 242.

- Ferrer i Cancho, R., Capocci, A. & Caldarelli, G. (2007a). Spectral methods cluster words of the same class in a syntactic dependency network. *International Journal of Bifurcation and Chaos*, 17, 7, 2453-2463.
- Ferrer i Cancho, R., Mehler, A., Pustynnikov, O. & Diaz-Guilera, A., (2007b). Correlations in the organization of large-scale syntactic dependency networks, in *TextGraphs-2: Graph-Based Algorithms for Natural Language Processing*, 65–72. Rochester: Association for Computational Linguistics.
- Ferrer i Cancho R. & Solé R. (2002). Least effort and the origins of scaling in human language. In Wachter, K. W. (Ed.), University of California. *PNAS*, 100, 3, 788-791.
- Ferrer i Cancho R. & Solé R. (2001). The small-world of human language. *The Royal Society B*, 268, 2261-2265.
- Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.
- Frova, A. (1999). *Fisica nella musica*. Bologna: Zanichelli.
- Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8, 396-403.
- Garcia-Lazaro, J.A., Ahemd, B. & Schnupp, J.W. (2006). Dynamical substructure of coordinated rhythmic movements. *Curr Biol* 16, 264–271.
- Guasti, M.T. (2007). *L'acquisizione del linguaggio. Un'introduzione*, Milano: Raffaello Cortina Editore.
- Gazzaniga, M.S., (2004). *The Cognitive Neurosciences*. 3rd Ed., Cambridge: The MIT Press.
- Gentilucci, M. & Corballis, M.C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience and Biobehavioral Reviews*, 30, 949-60.
- Geswind, N. (1972). Language and the Brain. *Scientific American*, 266, 76-83.
- Golinkoff, R. M., Mervis, C.B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, 21, 125-155.
- Gong, G., He, Y., Concha, L., Lebel, C., Gross, D. W., Evans, A. C. & Beaulieu, C. (2009). Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb Cortex*, 19, 3, 524–36.
- Granovetter, M. S. (1973). The strength of weak ties: A network theory revisited. *American Journal of Sociology*, 1360-1380.

- Greely, H. T., & Illes, J. (2007). Neuroscience-based lie detection: The urgent need for regulation. *American Journal of Law and Medicine*, 33, 377-431.
- Greimas, A. J. (1966). *Sémantique structurale, recherche et method*. Paris: Larousse.
- Hagmann, P., Cammoun, L., Gigandet, X., Gerhard, S., Ellen Grant, P., Wedeen, V., Meuli, R., Thiran, J.-P., Honey, C. J. & Sporns, O. (2010). MR connectomics: principles and challenges. *J. Neurosci. Methods*, 194, 34–45.
- Hagmann, P, Kurant, M, Gigandet, X, Thiran, P, Wedeen, V.J., Meuli, R., et al. (2007). Mapping human whole-brain structural networks with diffusion MRI. *PLoS ONE*, 2, e597.
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J. & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6, e159.
- Hebb, D. O. (1946). On the Nature of Fear. *Psychological Review*. 53, 259–276.
- Heidar, D. M. (2012). First Language Acquisition: Psychological Considerations and Epistemology. *Theory and Practice in Language Studies*, 2, 411-416.
- Hjelmslev, L. (1961). *Prolegomena to a Theory of Language*. Madison: The University of Wisconsin.
- Hiller, L., & Fuller, R. (1967). Structure and information in Webern's Symphonie, Op. 21. *Journal of Music Theory*, 11,1, 60–115.
- Hockett, C. F. (1960). The Origins of Speech. *Scientific American*, 203, 88-96.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26, 55–88.
- Hofstadter, D. R. (1999). *Gödel, Escher, Bach: An Eternal Golden Braid*. NY: Basic Books.
- Hopcroft, J.E. Motwani, R.& & Ullman, J.D. (2007). *Introduction to Automata Theory, Languages, and Computation*. (3rd Ed.), Boston, MA: Addison Wesley.
- Hsu, A. & Griffiths, T. (2009). Differential Use of Implicit Negative Evidence in Generative and Discriminative Language Learning. *Advances in Neural Information Processing Systems*, 22, 77-82.
- Hsu, A. S, Chater, N. & Vitányi, P. M. B. (2011). The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis. *Cognition*, 120, 380-390.
- Hughes, A.J., Daniel, S.E. & Lees, A.J. (2001). Improved accuracy of clinical diagnosis of Lewy body Parkinson's disease. *Neurology*, 57, 1497–1499.

- Iturria-Medina Y, Sotero, R.C., Canales-Rodriguez, E. J., AlemanGomez. Y. & Melie-Garcia, L. (2008). Studying the human brain anatomical network via diffusion-weighted MRI and graph theory. *Neuroimage*, 40, 1064-1076.
- Itzkovitz, S. & Alon, U. (2005). Subgraphs and network motifs in geometric networks. *Phys. Rev. E* 71, 26-117.
- Izard, C.E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.
- Izard, C.E. (1977). *Human emotions*. New York: Plenum.
- James, W. (1884). What is an Emotion? *Mind*, 9, 188-205.
- Karmiloff-Smith, A. (1995). *Oltre la mente modulare. Una prospettiva evolutiva sulla scienza cognitiva*. Bologna: Il Mulino.
- Ke, J. (2007). Complex networks and human language, unpublished paper retrieved at: <http://www.isrl.uiuc.edu/~amag/langev/paper/ke07complexNetworkLanguage.html>, 11/10/2012.
- Knopoff, L., & Hutchinson, W. (1981). Information theory for musical continua. *Journal of Music Theory*, 25, 1, 17-44.
- Krumhansl, C. L. (1997). An Exploratory Study of Musical Emotions and Psychophysiology. *Canadian Journal of Experimental Psychology*, 51,4, 336-352.
- Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115-129.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage Publications.
- Kunej, D. & Turk, I. (2000). New perspective on the beginnings of music: archaeological and musicological analysis of a middle Palaeolithic bone "flute". In Wallin, N. L., Merker, B. & Brown S. (Eds.), *The Origins of Music*. Cambridge, MA: The MIT Press, 235-268.
- Kubicki, M., Westin, C.F., Nestor, P.G., Wible. C.G., Frumin, M., Maier, S.E., et al. (2003). Cingulate fasciculus integrity disruption in schizophrenia: a magnetic resonance diffusion tensor imaging study. *Biological Psychiatry*, 54, 11, 1171-80.
- Kuhl, P. K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, 67, 713-727.
- Langton, C.G. (1986). Studying Artificial Life with Cellular Automata . *Physica D*, 22, 120-149.
- Langton, C.G. (1990). Computation at the edge of chaos. *Physica D*, 42, 12-37.

- Lazar, M., Weinstein, D.M., Tsuruda, J.S., Hasan, K.M., Arfanakis, K., Meyerand, E.M., Badie, B., Rowley, H., Haughton, V., Field, A., Witwer, B. & Alexander, A.L., (2003). White matter tractography using tensor deflection. *Hum. Brain Mapp.* 18, 306–321.
- LeDoux, J. (2002). *Synaptic self. How our brains become who we are.* NY: Penguin Putnam.
- Lennenberg, E.H. (1982). *Fondamenti biologici del linguaggio.* Torino: Boringhieri.
- Lerdahl, F., & Jackendoff, R. (1983a). *A Generative Theory of tonal music.* Cambridge, MA: the MIT Press.
- Lerdahl, F., & Jackendoff, R. (1983b). Generative Music Theory and Its Relation to Psychology. In *Journal of Music Theory* (25th Anniversary Issue) 25, 1, 45-90.
- Levitin, D.J. & Menon, V. (2003) Musical structure is processed in ‘language’ areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *Neuroimage*, 20, 2142–2152.
- Levitin, D.J. (2006). *This Is Your Brain on Music.* New York: Dutton.
- Levitin, D.J. (2010). Why music moves us. *Nature*, 464, 834–835.
- Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, 120, 2546–2556.
- Litvan, I., Agid, Y., Calne, D., Campbell, G., Dubois, B., Duvoisin, R.C., et al. (1996). Clinical research criteria for the diagnosis of progressive Supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDSSPSP international workshop. *Neurology*, 47, 1–9.
- Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., et al. (2008). Disrupted small-world networks in schizophrenia. *Brain*, 131, 945–961.
- Liu, X. F., Tse, C. K. & Small, M. (2010). Complex network structure of musical compositions: Algorithmic generation of appealing music. *Physica A* 389, 126-132.
- Lockhart, R.S., (2000). Methods of memory research. In Tulving, E. & Craik, F.I.M. (Eds.), *The Oxford Handbook of Memory.* Oxford: Oxford University Press, 45–57.
- Loftus, E.F. (1974). Reconstructing memory: The incredible eyewitness. *Psychology Today*, 8, 116-119.
- Loftus, E.F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7, 560-572.
- Lyons, J. (1982). *Lezioni di linguistica.* Bari: Laterza.
- Machado, P., Romero, J. & Manaris, B. (2007). *Experiments in Computational Aesthetics – An Iterative Approach to Stylistic Change in Evolutionary Art, The Art of Artificial Evolution.* New York: Springer-Verlag.

- Machlis, J. & Forney, K. (1990). *The Enjoyment of Music*. New York: W. W. Norton & Co Inc.
- Mac Lean, P. D. (1984). *Evoluzione del cervello e comportamento umano*. Torino: Einaudi.
- Maess, B., Koelsch, S., Gunter, T.C. & Friederici, A.D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neurosci.*, 4, 540–545.
- Magno Caldognetto, E., Cavicchio, F., & Cosi, P. (a cura di) (2004). Comunicazione parlata e manifestazione delle emozioni. *Atti del I Convegno GSCP*, Padova 29 novembre - 1 dicembre 2004, Napoli: Liguori Editore.
- Malloch, S. & Trevarthen, C. (2010). *Communicative Musicality: Exploring the basis of human companionship*. New York: Wiley.
- Manaris, B., Vaughan, D., Wagner, C., Romero, J. & Davis, R. B. (2003). Evolutionary Music and the Zipf–Mandelbrot Law – Progress towards Developing Fitness Functions for Pleasant Music, Applications of Evolutionary Computing. *LNCS 2611*, New York: Springer-Verlag, 522-534.
- Manaris, B., Romero, J., Machado, P., Krehbiel, D., Hirzel, T., Pharr, W., & Davis, R.B. (2005). Zipf's Law, Music Classification and Aesthetics. *Computer Music Journal*, 29,1, 55-69.
- Manaris, B., Roos, P., Machado, P., Krehbiel, D., Pellicoro, L., & Romero, J. (2007). A Corpus-Based Hybrid Approach to Music Analysis and Composition Proceedings of Twenty-Second *Conference on Artificial Intelligence (AAAI-07)*, Vancouver, BC, July 2007.
- Mandler, G. (1975). *Mind and Emotion*. New York: Wiley.
- Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. New York: Norton.
- Margulis, E. H., & Beatty, A. P. (2008). Musical style, psychoaesthetics, and prospects for entropy as an analytic tool. *Computer Music Journal*, 32,4, 64–78.
- Meyer, L., (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.
- Milgram, S. (1967). The Small World Problem. *Psychology Today*, 1,1, 60 – 67.
- Miller, G. A. (1967). Psycholinguistic approaches to the study of communication. In Arm, D. L. (Ed.) *Journeys in science: small steps-great strides*. Albuquerque: University of New, Mexico Press.
- Miller, G. A. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 6, 169-191.
- Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38, 39-41.

- Miller, G. A. (1972). English verbs of motion: a case study in semantics and lexical memory. In Melton, A. W. & Martin, E. (Eds.) *Coding processes in human memory*, Washington, D.C.: V. H. Winston and Sons.
- Miller, G. A. (1979). *A Very Personal History*. Cambridge, MA: Center for Cognitive Science.
- Miller, G., Beckwith, R., Fellbaum, C., Gross, D. & Miller, K. (1990). WordNet: An on-line lexical database. *International journal of lexicography*, 3,4, 235-244.
- Miller, G. A. (1998). Nouns in WordNet. WordNet - an electronic lexical database. In Fellbaum, C. (Ed.), *WordNet: An electronic lexical database*. Cambridge, MA: The MIT Press.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298, 824–827.
- Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R. & Alon, U. (2004a). Response to comment on Network motifs: Simple building blocks of complex networks and Superfamilies of evolved and designed networks. *Science*, 305, 1107.
- Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., et al. (2004b). Superfamilies of evolved and designed networks. *Science* 303, 1538–1542.
- Minsky, M.(1975). A Framework for Representing Knowledge. In Winston, P., (Ed.) *The Psychology of Computer Vision*. New York: McGraw-Hill, 211-77.
- Mithen, S. (2006). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. Cambridge, MA: Harvard Univ. Press.
- Molnar-Szakacs, I.,Kaplan, J., Greenfield, P. M. & Iacoboni, M.(2006). Observing complex action sequences: The role of the front-parietal mirror neuron system. *Neuro Image*, 33, 923-35.
- Mori, S., Crain, B.J., Chacko, V.P. & van Zijl, P.C. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals Neurol*, 45, 265–9.
- Morin E. (1983). *Il metodo. Ordine, disordine, organizzazione*. Milano: Feltrinelli.
- Moseley, M.E., Kucharczyk, J., Mintorovitch, J., Cohen, Y., Kurhanewicz, J., Derugin, N., et al. (1990). Diffusion-weighted MR imaging of acute stroke: correlation with T2-weighted and magnetic susceptibility-enhanced MR imaging in cats. *Ajnr*, 11,423–9.

- Moseley, M. E., Y. Cohen, Y. Kucharczyk, J., Mintorovitch, J. Asgari, H. S. Wendland, M. R. Tsuruda, J. & Norman. D. (1990). Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. *Radiology*, 176, 439-446.
- Mukherjee, A., Choudhury, M. & Kannan, R. (2009). Discovering Global Patterns in Linguistic Networks through Spectral Analysis: A Case Study of the Consonant Inventories, *EACL 2009*, 585-593.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Nespor, M. (1993). *Le strutture del linguaggio. Fonologia*. Bologna: Il Mulino.
- Nespor, M. & Bafile, L. (2008). *I suoni del linguaggio*, Bologna: Il Mulino.
- Newman, M. E. J., Strogatz, S. H. & Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64, 26-118.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45, 167-256.
- Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford: Oxford University Press.
- Nicoletti, G., Fera, F., Condino, F., et al. (2006a). MR imaging of middle cerebellar peduncle width: differentiation of multiple system atrophy from Parkinson disease. *Radiology*, 239, 825-30.
- Nicoletti, G., Lodi, R., Condino, F., et al. (2006b). Apparent diffusion coefficient measurements of the middle cerebellar peduncle differentiate the Parkinson variant of MSA from Parkinson's disease and progressive supranuclear palsy, *Brain*, 129, 2679-87.
- Nicoletti, G., Tonon, C., Lodi, R., et al. (2008). Apparent diffusion coefficient of the superior cerebellar peduncle differentiates progressive supranuclear palsy from Parkinson's disease. *Mov Disord*, 23, 2370-6.
- Nilsson, C., Markenroth, B.K., Brockstedt, S., Latt, J., Widner, H. & Larsson, E.M. (2007). Tracking the neurodegeneration of parkinsonian disorders—a pilot study. *Neuroradiology*, 49, 111–9.
- Nolfi S. & Mirolli M. (2010). *Evolution of Communication and Language in Embodied Agents*, Berlin: Springer Verlag.
- Nowak, M., Komarova, N. & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, 417, 611–617.
- Oba, H., Yagishita, A., Terada, H., et al. (2005). New and reliable MRI diagnosis for progressive Supranuclear palsy. *Neurology*, 64, 2050-5.

- Oliverio, A. (2004). Neuroscienze. Basi biologiche dei processi mentali, in *Enciclopedia del Novecento*. Terzo supplemento, Istituto della Enciclopedia Italiana, 2° vol., 230-37.
- Ortony, A., Clore, G. L. & Collins, A. (1988). *The cognitive structure of emotions*, Cambridge: Cambridge University Press.
- Panksepp, J., (2003). At the interface of the affective, behavioral, and cognitive neurosciences: decoding the emotional feelings of the brain. *Brain Cognition*. 52,1, 4 – 14.
- Papez, J. W. (1937). A proposed mechanism of emotion. *J Neuropsychiatry Clin Neurosci*, 7,1, 103-12.
- Parsons, L. M. (2003). Exploring the functional neuroanatomy of music performance, perception and comprehension. In Peretz, I. & Zatorre, R. J., (Eds.), *The Cognitive Neuroscience of Music*. Oxford: Oxford University Press, 247–268.
- Patel, A. D., Gibson, E., Ratner, J., Besson. M. & Holcomb, P. J. (1998). Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience*. 10, 717–33.
- Patel, A. D. & Balaban, E. (2000). Temporal patterns of human cortical activity reflect tone sequence structure. *Nature*, 404, 80–84.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Natural Neuroscience*. 6, 674–81.
- Paternoster, A. (2002). *Percezione e Ancoraggio simbolico, Linguaggio e Percezione, le basi sensoriali della comunicazione linguistica*. Roma: Carocci.
- Pearce, M., Conklin, D., Wiggins. G. & Wiil, U. K. (Ed.) (2004). Methods for Combining Statistical Models of Music. *LNCS 3310*, 295–312. Berlin Heidelberg : SpringerVerlag.
- Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, 100, 1-32.
- Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *J. Cognit. Neurosci*. 1, 153–170.
- Piaget, J., (1972). *La nascita dell'intelligenza nel fanciullo*. Firenze: Giunti.
- Pierce, J. R. (1992). *The Science of Musical Sound*. New York: Freeman and Company.
- Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: Harper & Collins.
- Pierpaoli, C., Jezzard, P., Basser, P.J., Barnett, A. & Di Chiro, G. (1996). Diffusion tensor MR imaging of the human brain. *Radiology*, 201, 637-48.

- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Plutchik, R. (1962). *The emotions: Facts, Theories and a new model*. New York: Random House.
- Plutchik, R. (1980). *Emotion: A psychoevolutionary synthesis*. New York: Harper & Row.
- Plutchick, R. (1995). *Psicologia e biologia delle emozioni*. Torino: Bollati Boringhieri.
- Plutchik, R. (1997). *The psychology and biology of emotion*. New York: HarperCollins.
- Plutchik, R. (2001). The nature of emotions. *American Scientist* 89, 344-351
- Preuss, T. M. (2012). Human brain evolution: From gene discovery to phenotype discovery. *PNAS*, 26-109, 10709-10716.
- Quattrone, A., Nicoletti, G., Messina, D., et al. (2008). MR imaging index for differentiation of progressive supranuclear palsy from Parkinson disease and the Parkinson variant of multiple system atrophy. *Radiology*, 246, 214-21.
- Quillian, M.R. (1967). Word concepts: a theory and simulation of some basic semantic capabilities. *Behavioral Science*, 12, 410-430.
- Quillian, M.R. (1968). Semantic memory. In Minsky, M. (Ed.), *Semantic Information Processing*, Cambridge, MA: The MIT Press.
- Quillian, M.R. (1969). The teachable language comprehender: a simulation program and theory of language. *Communications of the ACM*, 12, 8, 459-476.
- Rizzolatti, G., & Arbib, M.A. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188-194.
- Rizzolatti, G., & Sinigaglia, C. (2006). *So quel che fai. Il cervello che agisce e i neuroni specchio*. Milano: Cortina Raffaello.
- Roeper, T. (2007). *The Prism of Grammar: How Child Language Illuminates Humanism*. Cambridge, MA: The MIT Press.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General.*, 104, 192-233.
- Rosch, E. (1978). Principles of categorisation. In Rosch, E. & Lloyd, B.B. (Eds.), *Cognition and categorization*, NJ: Hillsdale.

- Rowe, M. L., Raudenbush, S. W. & Goldin-Meadow, S. (2012). The Pace of Vocabulary Growth Helps Predict Later Vocabulary Skill. *Child Development*, 83, 508–525.
- Rubinov, M., Knock, S.A., Stam, C.J., Micheloyannis, S., Harris, A.W.F., et al. (2009). Small-world properties of nonlinear brain activity in schizophrenia. *Hum Brain Mapp.*, 30, 403–416.
- Rubinov, M. & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52, 1059-69
- Schank, R.C. & Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Earlbaum Assoc.
- Schank, R.C. (1991). *Tell Me a Story: A New Look at Real and Artificial Intelligence*. New York: Simon & Schuster.
- Schachter, S. & Singer, J. E. (1962). Cognitive, Social, and Physiological Determinants of Emotional State. *Psychological Review*, 69,5, 379-399.
- Schacter, D.L., (1994). Priming and multiple memory systems: perceptual mechanisms of implicit memory. In Schacter, D.L. & Tulving, E. (Eds.), *Memory Systems*. Cambridge, MA: The MIT Press.
- Scherer, K. R., Shorr, A. & Johnstone, T. (2001). *Appraisal processes in emotion: theory, methods, research*. Canary, NC: Oxford University Press.
- Schocke, M. F., Seppi, K., Esterhammer, R., et al. (2002). Diffusion-weighted MRI differentiates the Parkinson variant of multiple system atrophy from PD. *Neurology*, 58, 575-80.
- Schocke, M. F., Seppi, K., Esterhammer, R., et al. (2004). Trace of diffusion tensor differentiates the Parkinson variant of multiple system atrophy and Parkinson's disease. *Neuroimage* 21, 1443–1451.
- Schore, A. N. (1994). *Affect regulation and the origin of the self: The neurobiology of emotional development*. Mahwah, NJ: Erlbaum.
- Seppi, K., Schocke, M. F., Esterhammer, R., et al. (2003). Diffusion-weighted imaging discriminates progressive supranuclear palsy from PD, but not from the Parkinson variant of multiple system atrophy. *Neurology*, 60, 922–927.
- Seppi, K., Schocke, M. F., Mair, K. J., et al. (2006). Progression of putaminal degeneration in multiple system atrophy: a serial diffusion MR study. *Neuroimage*, 31, 240–245.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423.

- Sigman, M. & Cecchi, G. A. (2002). *Global organization of the Wordnet lexicon*, 99, 1742-1747.
- Simone, R. (2008). *Fondamenti di linguistica*. Bari: Laterza.
- Skinner, B.F. (1957). *Verbal Behavior*. Acton, Massachusetts: Copley Publishing Group.
- Snow, C. E. (1994). Beginning from baby talk: Twenty years of research on input and interaction. In Gallaway, C. R. (Ed.), *Input and interaction in language acquisition*, Cambridge, UK: Cambridge University Press.
- Snow, C.E. & Ferguson, C.A. (1977). *Talking to children: Language input and acquisition*, Cambridge: Cambridge University Press.
- Sporns, O., Tononi, G. & Edelman, G. M. (2000). Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex*;10:127–41.
- Sporns, O. & Kotter, R., (2004). Motifs in brain networks. *PLoS Biol.* 2, 1910–1918.
- Sporns, O. & Zwi, J. D. (2004). The small world of the cerebral cortex, *Neuroinformatics*, 2,145–62.
- Sporns, O., Tononi, G. & Kotter, R. (2005). *The human connectome: a structural description of the human brain*, *PLoS Comput Biol*, 1, 4, e42.
- Sporns, O. (2010). *Networks of the brain*. Cambridge, MA: The MIT Press.
- Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125.
- Stam, C. J. & Reijneveld, J.C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed Phys*, 1, 3.
- Stam, C.,J., Jones, B.F., Nolte, G., Breakspear, M. & Scheltens, P. (2007). Small-world networks and functional connectivity in Alzheimer's disease. *Cereb Cortex*, 17, 92–99.
- Stam, C.J., de Haan, W., Daffertshofer, A., Jones, B.F., Manshanden, I., et al. (2009). Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimers disease. *Brain*, 132, 213–224.
- Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Sciences*, 7, Issue 7, 308–312.
- Steyvers, M. & Tenenbaum J. (2004). Small worlds in semantic networks. *Cognitive sciences*, 29, 41-78.
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, 29, 41-78.

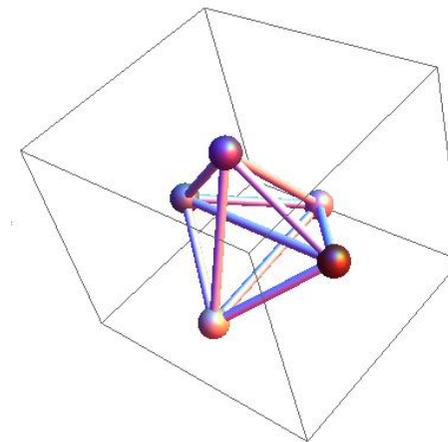
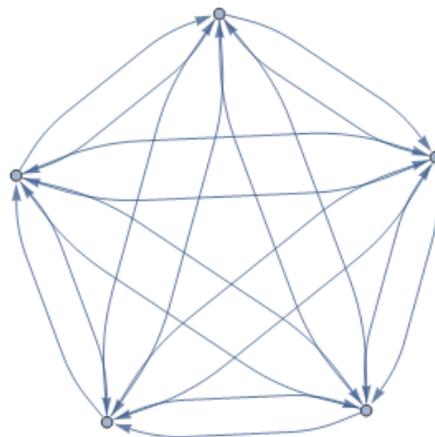
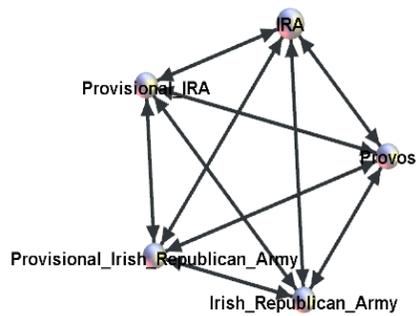
- Stewart, I. (2007). *Does God Play Dice? The New Mathematics of Chaos*. Oxford/London: Blackwell Publishing.
- Strogatz, S. H. (2001). *Exploring complex networks*. *Nature*, 410, 268-276.
- Sundberg, J. & Lindblom, B. (1991). Generative theories in language and music descriptions. In Howell, P. West, R. & Cross, I. (Eds.), *Representing musical structure*. London: Academic Press.
- Supekar, K., Menon, V., Rubin, D., Musen, M. & Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Comput Biol*, 4, e1000100.
- Thaut, M. H. (2005). *Rhythm, music and the brain: Scientific foundations and clinical applications*. New York: Taylor & Francis.
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*, Cambridge, MA: Bradford Books/The MIT Press.
- Tolman, E. C. (1981). *L'uomo psicologico*. Milano: Franco Angeli.
- Tomasello, M. (2003). *Constructing a Language*. Cambridge: Harvard University Press.
- Tomasello, M. & Slobin, D.J. (2005). *Beyond nature-nurture. Essays in honour of Elizabeth Bates*, Mahwah, N.J.: Lawrence Erlbaum.
- Tomkins, S. S. (1962). *Affect, imagery, consciousness: the positive affects*, Vol. 1, New York: Springer-Verlag.
- Travers, J. & Milgram, S., (1969). An experimental study of the small world problem. *Sociometry*, 32, 425-443.
- Trevarthen, C. (1992). *Le emozioni nell'infanzia: regolatrici del controllo e delle relazioni interpersonali*. Tr. It. Riva Crugnola, C., *Lo sviluppo affettivo del bambino*, Milano: Cortina Editore.
- Trevarthen, C. (1997). *Empatia e biologia: psicologia, cultura e neuroscienza*. Tr. it. Raffaello Cortina Editore : Milano, 1998.
- Tuch, D.S., Reese, T.G., Wiegell, M.R. & Wedeen V.J. (2003). Diffusion MRI of complex neural architecture, *Neuron*, 40, 885–95.
- Tulving, E., (1972). Episodic and semantic memory. In Tulving, E. & Donaldson, W. (Eds.), *Organization of memory*. New York: Academic Press.
- Tulving, E., (1989). Memory: performance, knowledge and experience. *European Journal of Cognitive Psychology*, 1, 3–26.

- Tulving, E. & Craik, F. I. M., (2000). *Handbook of Memory*. Oxford: University Press.
- Tulving, E. & Patkau, J. E., (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology*, 16, 83–95.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Clarendon Press.
- Tulving, E. (1985). How many memory systems are there? *American Psychologist*, 40, 385-398.
- Tulving, E. (2001). The origin of auto-noesis in episodic memory. In Roediger, H.L., Nairne, J.S., Neath, I. & Suprenant, A.M. (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp.17D34). Washington, D.C.: American Psychological Association.
- Turnbull, O. & Solms, M. (2002). *The brain and the inner world. An introduction to the neuroscience of subjective experience*, NY: Other Press.
- Valesio, P. (1969). *Il linguaggio*. Introduzione alla linguistica, Einaudi: Torino.
- Vihman, M. M. & McCune, L. (1994). When is a Word a Word? *Journal of Child Language*. 21, 517-542.
- Volterra, V., Caselli, M.C., Caprici, O. & Pizzuto, E. (2005). Gesture and the emergence and development of language. In Tomasello, M. S. (Eds.), *Beyond Nature-Nurture. Essays in Honor of Elizabeth Bates*, N. J.: Lawrence Erlbaum Associates.
- Voss, R.F. & Clarke, J. (1978). "1/f noise" in music: Music from 1/f noise. *J Acoust Soc Am*, 63, 258–263.
- Wallin, N.L., Merker, B. & Brown, S. (Eds.), (2000). *The origins of music*. Cambridge, MA: MIT Press.
- Wasserman, S. & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge Cambridge: University Press.
- Watson, J. (1985). *Il comportamentismo*. Firenze: Giunti.
- Watts, D. J., Strogatz, S. H. (1998). Collective dynamics of «small-world» networks. *Nature*, 393, 440-42.
- Wedeen, V.J. (1996). Diffusion anisotropy and white matter tracts. *Proc. Second Int. Conf. Funct. Mapp. Human Brain*, S146.
- Wedeen, V.J., Hagmann, P., Tseng, W.Y., Reese, T.G. & Weisskoff, R.M. (2005). Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magn Reson Med*, 54, 1377–86.

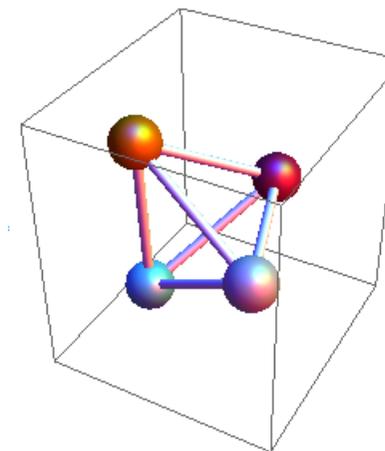
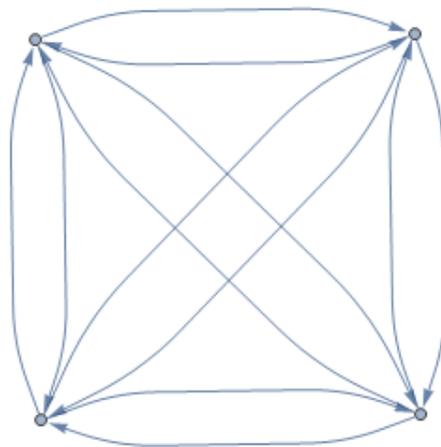
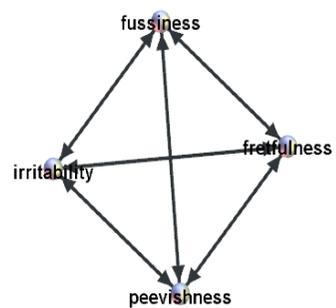
- Wedeen, V.J., Wang, R.P., Schmahmann, J.D., Benner, T., Tseng, W.Y., Dai, G., Pandya, D.N., Hagmann, P., D'Arceuil, H. & de Crespigny A.J. (2008). Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *Neuroimage*, 41,4,1267–77.
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, 34, 1289–1309.
- Wernicke, S. (2005). A faster algorithm for detecting network motifs. *Proc. WABI2005*, 1-12.
- Wernicke, S. & Rasche, F. (2006). FANMOD: a tool for fast network motif detection. *Bioinformatics* 22,1152-1153.
- Wiggins G. A. (2012). The Mind's Chorus: Creativity Before Consciousness. *Cognitive Computation*, 4, 3, 306-319.
- Wilson, R. J., Biggs, N. L. & Lloyd, E. K. (1999). *Graph Theory 1736-1936*. Oxford : Oxford University Press.
- Wittgenstein, L. (1953). *Philosophical Investigations*. New York : Macmillan.
- Zeki, S. & Shipp, S. (1988) The functional logic of cortical connections. *Nature*, 335, 311-317.
- Zhou, S., Hu, G., Zhang, Z. & Guan, J.(2008). An empirical study of Chinese language networks, *Physica A: Statistical Mechanics and its Applications*, 387, 12, 3039-3047.

Appendice: Particelle Emergenti

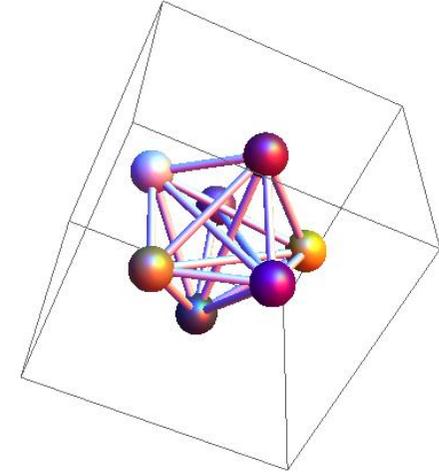
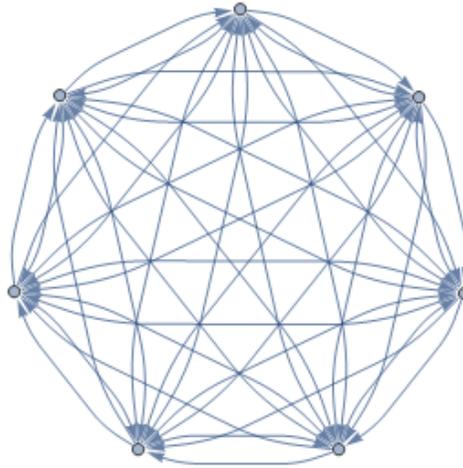
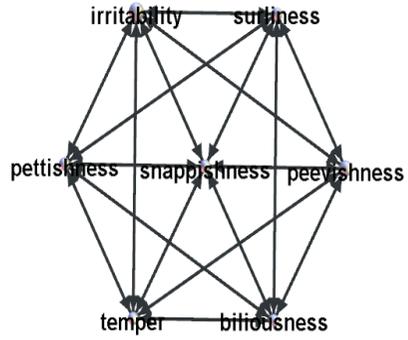
AngerPE_A1



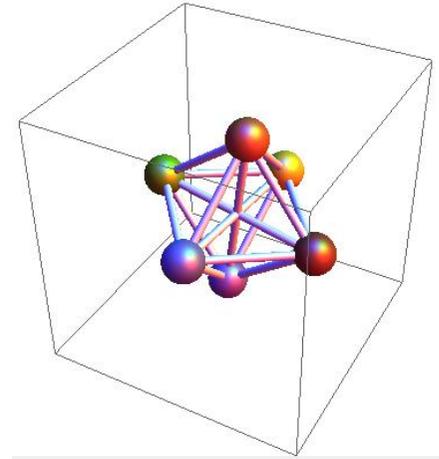
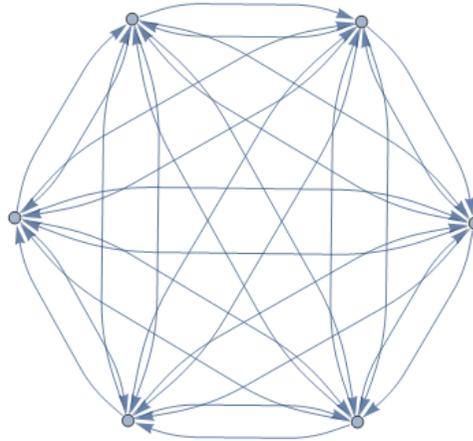
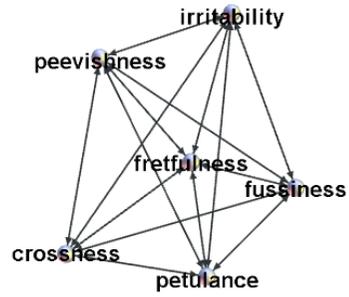
AngerPE_A2



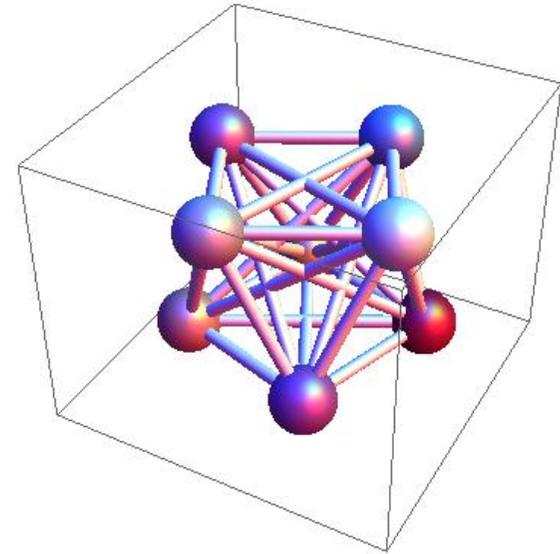
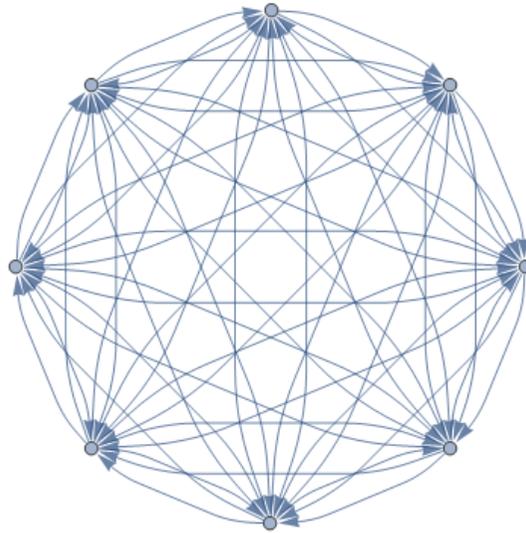
AngerPE_A3



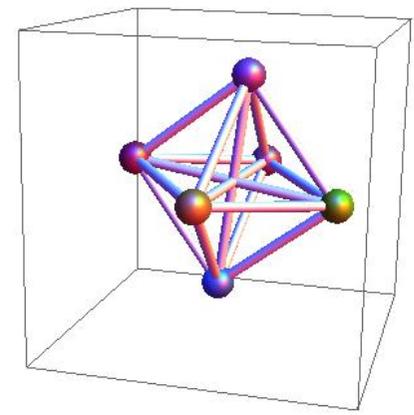
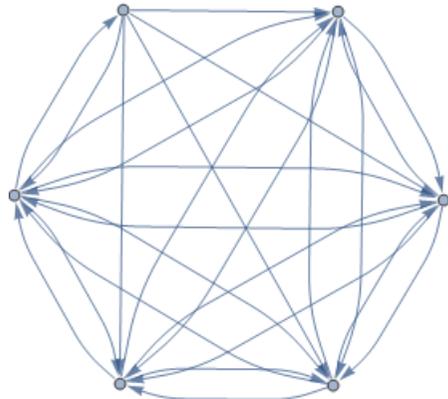
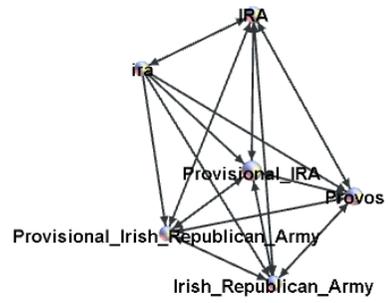
AngerPE_A4



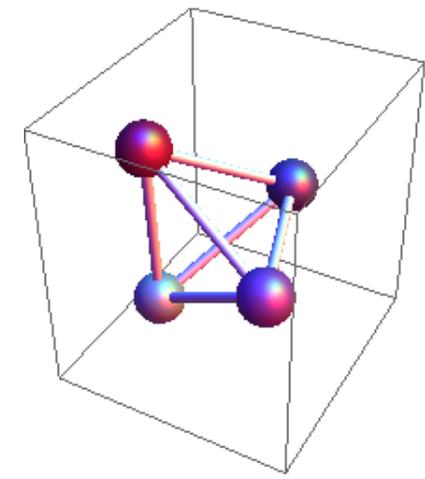
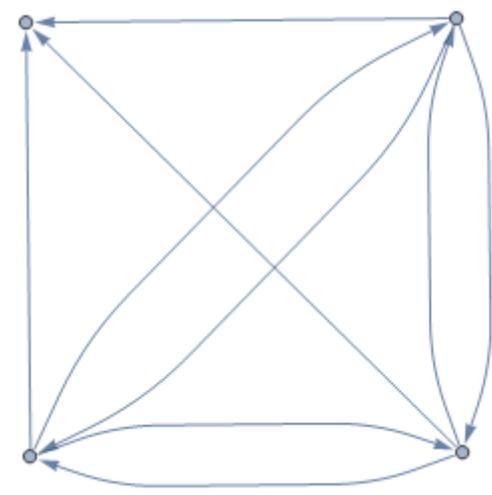
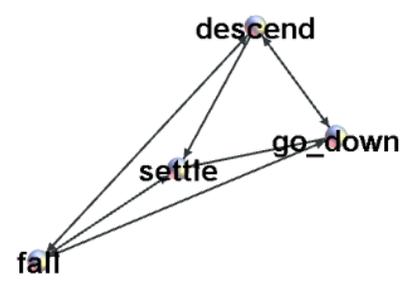
JoyPE_A1



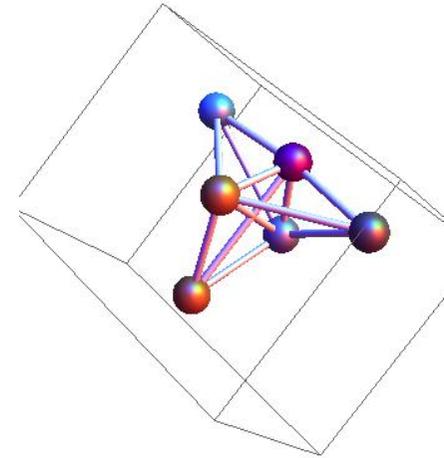
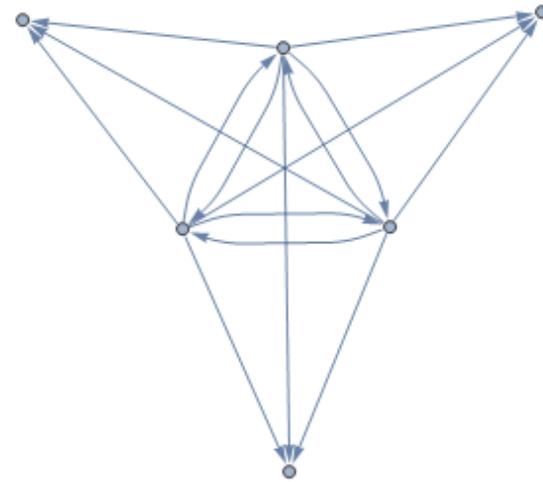
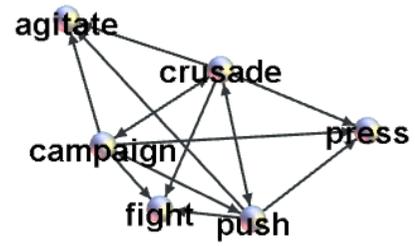
AngerPE_B1



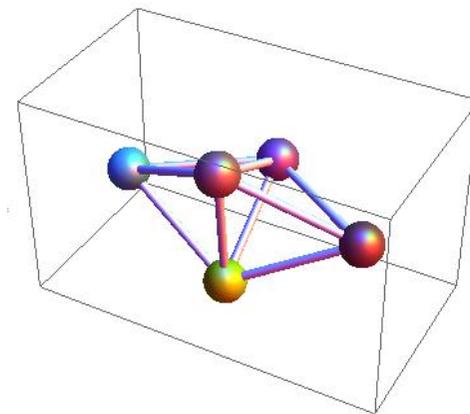
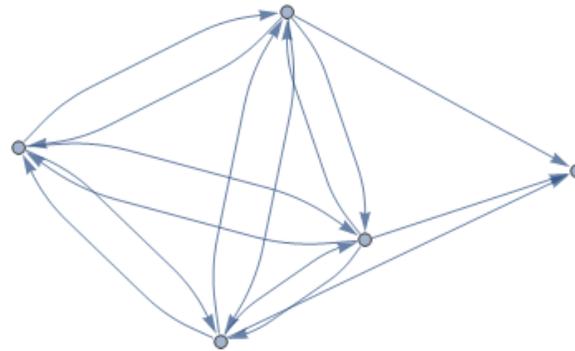
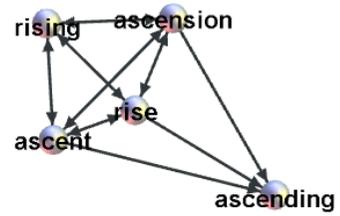
DisgustPE_B1



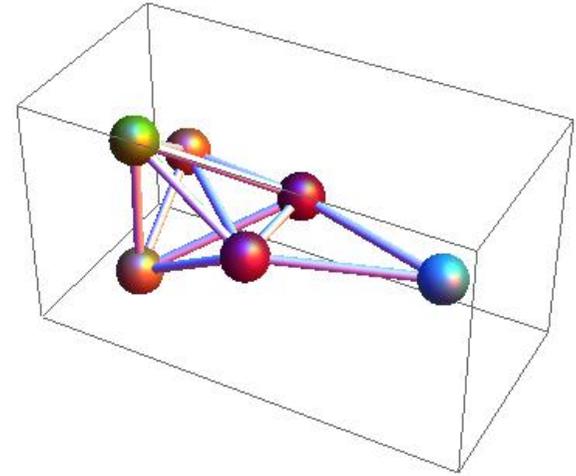
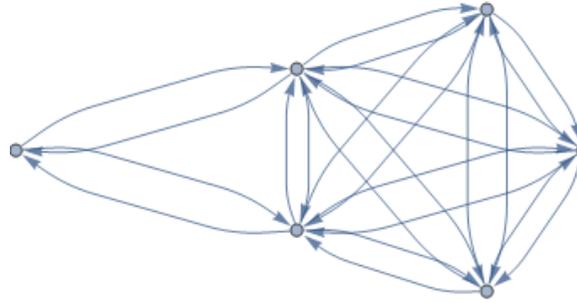
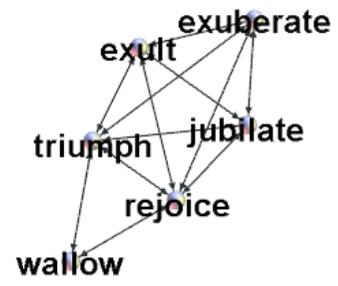
DisgustPE_B2



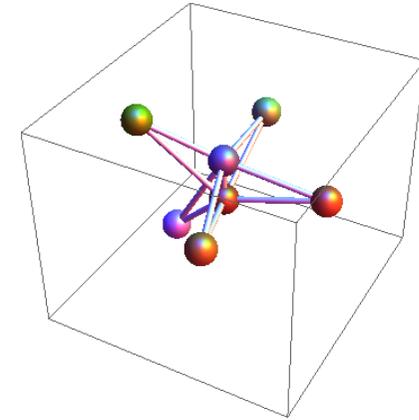
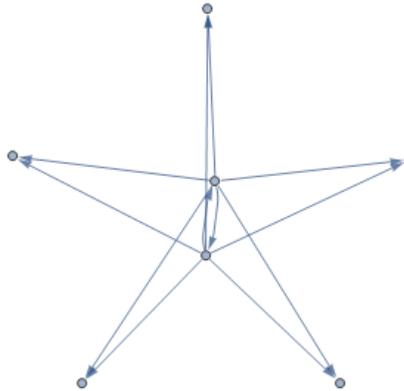
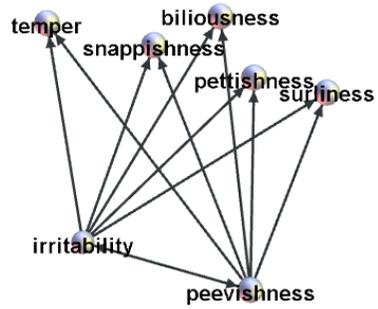
DisgustPE_B3



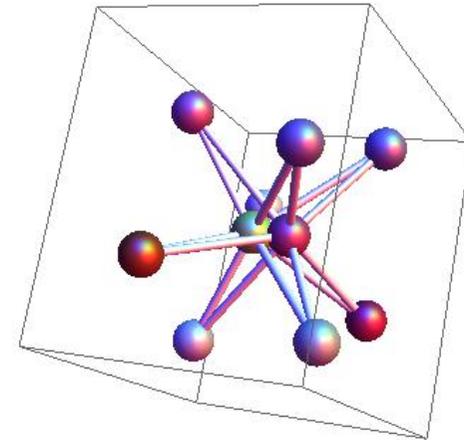
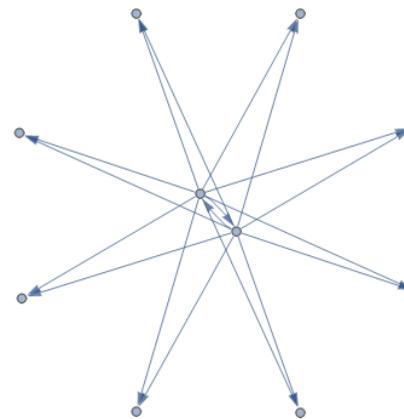
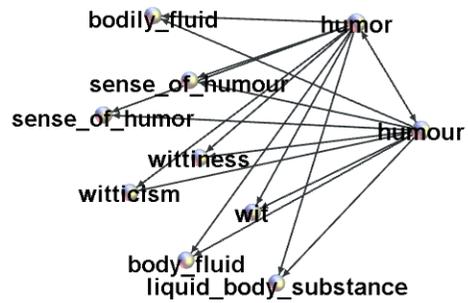
JoyPE_B1



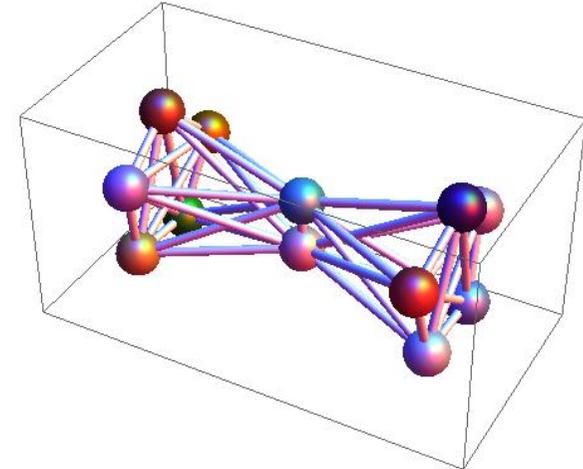
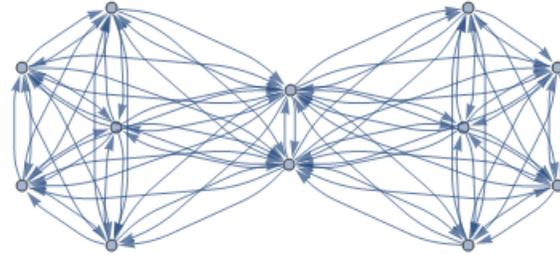
AngerPE_C1



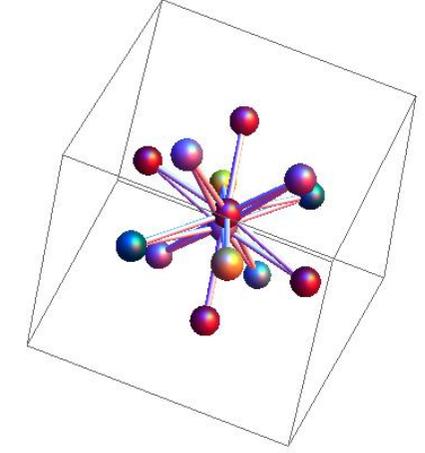
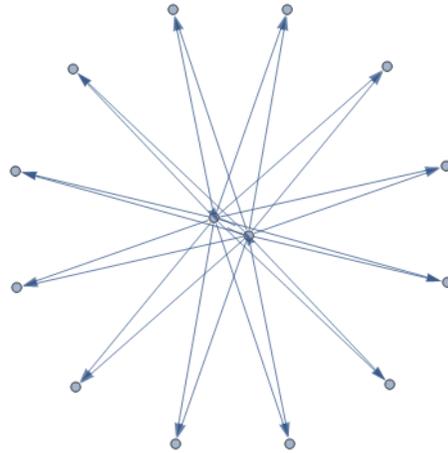
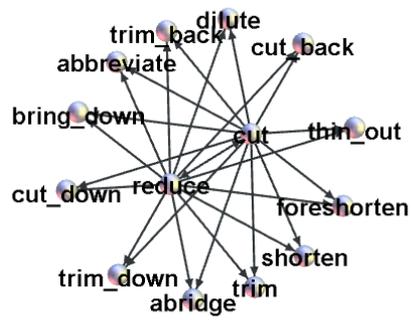
AngerPE_C2



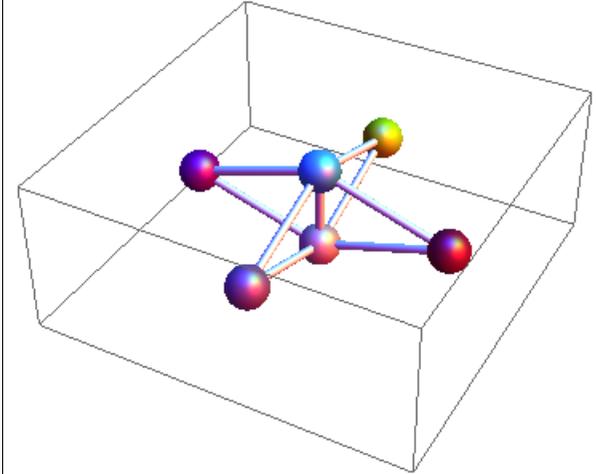
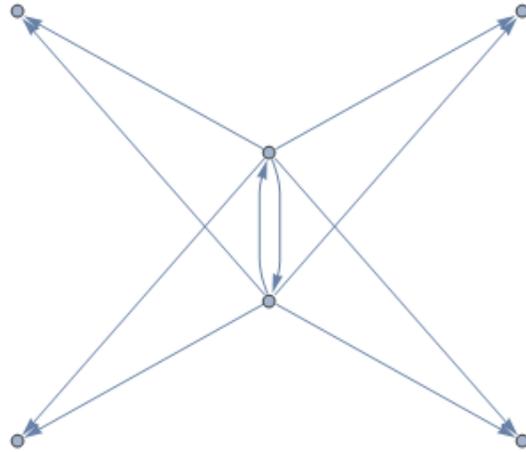
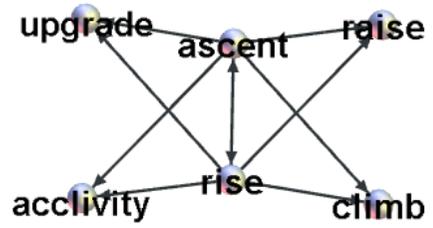
AngerPE_C3



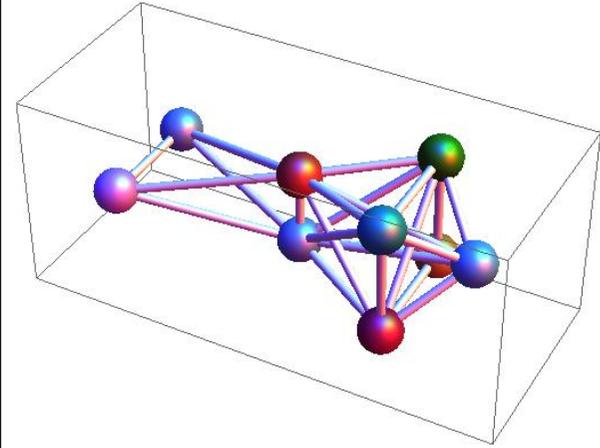
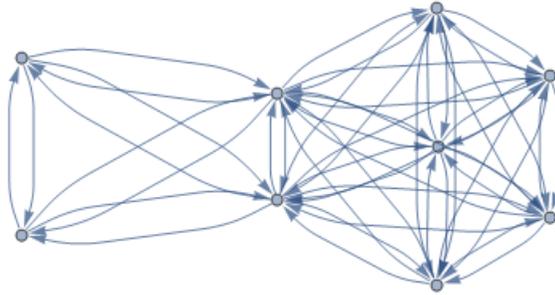
DisgustPE_C1



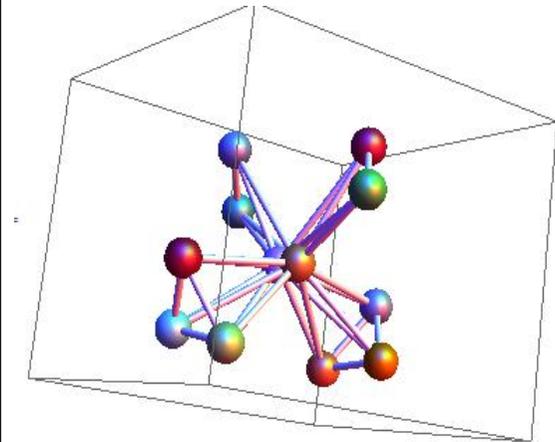
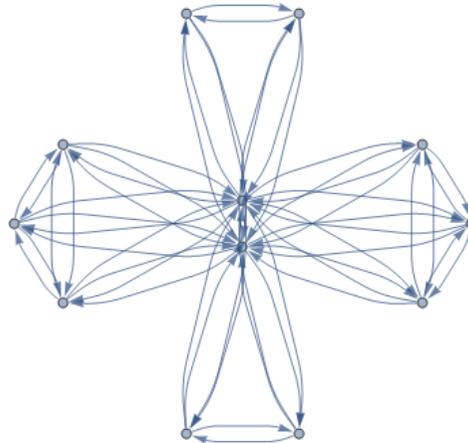
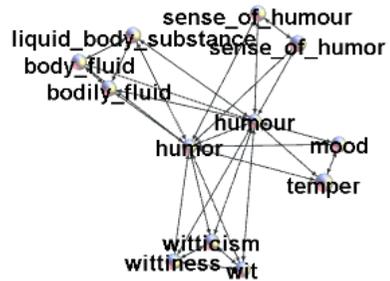
DisgustPE_C2



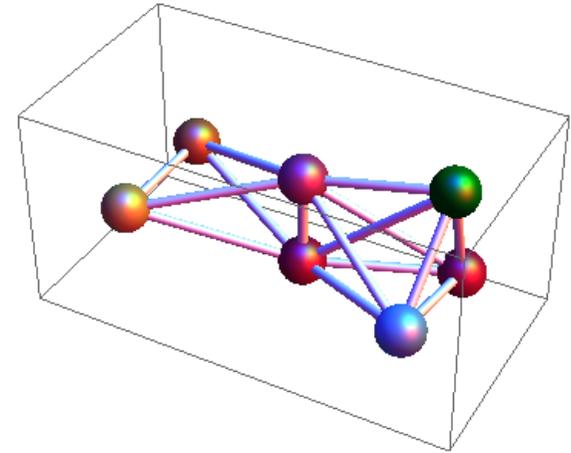
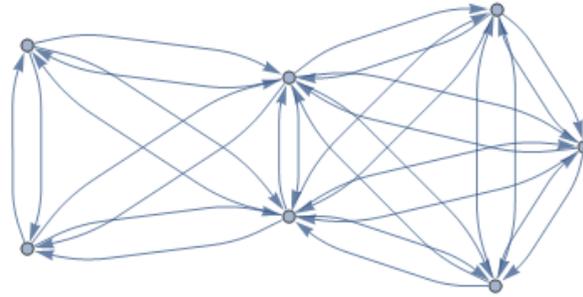
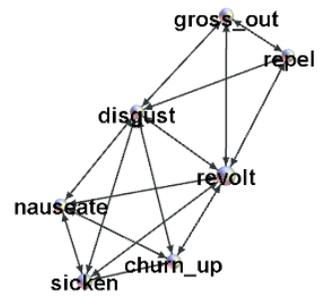
AngerPE_D1



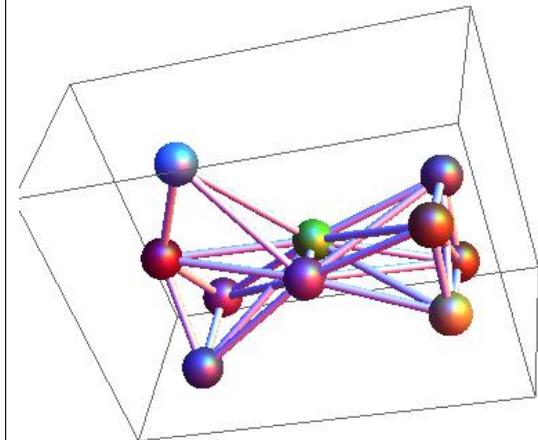
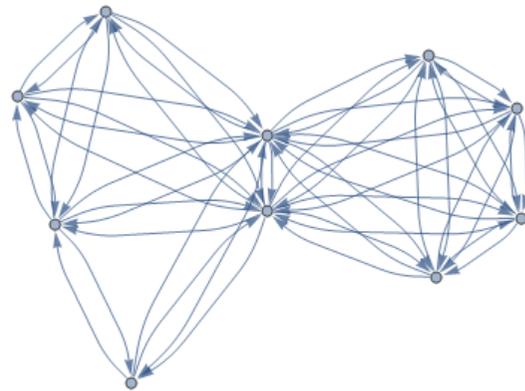
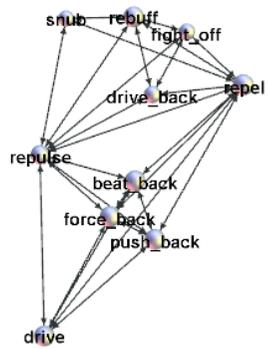
AngerPE_D2



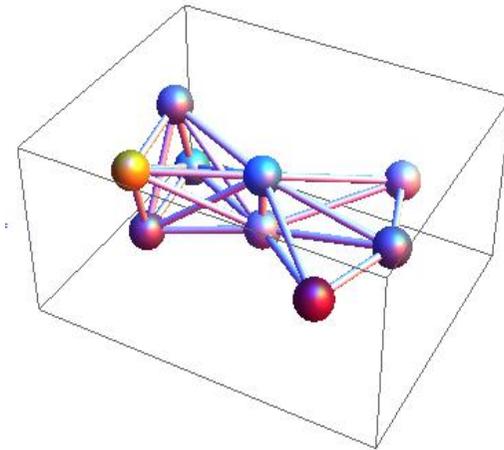
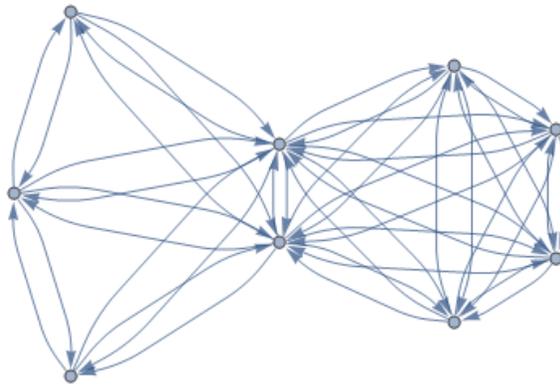
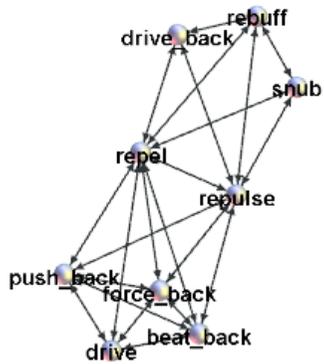
DisgustPE_D1



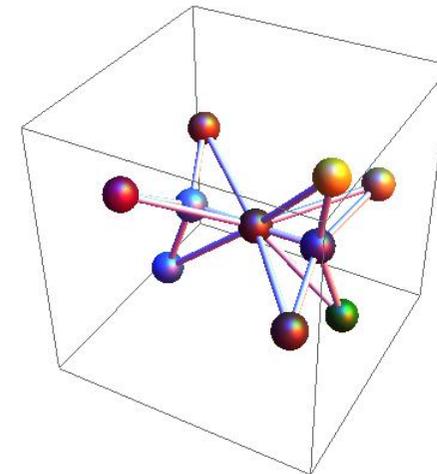
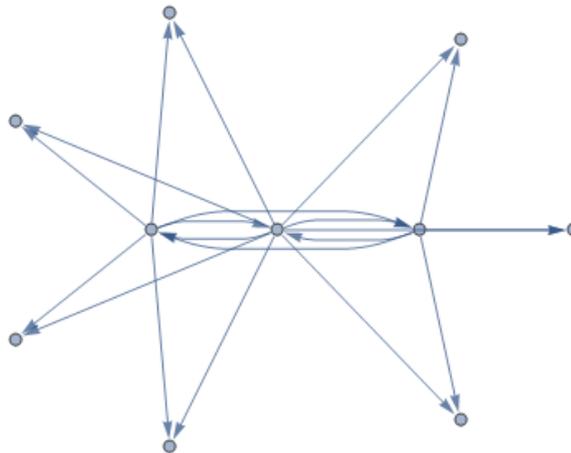
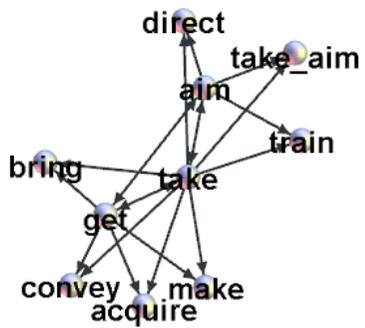
DisgustPE_D2



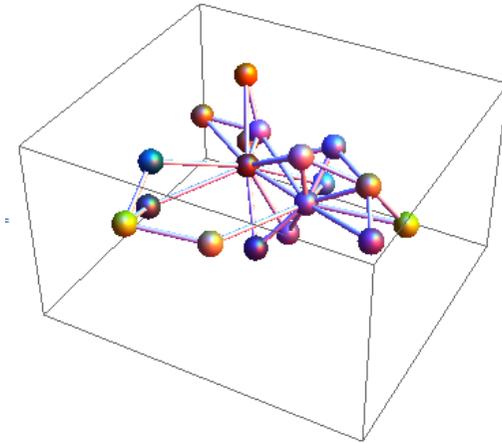
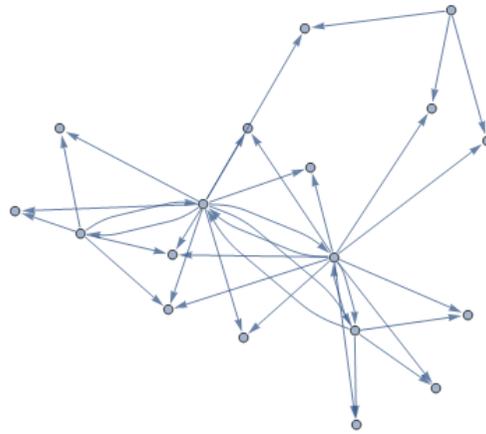
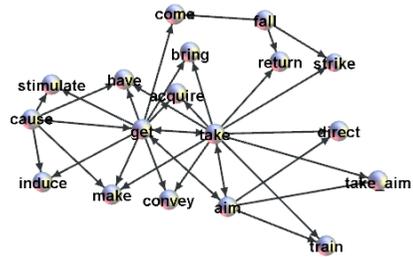
DisgustPE_D3



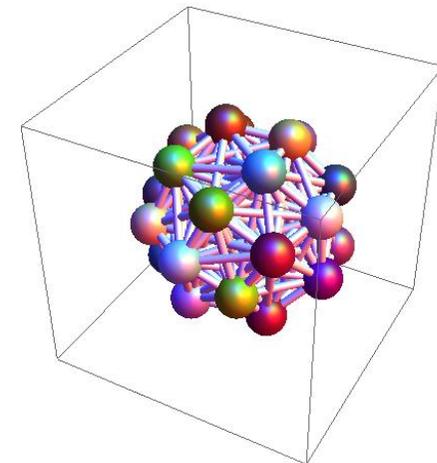
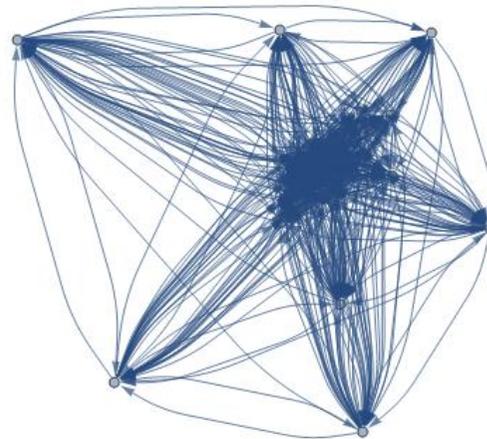
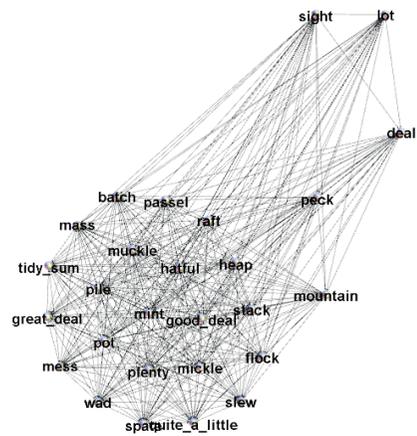
DisgustPE_D4



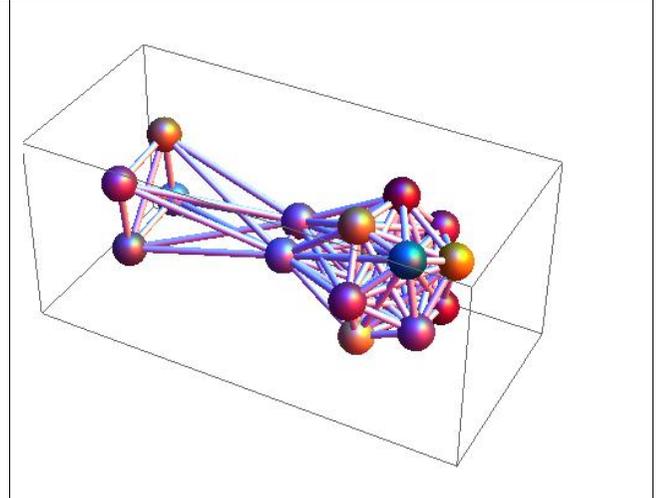
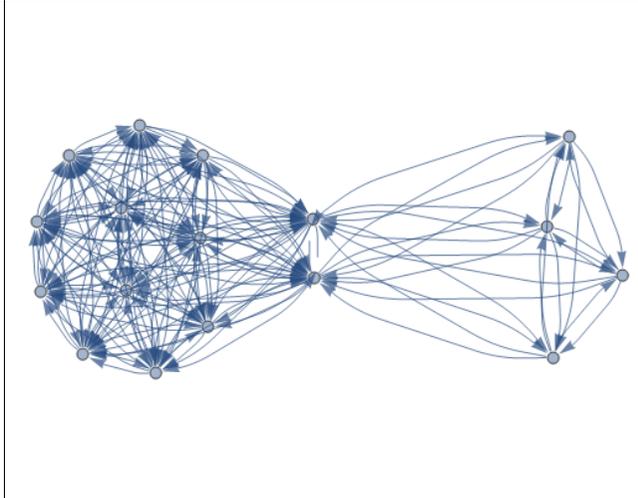
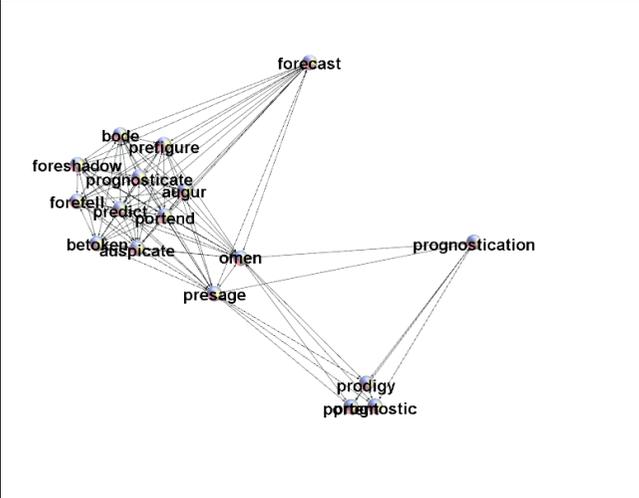
DisgustPE_D5



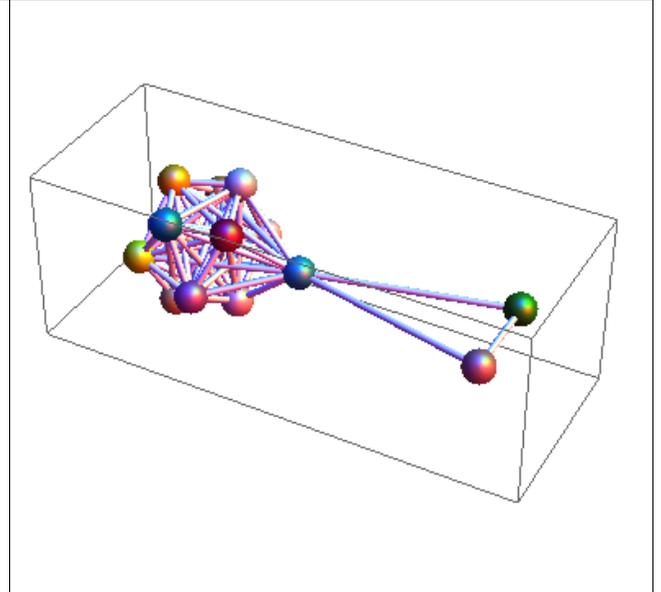
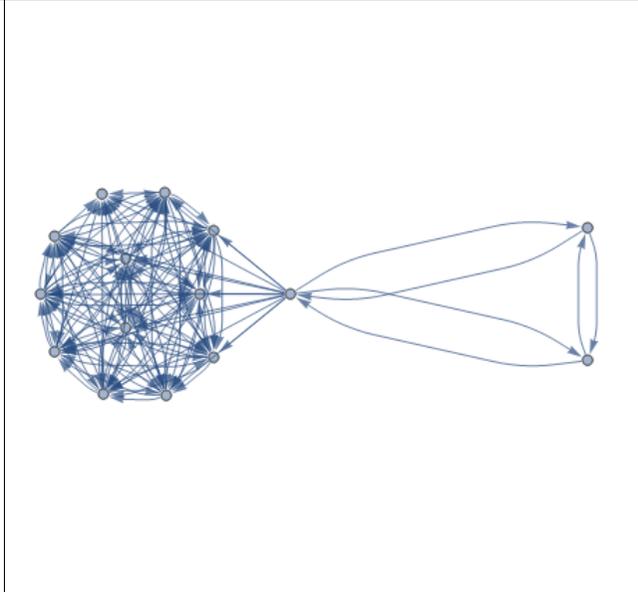
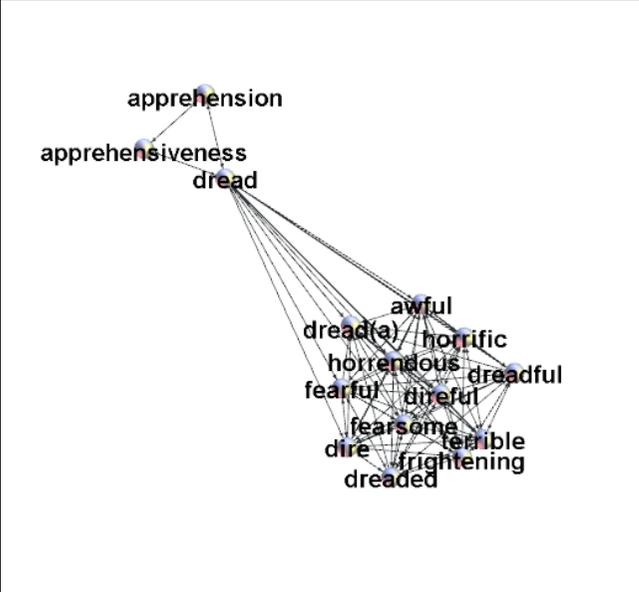
ExpectationPE_D1



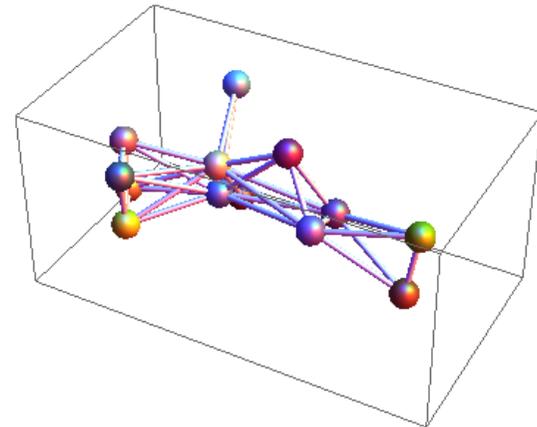
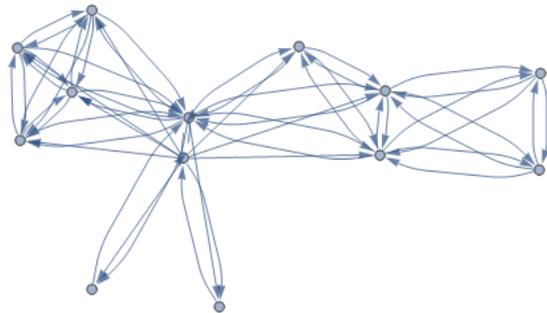
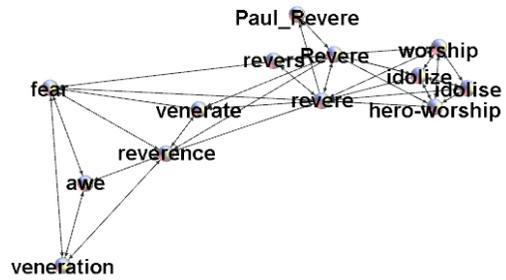
ExpectationPE_D2



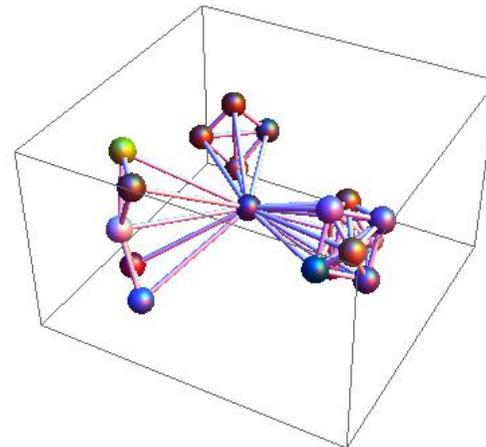
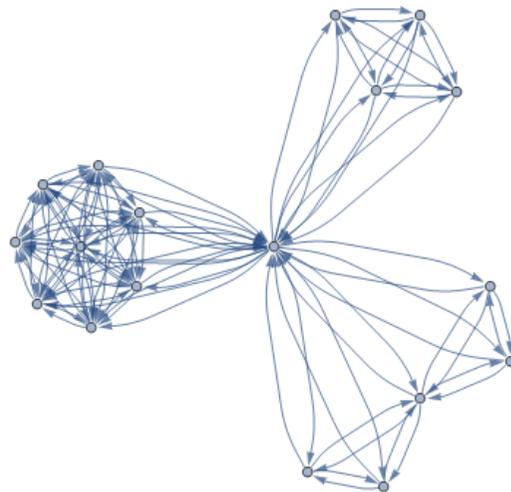
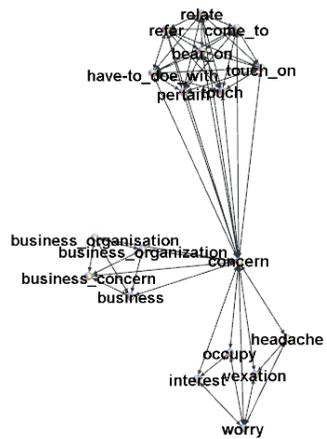
FearPE_D1



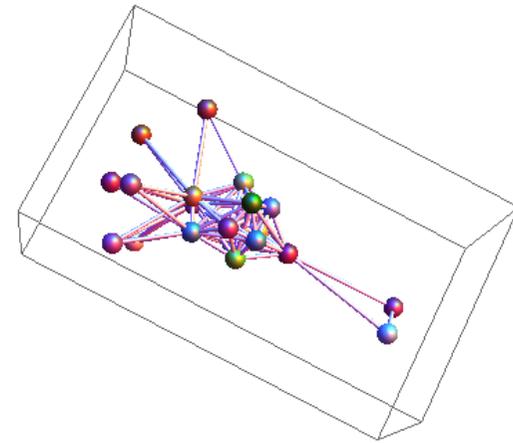
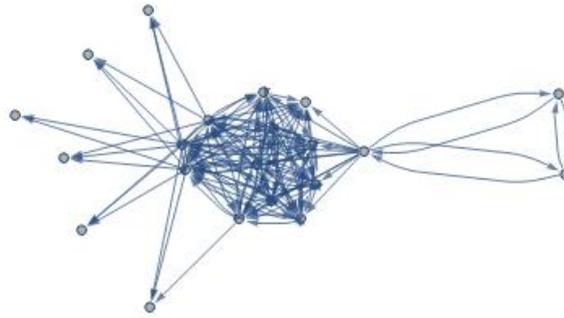
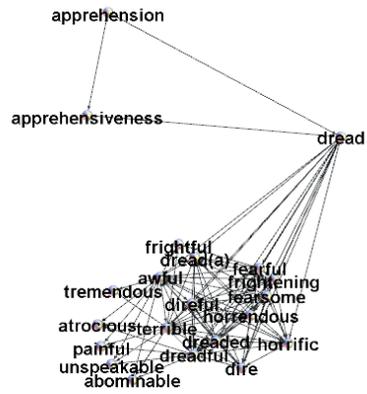
FearPE_D2



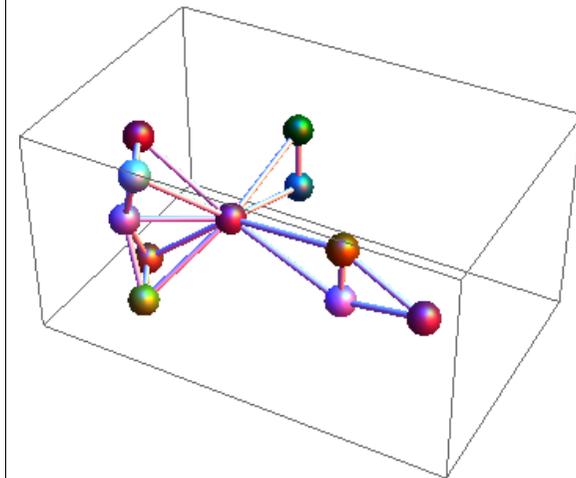
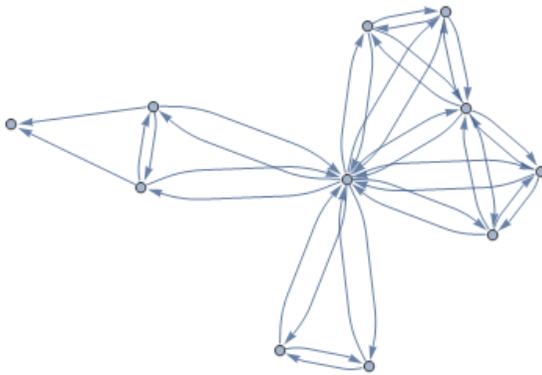
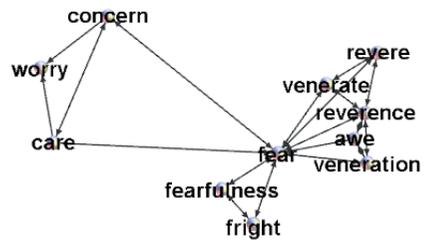
FearPE_D3



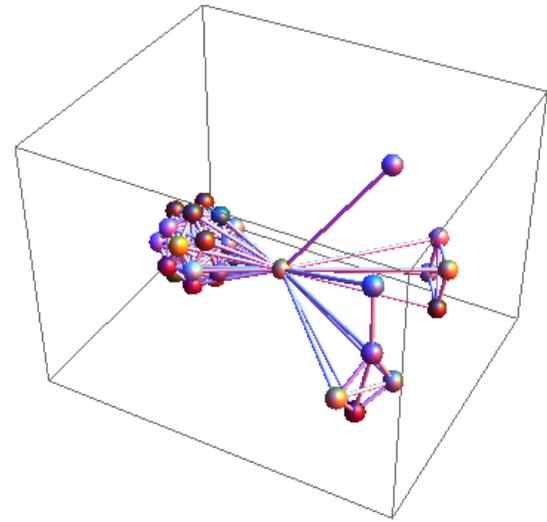
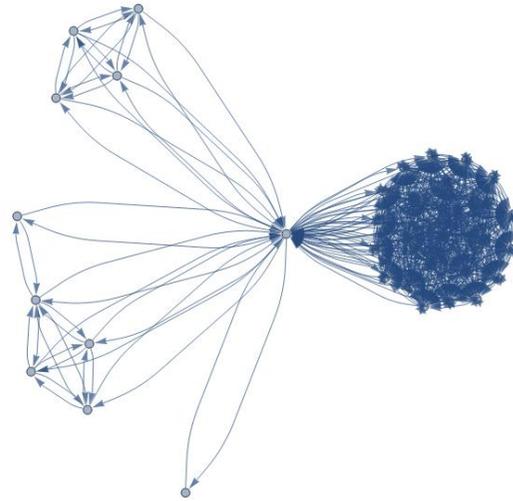
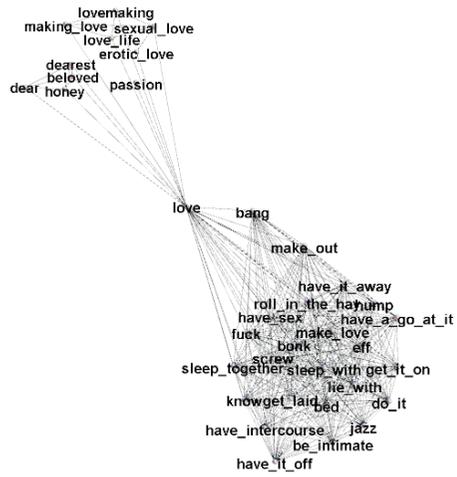
FearPE_D4



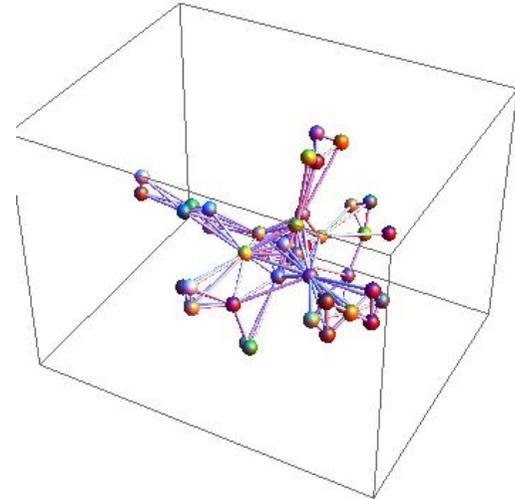
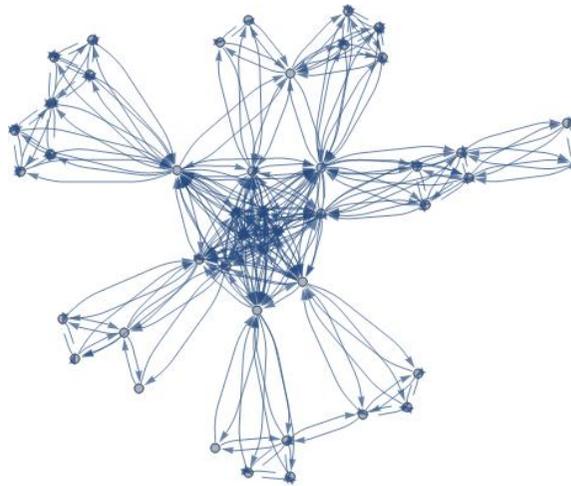
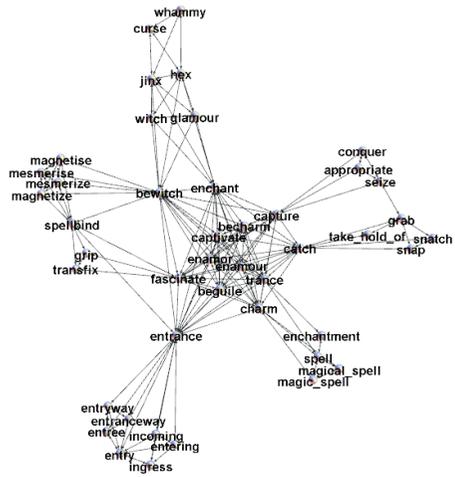
FearPE_D5



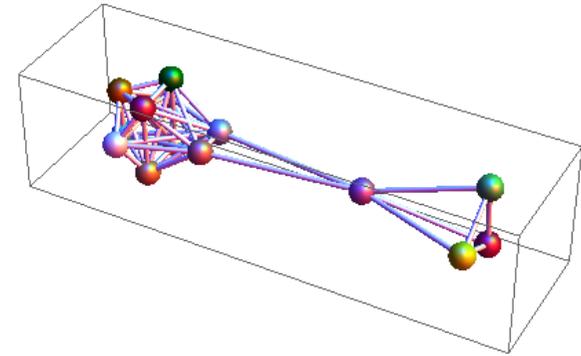
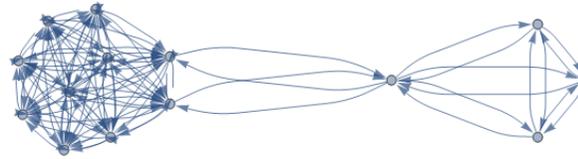
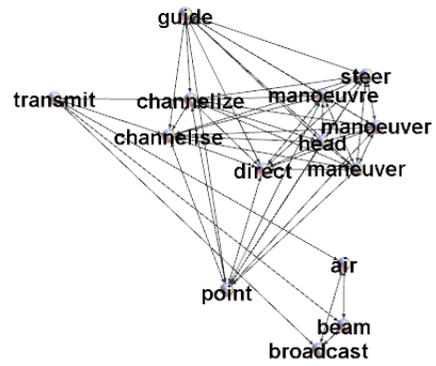
JoyPE_D1



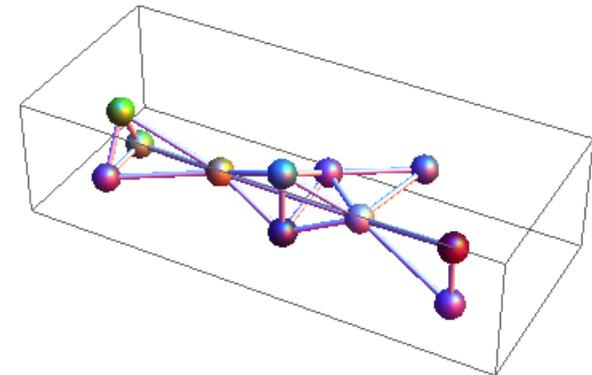
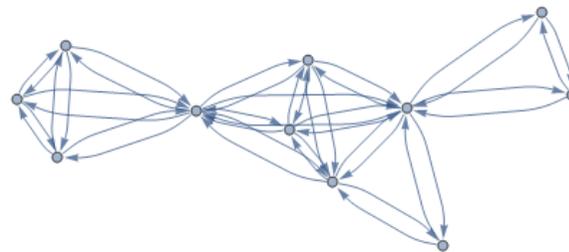
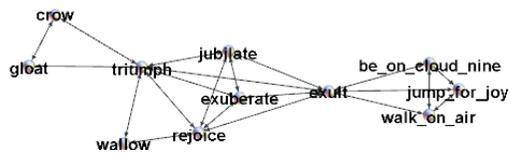
JoyPE_D2



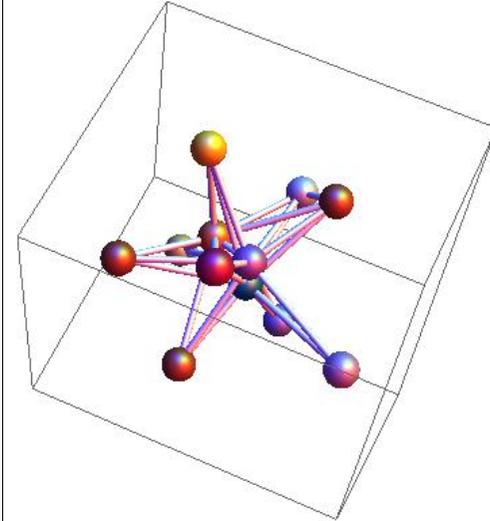
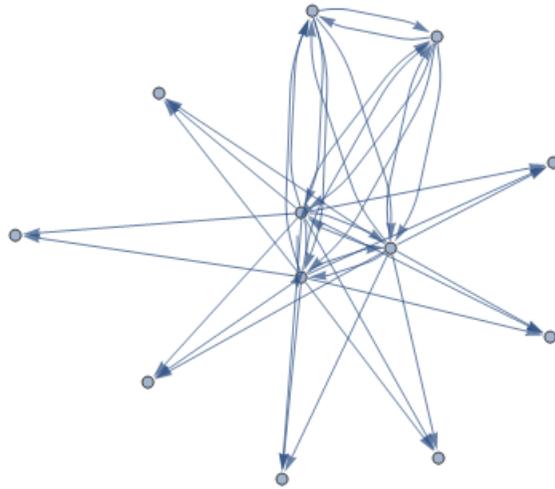
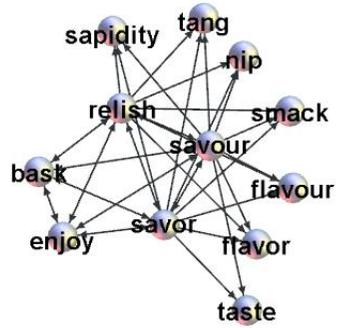
JoyPE_D3



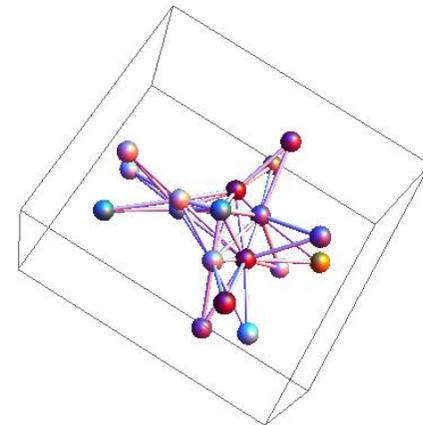
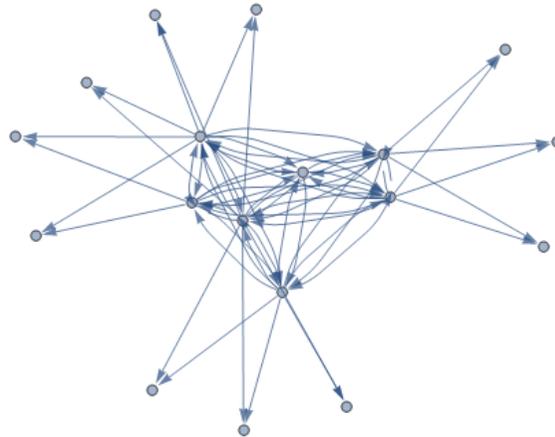
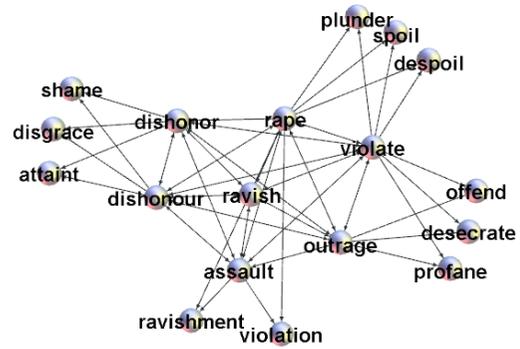
JoyPE_D4



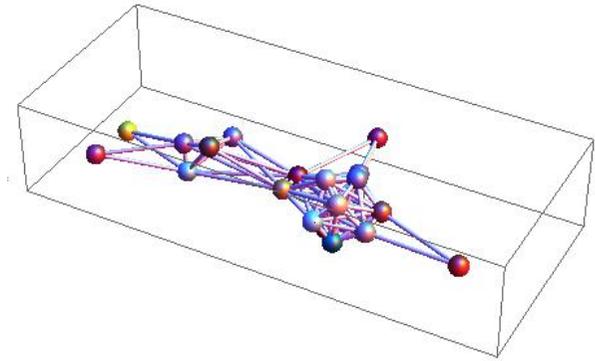
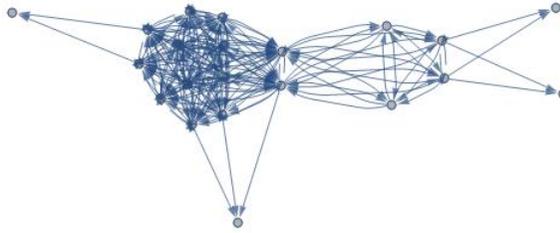
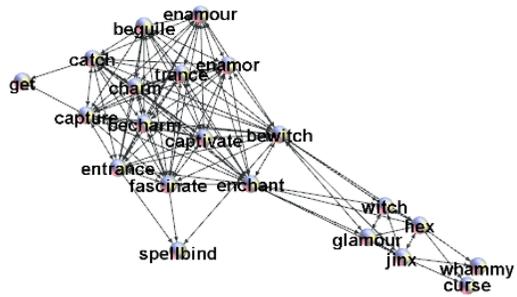
JoyPE_D5



JoyPE_D6



JoyPE_D7



JoyPE_D8

