UNIVERSITÀ DELLA CALABRIA

UNIVERSITA' DELLA CALABRIA

Dipartimento di Fisica

Scuola di Dottorato

Archimede

Indirizzo

Scienze e Tecnologie dei Sistemi Complessi

CICLO

XXVIII

SEMANTIC NETWORKS: MODELS AND APPLICATIONS

Settore Scientifico Disciplinare M-PSI/01

Direttore: Ch.mo Prof. (Pietro Pantano)
Firma

Supervisore: Ch.mo Prof. (Eleonora Bilotta)
Firma

Dottorando: Dott. (Antonio Procopio)
Firma

Abstract

Comprehend and model the human language is a problem of great importance for the modern society. Network science was already proved as a useful tool for this kind of study. In fact, many studies has been done in this direction, but none of them performed a deep investigation of the human memory and mental lexicon involving scripts. This thesis work propose a novel kind of network which maintain all the features observed in language and semantic networks, but is built in different steps and without classical approaches. This network is also used as a base to model a typicality score biased random walk model which performs good in language production and topic identification and can be seen as a prototype of an automatic system for these kinds of tasks.

Comprendere e modellare il linguaggio umano è un problema di grande importanza per la società moderna. La scienza delle reti è stata provata essere uno strumento utile per questo tipo di studi. Infatti, sono stati fatti molti studi in questa direzione, ma nessuno di questi esegue una profonda ricerca della memoria umana e del lessico mentale coinvolgendo gli scripts. Questo lavoro di tesi propone un nuovo tipo di rete che preserva tutte le caratteristiche osservate in reti linguistiche e semantiche, ma è costruita in diversi passi e senza l'utilizzo di metodi classici. Questa rete è stata anche utilizzata come base per modellare un random walk influenzato dal punteggio di tipicalità che dà buoni risultati nella produzione linguistica e nell'identificazione degli argomenti e che può essere visto come prototipo di un sistema automatico per questi tipi di compiti.

Contents

1 Introduction 3

2 Network Theory Background 6

2.1 Introduction 6

2.2 Definitions 7

2.3 Matrices 9

2.4 Measures 10

Hub and Authority 18

Page Rank 21

2.5 Communities 22

2.6 Small World and Scale Free 25

Small World 25

Scale Free 27

3 Script Based Processes 29

3.1 Introduction 29

3.2 Experiment 1 31

Results 32

3.3 Experiment 2 34

Results 35

Discussion 36

3.4 Network Analysis of Semantic Data 37

3.5 Memory for Scripts 40

Experiment 1 Networks: Grammatical Categories Occurences . . . 40
Experiment 2 Networks, First Stage Part A: Synonyms 40
Experiment 2 Networks, First Stage Part B: Synonyms of Synonyms 42
Experiment 2 Networks, Second Stage: Hyponyms and Hypernyms 42
Macro Level Analysis 43
Small World and Scale Free 44
Distance and Degree Correlations 49
3.6 Conclusions 52

4 Mental Lexicon and Language Structures 57
4.1 Introduction 57
4.2 Modularity and Community Detection 58
4.3 Modeling 59
Free Association Norms 59
Random Walk Model 61
Neighborhoods 62
Semantic Fields Features 65
Analysis 66
4.4 Different Random Walk Models 68
Simulation Runs 70
Results 70
The computational system 77
4.5 Conclusions 81

5 Conclusions 86

6 Appendix: Script tables and degree distribution fits 88

Chapter 1

Introduction

One of the main problems of contemporary society is the comprehension of natural language. While informatics tried to elaborate formal languages with the aim to model the human cognition's semantic structure, network science allowed to look at the linguistic production, organization and recall in several different ways.

This thesis work's main purpose is to provide mathematical models to analyze mental lexicon development and some of the processes needed to memorize linguistic configurations tied to different environments in which humans acts and live. These everyday life situations are referred to as scripts and are modeled with agents which extract the meaning from the above mentioned structures and configurations. In fact, these structures are the main cognitive database embodied in the human brain.

There are many and important advantages derived from network science in many different fields: for example, the comprehension of human brain connectivity (Sporns, 2011), the classification of psychological disorders (Cramer, Waldorp, van der Maas, & Borsboom, 2010), the semantics representation in the human brain (Huth, Nishimoto, Vu, & Gallant, 2012), the study of memory's semantic organization (Griffiths, Steyvers, & Firl, 2007; Steyvers & Tenenbaum, 2005) even at an early stage in children (Hills, Maouene, Riordan, & Smith, 2010; Hills, Maouene, Maouene, Sheya, & Smith, 2009) and the study of word associations even between seemingly unrelated words (De Deyne, Navarro, Perfors, & Storms, 2012).

Many models were proposed to study and represent this kind of word association: semantic knowledge networks (Collins & Loftus, 1975; Quillian, 1967), networks of language (Lamb, 1970; MacKay, 1992), neural networks (Rosenblatt, 1958), propositional networks which are able to grasp the meanings of sentences (Anderson, 2005).

Semantic networks seemed to be the best way to model how the human brain works (Collins & Quillian, 1969; Quillian, 1967). Every word forming the mental lexicon was connected with a set of suggestions (pointers) which form the connection with other terms in the memorized mental lexicon. Usually, many paths start with a simple word association rule and end with a complex one. If a concept is activated in the network, all of its relations are activated, too. This process is called *spreading activation* (Collins & Loftus, 1975): at first only nearest neighbors are activated, while going on with the process, even distant words are activated because they are neighbors of neighbors.

This model was not enough because it does not contain any long range connection and it is not suitable to organize a mental lexicon for artificial intelligences. Then a higher level structure was introduced: clusters of words directly connected according to some rule. These structures are called frames (Charniak, 1972; Minksy, 1975), schemata (Rumelhart & Ortony, 1977) or *scripts* (R. C. Schank & Abelson, 1977).

Scripts contain human behavior description, together with inferences and decision processes present in the human brain from the age of fifteen months (Onishi, Baillargeon, & Leslie, 2007). Because of the repetition of the experiences, inferences are created based on the expected events. This process is very important for the semantic and conceptual systems' development. But a main question remains: is this the main growth dynamic of these systems, or there are some more relevant?

Network science take into consideration every complex system as a set of interconnected elements and, as said before, brought advancements in several fields: economy and biology (Barabási et al., 2009), social science (Watts, 2004), psychology, with the introduction of the term *connectome* (Sporns, 2010) which changed the way of looking at the cognitive structures as a network (MacKay, 1992; Quillian, 1967; Rosenblatt, 1958).

Several studies were performed using these concepts, but a deep investigation which relate scripts and mental lexicon organization to model the whole system as a network is needed and is the subject of this thesis.

Every chapter of this thesis, with the exception for the next one which gives the needed network science formal background, is mainly divided in two parts: a theoretical one and an experimental one.

The theoretical parts give an introduction on semantic networks and mental lexicon development from the cognitive point of view, together with the state of the art about network science applied to linguistics.

The experimental parts report the data acquired in two main experiments performed at the University of Calabria: these experiments concern a creative linguistic process about words association in a specific context. This collection of words have been modeled as networks which have been analyzed and used as a basis for experimental mathematical models presented in this thesis.

Chapter 2

Network Theory Background

2.1 Introduction

In mathematics, a network (or graph) is a representation of objects interconnected with some relation. Leonard Euler introduced this mathematical framework in 1736 to solve the seven Konigsberg's bridges problem. Konigsberg (today known as Kaliningrad) is split in four zones by the Pregel river. These zones were connected by seven bridges (today two of those bridges are no more). The problem was telling if it was possible to walk through the whole city crossing each bridge exactly one time and arrive at the starting point of the walk. To solve the problem, Euler represented the city as a graph: the four parts were pictured as points (nodes or vertices) with seven segments (links or arcs) linking them in the same way the bridges connected the four zones. With this graph, Euler proved the non existence of the walk described in the problem.

During the last years, graphs acquired great importance in many research fields thanks to the advance in network science.

The approach of network science has spread out from its original domain of application to social science (Watts, 2004) to a great variety of domains, such as economy, biology and technology (Barabási et al., 2009). Psychological sciences have used network approach to improve the understanding of the brain "connectome" (Sporns, 2010), to diagnose mental disorders (Cramer et al., 2010), or to represent semantic memory (Steyvers & Tenenbaum, 2005), changing the network-analogy meaning, usually used in the cognitive domain (MacKay, 1992; Quillian, 1967; Rosenblatt, 1958). The power of network science methods has already been proven by different works in the cognitive domain, at the level of human semantic knowledge, trying to investigate the connectivity of semantic networks (Steyvers & Tenenbaum, 2005), at the word learning and lexical retrieval in the mental lexicon (Vitevitch, 2008), comparing networks in different languages (Arbesman, Strogatz, & Vitevitch, 2010a), offering a huge variety of statistical and computational tools. These studies put in evidence many interesting features of language-related networks, detecting large highly interconnected components and many isolated islands, isolated words with no neighbors, small world features (related to a short average path length and a high clustering coefficient, as showed for other systems in (Watts & Strogatz, 1998), degree distribution deviating from a power law distribution (i Cancho, 2005), so different from scale free networks (Barabási & Albert, 1999). Though some of these studies dealt with what a speaker/hearer knows about the form of the language entry (its phonology), its structural complexity (morphology), its meaning (its semantic representation) and its combinatorial properties (its syntactic, categorical properties), a deep investigation of cognitive organization that refers to scripts (R. Schank, 1982; R. C. Schank & Abelson, 1977) through networks has never been done. The script model, originally developed by R. C. Schank and Abelson (1977), stated that, for the purposes of text comprehension (Bower, Black, & Turner, 1979; Cellar & Barrett, 1987; Pollatsek, Ashby, & Clifton Jr, 2012), young and adult memory recall (Light & Anderson, 1983), language comprehension (Gernsbacher, 1991; Zwaan & Radvansky, 1998), memory organization (Anderson, 1983), methods of investigation of cognitive processes (Abelson, 1981) and behavior (Graesser, Gordon, & Sawyer, 1979), memory system in subjects with neurodegenerative diseases (Grafman et al., 1991), social behavior evaluation in interacting people (Abelson, 1976) and other high-level processing tasks, mind retrieves information from long-term memory in the organized form of scripts or schemata (Allington, 2005).

2.2 Definitions

Definition 2.1. A graph G over N is defined as an ordered couple of sets $G(N, L)$ where $N = \{n_1, n_2, \ldots, , n_n\}$ is the set of nodes, while $L = \{l_1, l_2, \ldots, l_m\}$ is the sets of the arcs. N can be finite or infinite and so will be the graph over it.

Every arc is represented by the couple of nodes which are its endpoints. These nodes are connected by the arc. So the arc linking n_i with n_j can be written as $n_i n_j$ or l_{ij}. An arc linking a node with itself is called loop.

Definition 2.2. Two nodes are adjacent (or neighbors) if they are linked by an arc. Two arcs are adjacent if they share an endpoint.

Definition 2.3. Given a graph $G(N, L)$, a path is a sequence of adjacent arcs. A minimum path from n_i to n_j is a minimal path starting from a node n_i and arriving to a node n_j. The minimum path is not unique.

Definition 2.4. A graph $G(N, L)$ is connected if $\forall\ n_i, n_j \in N$, with $i \neq j$, exists a path from n_i to n_j.

Definition 2.5. A graph is complete if all of its nodes are adjacent.

Definition 2.6. Given a graph $G(N, L)$, the set $N' \subseteq N$ and the set $L' \subseteq L$. The graph $G'(N', L')$ is a subgraph of G, while G is a supergraph of G'.

Definition 2.7. The neighborhood of a node n_i in a graph G is the subgraph over the set of its neighbors, itself included, with all the links among them existing in G.

Definition 2.8. An arc l_{ij} is directed if it links n_i with n_j, but not the opposite. Otherwise it is undirected.

Definition 2.9. A graph is directed if at least one of its arcs is directed. If all of its arcs are undirected, the graph is undirected, too.

Definition 2.10. A graph is weighted if every arc is given a non zero value (weight).

In weighted graphs, it is possible to tell the different importance of different arcs because of their weights. In unweighted graphs, instead, all the arcs have the same importance and can be considered as weighted graphs with unitary weights.

2.3 Matrices

It is possible and useful to represent a graph with a matrix.

Definition 2.11. Given an unweighted graph $G(N, L)$, with $|N| = n$, its adjacency matrix $A = \{a_{ij}\}$ is a $n \times n$ square matrix defined as:

$$a_{ij} = \begin{cases} 1 & \text{if } l_{ij} \in L \\ 0 & \text{otherwise} \end{cases}$$

If the graph is undirected, then its adjacency matrix is symmetric: in fact in an undirected graph if n_i is connected to n_j, then n_j is connected to n_i, so $a_{ij} = a_{ji}\ \forall\, i, j = 1, 2, \ldots, n$. If there are no loops, the adjacency matrix diagonal contains all zero entries: $a_{ii} = 0\ \forall i = 1, 2, \ldots, n$. The number of edges $|L| = m$ can be computed directly from A:

$$m = \begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} & \text{if } G \text{ is directed} \\ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} & \text{if } G \text{ is undirected} \end{cases}$$

If the graph is weighted, the definition is slightly different and the adjacency matrix is usually called weights matrix or weighted adjacency matrix $W = \{w_{ij}\}$, defined as:

$$W_{ij} \begin{cases} w_{ij} & \text{if } l_{ij} \in L \\ 0 & \text{otherwise} \end{cases}$$

where w_{ij} is the weight of the arc l_{ij}.

Another useful matrix associated to a graph is the Laplacian matrix.

Definition 2.12. Given a graph $G(N, L)$ with $|N| = n$ and adjacency matrix A, its Laplacian matrix $\Lambda = \{\Lambda_{ij}\}$ defined as:

$$\Lambda_{ij} \begin{cases} \left(\sum_{k=1}^{n} a_{kj}\right) - a_{ij} & \text{if } i = j \\ -a_{ij} & \text{if } i \neq j \end{cases}$$

The Laplacian matrix is useful if loops are not important: Λ does not change if loops are added to the graph. If the graph is undirected and it has k connected components (maximal connected subgraphs), 0 is an eigenvalue of Λ with algebraic multiplicity equal to k. Thus, Λ can be written in diagonal blocks form: nodes can be reordered to have k non zero squared blocks along the diagonal. In this case, there are k eigenvectors, one for each connected component, with non zero entries corresponding to the nodes belonging to the connected component and zero elsewhere. So, building a $n \times k$ matrix with these eigenvectors, each row represents a node and a direction of a k-dimensional space.

2.4 Measures

Defining some measures is necessary to study a network, both qualitatively and quantitatively. Measures can be local or global: the first characterize every node or every arc, while the second one describe the network as a whole.

Degree

Definition 2.13. In an undirected network with n nodes and adjacency matrix A, the degree k_i of the node n_i is the number of edges starting from or arriving to n_i:

$$k_i = \sum_{j=1}^{n} a_{ij}$$

While the degree is a local measure, it is also possible to consider the average degree as a global measure. The average degree of the network is the average of the degrees:

$$\bar{k} = \frac{\sum\limits_{i=1}^{n} k_i}{n}$$

If the network is directed, it is needed to distinguish ingoing arcs from outgoing ones: if an arc goes from n_i to n_j, then it is ingoing in n_j and outgoing from n_i. Thus, it is possible to define three different degree measures for directed networks.

Definition 2.14. In a directed network with n nodes and adjacency matrix A, for the node n_i are defined:

- indegree k_i^{in}, the number of ingoing arcs;
- outdegree k_i^{out}, the number of outgoing arcs;
- total degree k_i^{tot}, the sum of indegree and outdegree.

$$k_i^{in} = \sum_{j=1}^{n} a_{ij}$$

$$k_i^{out} = \sum_{j=1}^{n} a_{ji}$$

$$k_i^{tot} = k_i^{in} + k_i^{out}$$

In a weighted network with weights matrix W the same formulas are still valid if not the degree k_i, but the strength s_i of the node n_i is considered.

Definition 2.15. In a weighted network the strength s_i of the node n_i is defined as the sum of the weights of the arcs starting from or arriving to n_i.

For the directed and weighted case, the definition can be easily extended from the non weighted case.

Path length

The distance d_{ij} between two nodes n_i, n_j is computed as the length of the minimum path between them. Even if the minimum path is not, its length is unique by definition. If such a path does not exist, the distance is infinite.

Definition 2.16. The distance matrix $D_{ij} = \{d_{ij}\}$ of a given network is a matrix with the distances of every pair of nodes as entries.

The matrix D has only zeros on the diagonal. If all the entries of D are finite, the network is connected. For a connected network is useful to define the average path length

Definition 2.17. Given a connected network with n nodes and distance matrix D, the average path length L is the average of every nodes pair distances:

$$L = \frac{\sum_{i=1}^{n} \sum_{j \neq i} d_{ij}}{n(n-1)} \tag{2.1}$$

Given n nodes, the number of pairs is $n(n-1)$ if both n_i, n_j and n_j, n_i are counted $\forall\, i \neq j$. Thus, in a directed network, the maximum number of arcs is $n(n-1)$, while in an undirected one is $\frac{1}{2}n(n-1)$. The denominator in (2.1) is exactly the number of nodes pairs in the network. If the network is undirected $d_{ij} = d_{ji}\ \forall\, i, j$, so the distance is counted two times, but the two pairs (n_i, n_j and n_j, n_i) are both counted in the denominator. So, the formula (2.1) is valid for both directed and undirected networks.

The formula (2.1) is valid even for weighted networks, if the distance is generalized as the minimum sum of the weights (normalized over the maximum weight in the network) along the minimum path (in an unweighted network the weights are unitary).

Clustering

A node clustering coefficient measures how much the node is well connected with its neighbors.

Definition 2.18. The clustering coefficient C_i of the node n_i is the fraction of closed triangles over all the possible triangles, where the possible triangles are all the triplets of nodes containing n_i, and the closed ones are those in which all the three nodes are connected with the other two.

In an undirected network with n nodes, if the node n_i has degree k_i, the number of possible triangles containing n_i is $\binom{k_i}{2} = \frac{k_i!}{2(k_i-2)!} = \frac{1}{2}k_i(k_i-1)$. Given the adjacency matrix A, for the same graph, to count the number of closed triangles containing n_i is enough to compute $\frac{1}{2}\sum_{j\neq i}\sum_{l\neq(i,j)} a_{ij}a_{il}a_{jl}$: every addend is 1 only if all the three arcs exist and the sum is equal to $(A^3)_{ii}$, i.e. the i^{th} element of the diagonal of A^3. Thus, for an undirected, unweighted network:

$$C_i = \frac{(A^3)_{ii}}{k_i(k_i-1)}, \tag{2.2}$$

where $A^3 = AAA$. By definition, $C_i \in [0,\, 1] \;\; \forall\, i$.

The clustering definition changes in the case of a weighted network. Given w_{ij} the weight of the arc from n_i to n_j and $\widetilde{w}_{ij} = \frac{w_{ij}}{\max\limits_{i,j}(w_{ij})}$, instead of the number of triangles, it is better to consider the sum:

$$\frac{1}{2}\sum_{j\neq i}\sum_{l\neq(i,j)} \widetilde{w}_{ij}^{1/3}\widetilde{w}_{il}^{1/3}\widetilde{w}_{jl}^{1/3} = \frac{1}{2}\left(A^{1/3}\right)^3_{ii}, \tag{2.3}$$

where $A^{1/3} = \left(a_{ij}^{1/3}\right)$. Thus, for the node n_i:

$$C_i = \frac{\left(A^{1/3}\right)^3_{ii}}{k_i(k_i-1)} \tag{2.4}$$

Formula (2.4) holds for unweighted networks, too.

In directed network the situation is more complicated, in fact for every triplet of nodes, there are eight possible triangles. Chosen a node n_i, the number of bidirectional edges is:

$$k_i^{\leftrightarrow} = \sum_{i \neq j} a_{ij} a_{ji} = \left(A^2\right)_{ii} \tag{2.5}$$

The clustering coefficient of the node n_i is, like before, the number of triangles including n_i over the number of possible triangles:

$$C_i = \frac{\frac{1}{2} \sum_{j \neq i} \sum_{l \neq (i,j)} \left(a_{ij} + a_{ji}\right) \left(a_{il} + a_{li}\right) \left(a_{jl} + a_{lj}\right)}{k_i^{tot}(k_i^{tot} - 1) - 2k_i^{\leftrightarrow}} = \frac{\left(A + A^T\right)^3_{ii}}{2\left[k_i^{tot}(k_i^{tot} - 1) - 2k_i^{\leftrightarrow}\right]}, \tag{2.6}$$

where $2k_i^{\leftrightarrow}$ is the number of false triangles counted in $k_i^{tot}(k_i^{tot} - 1)$: only real triangles are needed.

Formula (2.6) holds for undirected networks because: $A = A^T$, $k_i^{tot} = 2k_i$, and $k_i^{\leftrightarrow} = k_i$.

Taking into account formulas (2.4) and (2.6), for directed and weighted networks:

$$C_i = \frac{\left[A^{1/3} + \left(A^T\right)^{1/3}\right]^3_{ii}}{2\left[k_i^{tot}(k_i^{tot} - 1) - 2k_i^{\leftrightarrow}\right]} \tag{2.7}$$

which is the same as the 2.6 when A is binary.

In the case of directed networks, the eight possible triangles for each triplet of nodes including the node n_i, can be classified into four categories:

- cycle, where every node has both an ingoing and an outgoing arc;
- middleman, where a node different from n_i has only ingoing or only outgoing arcs;
- in, where n_i has only ingoing arcs;
- out, where n_i has only outgoing arcs.

With this categorization, it is possible to define a clustering coefficient for every class of triangles:

$$C_i^{cyc} = \frac{(A^3)_{ii}}{k_i^{in}k_i^{out} - k_i^{\leftrightarrow}} \tag{2.8}$$

$$C_i^{mid} = \frac{\left(AA^TA\right)_{ii}}{k_i^{in}k_i^{out} - k_i^{\leftrightarrow}} \tag{2.9}$$

$$C_i^{in} = \frac{\left(A^TA^2\right)_{ii}}{k_i^{in}\left(k_i^{in} - 1\right)} \tag{2.10}$$

$$C_i^{out} = \frac{\left(A^2A^T\right)_{ii}}{k_i^{out}\left(k_i^{out} - 1\right)}. \tag{2.11}$$

These formulas hold for both directed and undirected networks. Their sum gives the clustering coefficient of the node n_i.

For the clustering coefficient is possible to define a global measure, too: in a network with n nodes the average clustering coefficient C of the network is the average of the clustering coefficients of the nodes

$$C = \frac{\displaystyle\sum_{i=1}^{n} C_i}{n}. \tag{2.12}$$

Density

Definition 2.19. Given an undirected network with n nodes and adjacency matrix A, its density is the fraction of existing arcs over the maximum number of arcs:

$$\delta = 2\frac{\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}}{n(n-1)}. \tag{2.13}$$

If the network is directed:

$$\delta = \frac{\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}}{n(n-1)}. \tag{2.14}$$

In the undirected case, the number of possible arcs is the half of the directed case: this is the reason why the 2 factor vanishes in the directed case.

Efficiency

The efficiency is a measure which comes in two versions: global and local.

Global efficiency tells how good is the information flow is in a network. Usually the assumption is that the communication among nodes is inversely proportional to their distance.

Definition 2.20. Given a network G with n nodes, and distance matrix $D = \{d_{ij}\}$, its efficiency $E(G)$ is defined as:

$$E(G) = \frac{\sum_{i=1}^{n}\sum_{i=1}^{n}\frac{1}{d_{ij}}}{n(n-1)} \tag{2.15}$$

Obviously, like said before, if a path from a given node n_i to another given node n_j does not exist, then $d_{ij} = +\infty$. Consequently, in this case, it is safe to assume $\frac{1}{d_{ij}} = 0$.

By definition, $0 \leq E(G) \leq 1$ and $E(G) = 1$ if and only if every pair of nodes is connected.

In the case of weighted networks, a normalization of formula (2.15) is enough: the obtained efficiency must be divided by the efficiency of the complete graph with the same nodes of G.

Definition 2.21. Be G^{ideal} the graph with the same nodes of G, but with all the possible arcs, the global efficiency (normalized and generalized) is given by:

$$E_{glob}(G) = \frac{E(G)}{E(G^{ideal})}. \tag{2.16}$$

Chosen a node n_i among the n nodes in the network G, its neighborhood has k_i+1 (or $k_i^{in}+k_i^{out}-k_i^{\leftrightarrow}+1 = k_i^{tot}-k_i^{\leftrightarrow}+1$ for directed networks) nodes. Be G_i^{ideal} the neighborhood of n_i completed with all the $\frac{k_i(k_i+1)}{2}$ (or $(k_i^{tot}-k_i^{\leftrightarrow}+1)(k_i^{tot}-k_i^{\leftrightarrow})$

for directed networks) arcs, the local efficiency of the network G is:

$$E_{loc}(G) = \frac{1}{n}\sum_{i=1}^{n}\frac{E(G_i)}{E(G_i^{ideal})}. \tag{2.17}$$

Vulnerability

Knowing the definition of efficiency given by formula (2.16), it is straightforward to define another measure: the vulnerability.

Definition 2.22. Chosen a node n_i in the network G, the vulnerability relative to node n_i measures the damage of the elimination of the node n_i and is computed according to the formula:

$$V_i = \frac{E_{glob}(G) - E_{glob}^{i}(G)}{E_{glob}(G)}, \tag{2.18}$$

where $E_{glob}^{i}(G)$ is the global efficiency of the graph obtained from the elimination of n_i in G.

The global version of this measure, i.e. the network vulnerability is:

$$V = \max_{i=1,2,\dots,n}(V_i). \tag{2.19}$$

Centrality

In network science, several measures of centrality exist, and each determines how much a node is central in the network. The first and simplest idea of centrality, proposed by Freeman (1978) was the degree centrality which classify the nodes according their degree: higher the degree of a node, higher the number of its neighbors, higher its importance in the communication in the network. It is necessary to be normalized to be used in different network with the same efficacy: it is normalized over the maximum possible degree.

Definition 2.23. Given a network with n nodes, be k_i the degree of the node n_i, the degree centrality of the node n_i is defined as:

$$C_i^{DEG} = \frac{k_i}{n-1}. \tag{2.20}$$

For directed networks, it is enough to substitute k_i with k_i^{tot}.

The second centrality measure proposed by Freeman (1978), is the betweenness centrality. A node is considered more or less central according to the number of shortest paths passing through it.

Definition 2.24. Be l_{jk} the number of shortest paths between nodes n_j and n_k in an undirected network with n nodes. Be $l_{jk}(i)$ the number of shortest paths between n_j and n_k passing through n_i. The betweenness centrality of the node n_i is:

$$C_i^{BET} = \frac{\sum\limits_{\substack{j=1\\ j\neq i}}^{n} \sum\limits_{\substack{k=1\\ k\neq i,j}}^{n} \frac{l_{jk}(i)}{j_{jk}}}{\frac{n^2-3n+2}{2}}, \tag{2.21}$$

The denominator of (2.21) is equal to $\left(\binom{n}{2} - 2(n-1)\right)$, that is the maximum number of shortest path for a node used as a normalizing factor. In a directed network this number is doubled, so in the directed case:

$$C_i^{BET} = \frac{\sum\limits_{\substack{j=1\\ j\neq i}}^{n} \sum\limits_{\substack{k=1\\ k\neq i,j}}^{n} \frac{l_{jk}(i)}{j_{jk}}}{n^2-3n+2}. \tag{2.22}$$

The third centrality measure proposed by Freeman (1978) is the closeness centrality which measure how much a node is close to the others: a node is more central if it is less distant from the others, that is if the sum of the distances from the other nodes is low.

Definition 2.25. Given a network with n nodes and distance matrix $D = \{d_{ij}\}$, the closeness centrality of the node n_i is:

$$C_i^{CLO} = \frac{n-1}{\sum\limits_{j=1}^{n} d_{ij}}, \tag{2.23}$$

where $n-1$ is the minimum sum of the distances from the other nodes of n_i.

This definition is like the inverse of the ratio which indicates how much the sum of the distances is higher than the minimum possible. This definition holds only if the network is connected.

Many other centrality measures exists, like the eigenvector centrality (Bonacich, 1991; Bonacich & Lloyd, 2001), the Katz centrality (Katz, 1953), the Bonacich centrality (Bonacich, 1987), and the Hubbel centrality (Hubbell, 1965), but these were not used in the rest of this thesis. The papers in which these measure were presented are in the bibliography, if needed.

Hub and Authority

Hub and authority are two intertwined concepts: it is not possible to define one without the other. Hub and authority definition overlap in the case of undirected networks. Both hubs and authorities are important nodes in a network, when the communication among nodes matters: a hub is a node to which many nodes are connected and a authority is a node connected to many other nodes. But this is not enough, in fact this would be just a directed degree centrality. Thus, an informal definition is: to a good hub are connected many good authorities and a good authority is connected to many good hubs.

Kleinberg (1999), studying the World Wide Web network, proposed the HITS (Hyperlink Induced Topic Search) algorithm to identify hubs and authorities. If the network has n nodes, two vectors $\vec{x} = (x_1, x_2, \ldots, x_n)$ and $\vec{y} = (y_1, y_2, \ldots, y_n)$, with $\sum_{i=1}^{n} x_i^2 = 1$ and $\sum_{i=1}^{n} y_i^2 = 1$ are needed. x_i is the initial non negative score as an authority for the node n_i, while y_i is the non negative score as a hub. Be

N_i^{in} the set of indexes of the nodes with an outgoing arc towards n_i and N_i^{out} the set of indexes of nodes with an ingoing arc from n_i. Two operations are defined:

$$\mathcal{I} : \sum_{j \in N_i^{in}} y_i \mapsto x_i \tag{2.24}$$

$$\mathcal{O} : \sum_{j \in N_i^{out}} x_i \mapsto y_i. \tag{2.25}$$

At every step of the algorithm, $\vec{x}$ is updated using (2.24) and $\vec{y}$ using (2.25). Then, the two vectors are normalized to obtain $\sum_{i=1}^{n} x_i^2 = 1$ and $\sum_{i=1}^{n} y_i = 1$. At some point, a fixed point, independent from the initial vectors, is reached and the final $\vec{x}$ and $\vec{y}$ are computed. Higher the i^{th} component of $\vec{x}$ (or $\vec{y}$), better n_i is as authority (or hub). Thus, it is possible that a good hub is a good authority, too.

This algorithm shows a good convergence speed, but there is an algebraic alternative ot obtain the authority and hub scores. A possible initial choice, before the normalization, for $\vec{x}$ and $\vec{y}$, is $\vec{z}$ with components all equal to 1. If A is the adjacency matrix of the network, the two operations (2.24) and (2.25) can be rewritten as:

$$\mathcal{I} : A^T\vec{y} \mapsto \vec{x} \qquad \mathcal{O} : A\vec{x} \mapsto \vec{y}.$$

After k iterations, $\vec{x} = \left(A^T A\right)^{k-1} A^T \vec{z}$ and $\vec{y} = \left(AA^T\right)^k \vec{z}$. $\vec{z}$ is not orthogonal to the principal eigenvector of the symmetric matrix AA^T, then it converges, as k increases, to the principal eigenvector of AA^T, which is non negative. So the vector $\vec{y}$ converges to the principal eigenvector of AA^T. The same holds for $\vec{x}$ which converges to the principal eigenvector of $A^T A$.

Independently from the algorithm used to compute hub and authority scores for the nodes of a network, a problem still remains: there are no fixed thresholds to tell if a node is really a hub or an authority. This threshold needs to be careful selected case by case, taking into consideration the number of nodes and arcs.

For undirected networks, a different method is proposed by van den Heuvel, Mandl, Stam, Kahn, and Pol (2010). The method, tested in a study on schizophrenia, is based on characteristics a good hub should have:

- high degree (with direct connections, information flow control is easier);
- high betweenness centrality (under the assumption that information prefer to travel on shortest paths);
- low average distance (to be easily reached from other nodes);
- low clustering coefficient (if two nodes in a triplet are not connected, they have to communicate necessarily through the other node).

After the computation of the four measures a point is assigned to the 20% of the nodes with:

- the higher degree,
- the higher betweenness centrality,
- the lower average shortest distance,
- the lower clustering coefficient.

Every node with at least one point is considered a hub. This method seems valid for brain networks, but a different threshold could be needed for different cases.

Page Rank

Page Rank (Page, Brin, Motwani, & Winograd, 1998) was introduced to measure the importance of web pages giving a rank to every of them. This measure take into consideration that the World Wide Web is a directed network, with the pages as nodes and the hyperlinks as arcs. Let n_i be a node, with O_i the set of nodes with an outgoing arc pointing to n_i and I_i the set of nodes with an incoming arc from n_i. Let $c < 1$ be a normalization factor and $N_i = |O_i|$. A rank for every page, which depends on the rank of the other pages, is a simplified definition of Page Rank:

$$R(n_i) = x \sum_{n_j \in I_i} \frac{R(n_j)}{N_i} \tag{2.26}$$

In formula 2.26, this concept is respected: a node has a high rank if many high ranked nodes point to it. Thus, $c < 1$, because there are nodes without outgoing arcs. Equation 2.26 is recursive, but convergent and its computation may start from a ranking where every node has the same rank. If there are two nodes which point towards each other and not to any other nodes, but with one node pointing to one of them, during the computation, this loop will become a rank sink which accumulate rank without distributing any of it. To solve this problem, a rank source is introduced:

Definition 2.26. Let S be a rank source, i. e. a vector of ranks for every node. The Page Rank of the node n_i is defined as

$$PR(n_i) = c \sum_{n_j \in I_i} \frac{PR(n_j)}{N_i} + cS \tag{2.27}$$

with c maximized and $||PR||_1 = 1$.

An alternative formula, used for the models in this thesis, is:

$$PR(n_i) = \frac{1-d}{N} + d \sum_{n_j \in I_i} \frac{PR(n_j)}{k_j^{out}} \tag{2.28}$$

where d is a damping factor set as 0.85 and N is the number of nodes in the network. The damping factor d, in the original idea, represents the probability that a user will continue to click on a hyperlink while browsing the web. So $1 - d = 0.25$ is the probability that the user stops browsing.

2.5 Communities

It is often important to identify different groups of nodes in a network. Usually these groups are called communities. A community is a subset of nodes which are closer together in the network. Unfortunately, a definition of community does not exist, but it is clear that the nodes in a community must have many connections among them and few with the other nodes of the network.

A good idea to try to identify a partition of the network (a subgraph division of the network), dividing it in communities, is to define a null model: a random graph having the same structural properties of the original one. After establishing a null model, a comparison with the original one is performed to highlight the network's community structure.

First of all, the definition of random graph is needed. Its definition was given by Erdös and Rényi (1959) as the first try to represent real networks. Be $N = \{n_0, n_1, \ldots, n_{n-1}\}$ as a set of nodes, and $\mathcal{G}_N$ as the set of all graphs over N, we need to transform $\mathcal{G}_N$ in a probability space. Assume every possible arc to have a probability $p \in [0,\ 1]$ of being an arc of a selected graph $G \in \mathcal{G}_N$ and a probability $q = 1 - p$ to not belong to G. Fixed $G_0 \in \mathcal{G}_N$ with m arcs, the probability of the event $\{G_0\}$ is $p^m q^{\binom{n}{2}-m}$: this is the probability to generate G_0 among all the possible graphs in $\mathcal{G}_N$. For every possible arc l, a probability space $\Omega_l = \{0_l, 1_l\}$, where $P_l(1_l) = p$ and $P_l(0_l) = q$, is defined. The desired probability space is, then, $\Omega = \prod \Omega_l$. Every element of Ω is a map ω which maps every arc l to the values 0_l or 1_l. The probability measure P over Ω is the product of the probability measures P_l.

Definition 2.27. ω, as constructed above, is a random graph G over N with connection probability p and arcs set $L = \{l | \omega(l) = 1_l\}$.

Random graph model is based on two parameters: the number of nodes n and the connection probability p: every pair of nodes has the same probability p to be connected, independently from the other pairs. The expected number of arcs is $\frac{pn(n-1)}{2}$ and the average degree is $p(n-1)$.

The independent probability of every pair to be connected, make the random graph a good null model: a community structure is not expected.

Many algorithm to find graph partitions exists, but it is necessary to establish if the results are reliable. To do so, a quality function is needed. The most famous one is the modularity (Newman & Girvan, 2004), based on the comparison between the network density and the null model density.

Definition 2.28. Given a network G with n nodes, m arcs, and adjacency matrix A, modularity is:

$$Q = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(a_{ij} - P_{ij} \right) \delta \left(\mathcal{C}_i, \mathcal{C}_j \right), \tag{2.29}$$

where P_{ij} is the expected number of arcs between n_i and n_j in the null model, $\mathcal{C}_i$ and $\mathcal{C}_j$ are the community of n_i and n_j, respectively, and $\delta \left(\mathcal{C}_i, \mathcal{C}_j \right) = 1$ if $\mathcal{C}_i = \mathcal{C}_j$, 0 otherwise.

Using $P_{ij} = \frac{2m}{n(n-1)}$ $\forall i, j$, i.e. choose a Erdös-Rény random graph as a null model, is the easiest choice, but unfortunately almost every real network does not have this kind of distribution. A better null model, because it reflects more the structure of the network, is the configuration model (Van Der Hofstad, 2009). The probability of a node n_i with degree k_i, to be connected with the node n_j with degree k_j is $p_i p_j = \frac{k_i k_j}{4m^2}$, where $p_i = \frac{k_i}{2m}$ is the probability of having an arc starting from n_i. So the expected result is $P_{ij} = 2m p_i p_j = \frac{k_i k_j}{2m}$. Using this null model in the definition (2.29):

$$Q = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(a_{ij} - \frac{k_i k_j}{2m} \right) \delta \left(\mathcal{C}_i, \mathcal{C}_j \right). \tag{2.30}$$

Defining n_c as the number of communities, l_i as the number of arcs among the nodes of the i^{th} community, and d_i as the sum of the degrees of the nodes of the i^{th} community, and noticing that the only non zero addends are the ones regarding the pairs of nodes in the same communities, the formula (2.30) can be rewritten as:

$$Q = \sum_{i=1}^{n_c} \left[\frac{l_i}{m} - \left(\frac{d_i}{2m} \right)^2 \right]. \tag{2.31}$$

The first ratio is the fraction of arcs which are in the community, while the second one represents the expected fraction in a random graph with the same degree distribution. According (2.31), a subgraph is a community if and only if its addend in the formula is positive.

To find a good community partition, i.e. a partition with a modularity value as high as possible, there are many algorithms. The one which represents the best compromise among execution time, memory usage, and good results is the Louvain method (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008), which is a greedy algorithm divided in two parts.

The initial step of the first part is considering every node in the network as a community. Then, every node's neighborhood is considered: every neighbor of n_i is tested as a new member of C_i and the consequent modularity variation ΔQ is recorded. If such a move, with the maximum positive ΔQ exists, the move is done. When the modularity cannot be increased anymore, the first part of the algorithm is halted.

The second part starts building a network in which the nodes are the found communities linked together only if at least an arc between the two communities exists. The arcs are weighted with the sum of the weights of the arcs between the communities. Loops are introduced, weighted with the sum of the weights of the arcs in the community.

The first part is repeated on the new network, with the modularity computed on the original network, and so on, until no changes are made to the network.

The best part of this algorithm is the easy and fast computation of the variation of the modularity when the node n_i is moved into a community:

$$\Delta Q = \left[\frac{\Sigma_{in} + \sigma_{i,in}}{2m} - \left(\frac{\Sigma_{tot} + \sigma_i}{2m}\right)^2\right] - \left[\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m}\right)^2 - \left(\frac{\sigma_i}{2m}\right)^2\right], \qquad (2.32)$$

where Σ_{in} is the sum of the weights inside the community, Σ_{tot} is the sum of the total strengths of the nodes of the community, σ_i is the total strength of the node n_i, $\sigma_{i,in}$ is the sum of the weights of the arcs between n_i and the nodes of the community, m is the sum of the weights of all the arcs in the network.

2.6 Small World and Scale Free

In many studies, real networks shown two peculiarities: often real world networks are both scale free and small world.

Small World

A random network is a network in which every node is connected with another one according to a fixed probability. The opposite of the random networks are the regular ones: every node has the same degree and is topologically indistinguishable from the others. While random graphs have a low clustering coefficient and a low average path length, regular graphs have high value for both measures.

Real networks are different from both these kinds of graphs: they have no completely chaotic structure (like random ones), and they do not show a perfectly regular and ordered topology (like regular graphs). Real networks are, somehow, in the middle of these two models. Small world networks have a low average path length (like random graphs) and a high clustering coefficient (like regular graphs): these values favor both the network aggregation and its exploration.

In 1960s, the sociologist Milgram proved empirically that every person get to know every other person, with an average of five intermediaries: this was the experiment about the six degrees of separation. A group of people in the Midwestern United States was randomly selected to send a package to a stranger inhabitant of the Massachusetts. The addresses were unknown to the senders: only the names were known. Thus, every sender had to send the package to someone they knew and they believed to be an acquaintance of the target. Every time a package was received, the rules were the same. Some scientists thought about a hundred of passages were necessary, but an average of six were enough: this really is a small world.

Watts and Strogatz Watts and Strogatz (1998) in 1998 focused on the small world concept and formalized it. They started from a regular graph and applied a rewiring procedure to it: every arc was rewired (moved from a pair of nodes to another one randomly selected) with a fixed probability p. If $p = 0$ the network do not change after the procedure, while with $p = 1$ a random graph is obtained. When $0 < p < 1$ there is the possibility to obtain a small world network.

Two additional parameters are needed to fully formalize the small world characteristics:

$$\gamma = \frac{C^{real}}{C^{rand}}, \tag{2.33}$$

$$\lambda = \frac{L^{real}}{L^{rand}}, \tag{2.34}$$

where C^{real} and L^{real} are the clustering coefficient and the average path length of the network, respectively, while C^{rand} and L^{rand} are the averages of same measures for a population of rewired networks.

Definition 2.29. A network is said to be small world if:

$$\begin{cases} \gamma \gg 1 \\ \lambda \approx 1 \end{cases} \tag{2.35}$$

Sometimes the parameter $\sigma = \frac{\gamma}{\lambda}$ is used: if $\sigma > 1$ the network is small world. This classification can results in false positives when $\lambda < 1$.

A year later, Watts Watts (1999) specified that a network, to be considered small world needs to:

- have a number of nodes $n \gg 1$, because it is normal for the people of a small town to know every each other, while a world is needed;
- be sparse, meaning that $\bar{k} \ll n$, because not every people know each other;
- be decentralized, i.e. there are no hubs and $k_{max} \ll n$, otherwise a person who knows almost everyone else unrealistically shorten the paths;
- be very clustered, with many nodes sharing some of their neighbors.

Scale Free

The degree distribution is a key characteristic of every network: it represents how many nodes have a high or low degree. It is the probability distribution of the number of nodes with degree $k \in \mathbb{N} \cap [k_{min}, k_{max}]$. If the network has n nodes and $\eta_k \leq$ is the number of nodes with degree k, the probability distribution is:

$$P(k) = \frac{\eta_k}{n} \tag{2.36}$$

Usually three degree distribution are taken into consideration (Amaral, Scala, Barthelemy, & Stanley, 2000): scale free, broad scale and single scale.

Scale free networks have a degree distribution following a power law:

$$P(k) = ck^{-\alpha}, \tag{2.37}$$

which is a straight line when plotted in log-log scale.

(Barabási & Albert, 1999) proposed a growth model for networks which brings to the creation of a scale free network: the preferential attachment. Starting with a single node, at every time step a node is added and is linked with an arc to an existing node proportionally to its degree. The probability of the new node to be connected to the node n_i is:

$$\Pi(i) = \frac{k_i}{\sum_j k_j}. \tag{2.38}$$

In this model, riches get richer: a node prefers to attach itself to a node with many connections and it is like these to be the older. As example of real world networks following this rule, there are the social networks or the World Wide Web network: someone new in a group of people try to become acquaintance with the most popular people to have the possibility to know many people; likewise, a new web page prefer to share links with a visited page to become much popular.

It is proved that a network resulting from a long enough simulation of this growth model, is a scale free network with its degree distribution following a power law. The name scale free means that its law has the same form at every scale: $f(\alpha x) = \beta f(x)$, with $\alpha, \beta \in \mathbb{R}$.

Broad scale networks are usually referred to as scale free network, because the tail of their degree distribution follow a power law. The whole degree distribution, instead, follow a truncated power law, i.e. a power law with an exponential cut:

$$P(k) = ae^{-\frac{k}{k_c}}k^{-\alpha}, \tag{2.39}$$

with $a, \alpha, k_c \in \mathbb{R}$.

Single scale networks have a degree distribution with a very fast decay, like an exponential law:

$$P(k) = ae^{-\frac{k}{k_c}}. \tag{2.40}$$

Chapter 3

Script Based Processes

3.1 Introduction

One of the main problems about semantic networks is the study of words associations chains to create an associative hierarchy.

When a connection between words is activated, the starting node is recorded in memory. While the activation process continues, every time a word referring to a starting node is reached, an intersection between the two nodes is formed. Then a validation of the path which brought to this intersection is needed to check if the syntax and context rules are respected.

The dynamic process which build semantic networks which brings chains of words associations, from the cognitive viewpoint, is the focus of this study.

Collins and Loftus (1975) considered four basic assumptions in the semantic network building creation process:

1. A concept activation spreads as it follow a path between nodes. The activation should be treated as a signal from a fixed source which dissipates as it flows;

2. When a concept is activated repeatedly, the activation starts from the beginning of the path with the same pattern. Only one concept at a time can be activated and only one connection at a time can be activated;

3. The activation is a decreasing signal;
4. Every term activation and intersection brings to the path taken an activation level and the path is considered as a whole only when the total amount of activation level reaches a threshold.

Thus, this creation process is based on semantic similarity: if two words are semantically similar, their connection is almost certain and their concepts are linked. So, semantic connections derive from concepts ones.

What still is missing in the literature is a systematic study of these semantic network's organization. To fill this hole, in this thesis, complex networks have been used as a base model.

Complex networks can be used as a theoretical and as a methodological tool (Neal, 2012) and are useful to model linguistic structures (Solé, Corominas-Murtra, Valverde, & Steels, 2005). Thus, complex networks are widely employed in several linguistics sectors: word sense disambiguation (Mittal & Jain, 2015; Stevenson & Wilks, 2003), text summarization (Polepalli Ramesh, Sethi, & Yu, 2015), Natural Language Processing (Jurafsky & Martin, 2000), sentiment analysis (Pang & Lee, 2004), and as statistical physics methods (Krylov, 2014), too.

Linguistic networks can be classified into two main categories: semantic and superficial (Costa et al., 2011). The former category highlight semantic relations between terms (for example, synonymy and antonymy in dictionaries). The latter includes words structure and position.

As said before, words associations are good to create semantic networks: in fact, with this method a scale free and small world network can be obtained, and usually linguistic networks shows these behaviors. For example, many famous networks, such as synonyms networks (de Jesus Holanda, Pisa, Kinouchi, Martinez, & Ruiz, 2004; Makaruk & Owczarek, 2008; Motter, de Moura, Lai, & Dasgupta, 2002; Steyvers & Tenenbaum, 2005; Strori, Bombaci, & Bingol, 2007), English Wordnet (Sigman & Cecchi, 2002; Steyvers & Tenenbaum, 2005) and the word-association networks (Da Fontoura Costa, 2004; Ferreira, Corso, Piuvezam,

& Alves, 2006; Steyvers & Tenenbaum, 2005), are scale free and small world networks. Steyvers and Tenenbaum (2005) modeled lexicon development using a preferential attachment algorithm Barabási and Albert (1999) which brings both a scale free and small world network. It is obvious that a good model needs to have the same properties of real world data networks (Gravino, Servedio, Barrat, & Loreto, 2012; Nelson, McEvoy, & Schreiber, 2004; Steyvers & Tenenbaum, 2005).

These approaches still lack the comprehension of the intermediate level in these networks, that is how every level influence the others: the topology may affect the thinking process and vice versa(Barabási, 2011; Jasny, Zahn, Marshall, & Cho, 2009; Watts, 2004). Recent studies about language data produced by human subjects (Friederici & Gierhan, 2013) used network science for the data analysis, but there are no experiments about the topology analysis of knowledge based scripts and words associations.

Two experiments were performed in order to understand word production processes based on scripts and words associations and establishing the semantic structure of a script as a base. The data produced by the subjects were transformed into networks to study the relationship between lexicon organization in memory and the production processes.

3.2 Experiment 1

Experiment 1, divided in two stages, was performed to gather data from students to create the base for a memory system. In the first stage, 161 students (139 females and 22 males; mean age = 21.46 years, range = 21÷30 years) generated scripts as partial fulfillment of an Introductory course of Psychology, at the University of Calabria. Their education age was 14,47 mean years of schooling. Students achieved high scores on the 30-items vocabulary subtest from the WAIS (Wechsler, 2008) (mean = 36.50).

The subjects had to generate 4 scripts based on 4 different specified situations among 39 (reported in table 6.1) with a minimum of 20 words per script. The instructions given to the subjects were adapted from those used by Bower et al. (1979). The words had to be produced as if the subjects were telling a story. No time limit was given and the subject completed their task within a time range of 20÷60 minutes. The words could belong to the traditional grammatical categories. The actions and nouns produced in the first stage, were selected to be given a typicality score. Inconsistent entries were discarded. The number of actions to be rated ranged from 20 to 30 for each script. Each subject rated 3 among the scripts produced by the other subjects.

Following the method used by Smith and Graesser (1981), the typicality score score ranged from 1 to 6: 1 if the action was definitely not pertinent to the script; 2 if the subject was fairly sure that the action was not pertinent to the script, 3 if the subject was uncertain, but thought that the action was not pertinent to the script; 4 if the subject was uncertain, but thought that the action was pertinent to the script; 5 if the subject was fairly sure that the action was pertinent to the script; 6 if the action was definitely pertinent to the script. Given the total number of nouns, those with typicality score greater than 4.4 were considered typical (with a mean of 5.31) while the other were considered atypical (with a mean of 2.13).

Results

Every incorrect word (typos or non existent words) was discarded. Then, for each script, the number of words belonging to each different grammatical categories was computed (see table 6.1 for details), together with the mean value of the typicality score (see table 6.2 for details), divided per grammatical category.

Each script was different both in the details with which the subjects told the story about the different situations and in the kind of story they told to build the scripts.

An more accurate analysis was made on the words produced. The proportion

	AJ	AD	N	V	TOT
AJ	1	.38*	.93**	.87**	.78**
AD	.38*	1	.52**	.53**	.37*
N	.93**	.52**	1	.87**	.83**
V	.87**	.53**	.87**	1	.86**
TOT	.78**	.37*	.83**	.86**	1

Table 3.1: * Correlation was significant at the .05 level, ** Correlation was significant at the .01 level. Table legend: AJ stands for adjective, AD stands for adverb, N stands for noun, V stands for verb and TOT is the statistic considering all the terms. The table has been realized considering the Pearson product-moment correlation coefficient between the typicality score of the different categories of terms written by the students during the experiments.

of nouns over verbs generated were computed (see table 6.1 for details) and ranged from 0.37 to 1.95. For each word, script and subject, the mean typicality score was computed and for the whole group of subjects combined, too. The mean typicality score for each subject shown low variability in every script: correlations between the sets of typicality scores ranged from 0.37 to 0.93 as shown in table 3.1 and in figure 3.1. All the correlation coefficients were positive and were significant at the 0.01 level, except the ones between adjectives and adverbs and between the adjectives and all the terms which are significant at the 0.05 level.

This experiment gave a mental representation of every day situations and activities in the memory structure of the subjects. Indeed, some of the scripts were more accessible to the subjects than others: some situations are more familiar to the an average person respect to others. In fact, the positive correlations states that when a grammatical category received a high mean typicality score in a script, then the other categories had high scores, too. Even if the scale is discrete and assumes only 6 values, it was enough to show the higher familiarity of some scripts as it is easy to see from figure 3.1.

The number of terms produced greatly varied from script to script and from subject to subject. This number ranged from 30 to 210, with a mean value of 72 terms produced. The number of terms in each grammatical category gave an outline of how these categories are organized in the subjects' memory: the verbs category was the one with more words, while the adverbs category was the one

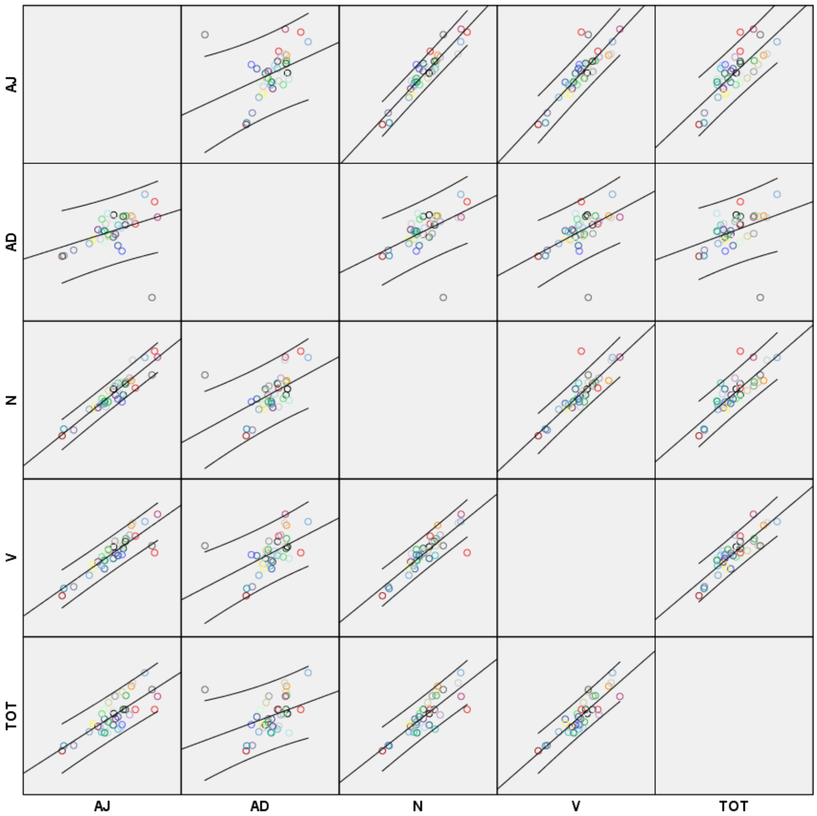


Figure 3.1: Correlations between different Grammatical Categories. Chart legend: AJ stands for adjective, AD stands for adverb, N stands for noun, V stands for verb and TOT is the statistic considering all the terms. This correlation matrix was obtained computing the Pearson product-moment correlation coefficient between the typicality scores of the different grammatical categories. Every circle, represented with a color, is one of the 39 different scripts, analyzed in this experiment.

with the lower number of words, without considering the conjunctions and the determiners categories, is the adverbs category. Thus, from this experiment, is evident that the subjects' lexicon, related to a script production task, is mainly composed of nouns and verbs. These two categories shows the lower mean typicality score, but this does not low their importance in the script: the low score is due to the very high number of elements belonging to these categories.

3.3 Experiment 2

The same subjects participating in the first experiment, took part in the second one. Experiment 2, divided in two stages, was performed to collect more data on subjects' semantic knowledge, by the introduction of superordinate and subordinate categories, as well as synonyms of each of the terms previously produced.

In the first stage, divided in two parts as well, the data driven retrieval method was used. This method is found in some memory recognition models with the aim to retrieve related items as a semantic *copy cue* (Rabinowitz, Mandler, & Patterson, 1977) and to allow the *detection of familiarity* (Atkinson & Juola, 1974) or "intraitem elaboration" (Mandler & Johnson, 1980) in the subjects' memory. In part A of the first stage, a booklet with the words produced in the first experiment was given to each subject. Every subject had to write synonyms of every word found in the booklet to expand the script's semantic space. Then, again, in part B of the first stage, for each synonym they produced, the subjects had to write other synonyms. In this second part of this stage, the use of an online dictionary was allowed, when no synonyms were found, to further expand the script space.

In the second stage, the conceptually driven recall took part: each of the word of the first experiment was embedded in a context as stimulus for replaced memory for scripts. Another booklet, with the words produced in the first experiment and in the first stage of this experiment, was given to each subject. Then, they were asked to write down hyponyms and hypernyms for each term, if possible.

Results

The detailed results are shown in tables 6.3 and 6.4. In figure 3.2 the synonyms production step is shown. The average number of new terms, per script was 213 in part A (with an increase of 336%) and 462 in part B (with an increase of 176%). The total number of new terms was 4825 in part A (increase of 273%) and 7910 in part B (increase of 120%). Thus, search for synonyms increase rapidly the number of terms in a mental lexicon, with the risk, although, of contaminating the semantic area of the script. The synonyms found helped to organize subjects' memory and gave a method to understand the contamination of the semantic field, by analyzing the linking variability among them: the synonyms could be very similar or very dissimilar from the starting situation.

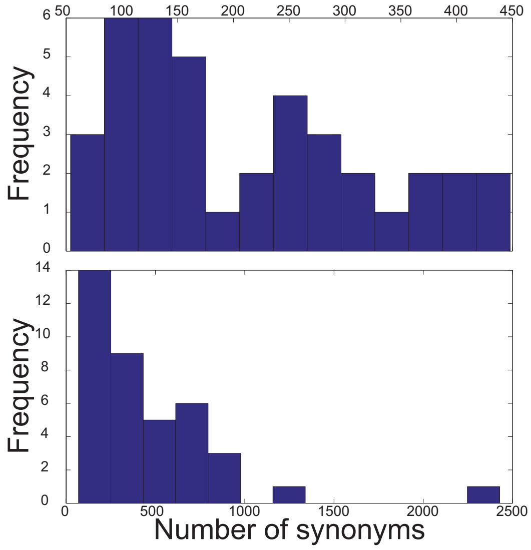


Figure 3.2: Experiment 2, part A (top) and part B (bottom) synonyms production. The histograms were realized counting the number of synonyms produced during the second experiment and counting how many scripts had that number of new terms. The binning was chosen to highlight the peaks and the outliers.

Discussion

The second experiment was about the search of higher level actions sequences in the scripts. With this experiment, the analysis of the scripts' boundaries was possible: it was

In the generation phase of the first experiment, normative data were gathered on 39 routine activities. This word production gave an outline of the subjects' memory system organization. To improve this picture, experiment 2 was designed to segment the higher-level action sequences in a script. If the subjects could change the words of the skeleton-like script, are there possible boundaries? At what level of the script's boundaries, picking up synonyms after synonyms for each word of the basic script, are there boundary locations? Starting from experiment 1 results, a first measure of the typicality of the words (Verb, Noun) was obtained, rated according to the 6-point recognition scale used by Smith and Graesser (1981).

In literature, memory representation for scripts consists of a *pointer* to a script together with a set of typical action, a set of not so typical actions and a set of atypical actions (Bower et al., 1979; Davidson & Hoe, 1993; Graesser et al., 1979; Graesser, Woll, Kowalski, & Smith, 1980; Smith & Graesser, 1981). In these works, the pointer links to a whole script which implicitly contains all the important information. Thus, discriminate between present and absent atypical action is easier than discriminate between present and absent typical ones(Bower et al., 1979; R. C. Schank & Abelson, 1977). Moreover, these works conclude that recall memory tasks perform better for atypical actions. The scripts, instead, usually contain typical actions, but not the atypical ones. This has been found false in preschool children (Hudson, 1988) who recognize atypical actions better, which suggests a different ways of memorizing script actions: in this way it is possible to recognize atypical actions as isolated ones when compared to an homogeneous background of typical ones (Bower et al., 1979).

Analyzing the data obtained in the experiment, some conclusions can be drawn. The spreading activation dynamics caused the subjects to follow the paths that made them remain in the boundaries of the script's semantic area, or to continuously choose the paths which brought them outside of the script to reach another one. The search for synonyms, corrupted the connections with the starting script and the search for hyponyms and hypernyms almost destroyed most of them, highlighting the borders among scripts. The different connection patterns involved in this experiment changed, in both a positive and negative way, the script organization, giving information on both the global and meso level of the lexicon and memory structure (Kintsch, 1998; Mannes & Kintsch, 1987; R. Schank, 1982; R. C. Schank & Abelson, 1977).

3.4 Network Analysis of Semantic Data

The sample collection of 17.810 words, obtained from previous two experiments, has been turned into networks. Starting with the script's initial situation as the core node, every produced word of the first experiment is connected to it. Then,

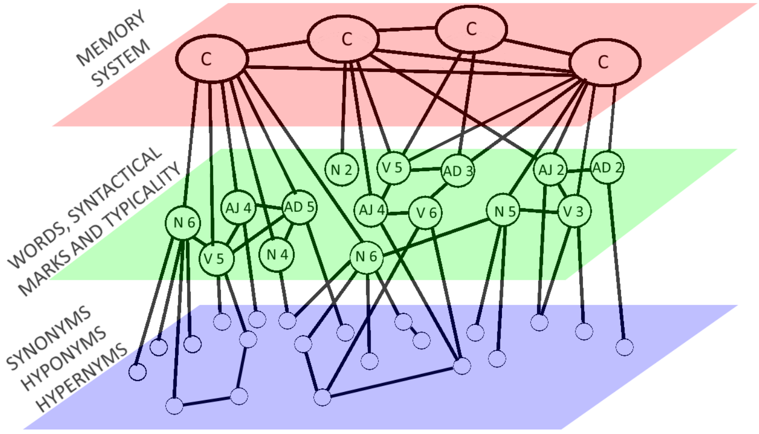


Figure 3.3: Representation of different levels of memory: on the highest level there is the semantic memory composed by cognitive structures; on the intermediate one the concepts space with the relative grammatical category and a typicality score; the lower plane consists of synonyms, hyponyms and hypernyms that widen the memory space.

for every other stage of the second experiment, every synonym is connected with the related term. At last the same thing was made for hyponyms and hypernyms. The resulting network contained no loops and was considered as undirected and unweighted. The network obtained was thought as a collection of words or concepts, and formed a complex system made up of different levels as shown in figure 3.3.

The highest level is constituted by semantic memory, represented by cognitive structures relating themselves to a part of memory for actions. These structures are composed by big clusters of concepts and, depending on the experience during one's life, they are interconnected and repeated. The middle level comprises single concepts together with their typicality scores for every situation. The last level is composed by synonyms, hyponyms and hypernyms. When information retrieval is needed, a search in this three level space is performed using the links, formed by experience, as a guide. Of course, depending on the starting point and the situation, the paths followed are different.

Networks' parameters, such as total number of nodes n, total number of edges m, average degree $\bar{k}$, clustering coefficient C, efficiency E, average path length L, network diameter d, density δ, small-world parameters $(\lambda, \gamma, \sigma)$, degree distribution $P(k)$ considering the small-world and scale-free models have been used to analyze the data obtained from the experiments. Other parameters such as K-Nearest-Neighbor K_{NN} and communities have been employed to understand the organization of the single words and groups of words in the parameter space of the students memory system. All these parameters have correspondent functional implications in the students memory systems.

The number of nodes of the network represents the number of different terms which compose the subject's memory. It can be considered for a single script, for a single subject or for the whole dataset, depending on which kind of information one is interested in.

The number of edges and the density, give an idea on the connectivity among the concepts. In fact, with more edges (and with a higher density), the network exploration is easier and it is easier to navigate through different scripts and situations as well.

The average degree, i.e. the average number of connection for each node, tells how far are the different areas in the memory: if $\bar{k}$ is low, areas which are very different, may be completely separated.

The clustering coefficient of the network measures how much the concepts are interconnected among themselves. So, even if there is a low number of concepts, with a high average clustering coefficient, it is possible to move easily among them. A low density coupled with a high clustering coefficient, is a good strategy to avoid unnecessary and too long paths (which can comprehend general and misleading terms) when searching the memory for information. Having a too low density and a too high clustering coefficient, by the way, is not always a good thing: hubs, which are usually general terms, are important because the play as a bridge between completely different semantic areas.

The average path length is also a good measure of separation and definition of the semantic areas. If it is too high, the areas are distant and well separated, so it is difficult to pass from one to another. On the other hand, if it is too low, the concepts may be too close and not well defined, so the concepts may result confused and overlapping with a possibility of information loss.

The diameter of the network represents how much two different areas can be dissimilar and distant.

The global efficiency allows to understand how much is easy to explore the memory, passing from one concept to another. The efficiency could be similar to the Quillian's spreading of information (Collins & Loftus, 1975).

If the memory network is a small world network, it contains some long range connections and strategic hubs to travel through different areas, but it contains even clusters of terms which define better every concept. Thus, having a small world behavior seems to be a good strategy to reduce the distance between different concepts without losing the definition of the concepts.

A scale free network has more low degree nodes than hubs. But usually, in a growing scale free network, new nodes, tend to link to the hubs. When a new term or a new concept is apprehended, generally it is connected to a preexisting more general term or concept which results better defined. So, it is safe to assume this growing behavior as a good one for a memory network.

K-Nearest-Neighbors measures how much the nodes are linked to nodes with the same degree. So, in a memory network, it tells how much the polysemic terms share meanings among them (hubs linked to hubs) and how much the terms of a semantic area tend to specialize (low degree nodes connected to low degree ones). After having computed the average degree of the neighbors for every node, averaging this quantities over the nodes with the same degree, is a good method to compute $K_{NN}(k)$.

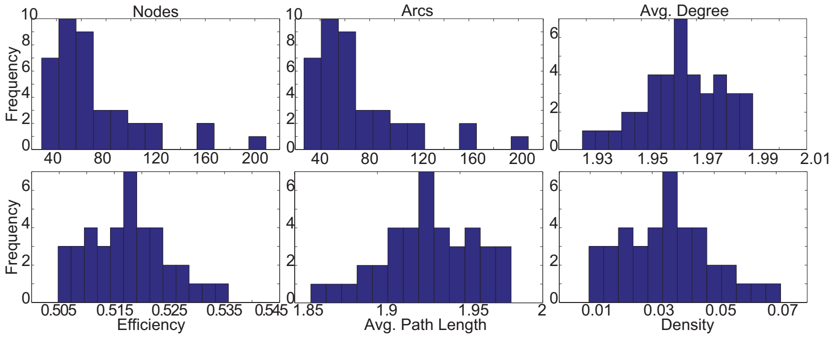


Figure 3.4: From left to right and from top to bottom, histograms of frequencies of: number of nodes, number of arcs, average degree, efficiency, average path length and density of networks for the production of scripts in experiment 1. The histograms have been created with an ad hoc developed Matlab code derived by using the package Brain Connectivity Toolbox (Rubinov & Sporns, 2010), (http://www.brain-connectivity-toolbox.net/).

3.5 Memory for Scripts

Experiment 1 Networks: Grammatical Categories Occurences

Script networks built from experiment 1 data shown a great variability in both number of edges and nodes depending on the subjects. The average degree, by construction, had values close to 2. The average clustering coefficient was obviously zero since in this stage, all nodes but one had degree equal to 1, and it was therefore impossible to form triangles. The efficiency however, was inversely proportional to the number of nodes n: in this stage, each starting node had distance 1 from the other $n-1$ nodes and the other $n-1$ nodes had distance 1 from the starting one and distance 2 from the remaining $n-2$ nodes. The average shortest path was equal to the average degree, while the diameter was obviously equal to 2. The density was inversely proportional to n. A schematic representation of the results is shown in figure 3.4.

Experiment 2 Networks, First Stage Part A: Synonyms

The networks built from data obtained in the first part of the first stage of the second experiment, had a great structural variability. In any network there was an increase in the number of nodes and in the number of edges m. For all the networks

$2 \leq \bar{k} < 3$ and was directly proportional to the difference $m - n$. The average clustering coefficient was still very low, remaining even zero in some networks. There was a decrease of the efficiency in each network, because each new node was not connected to all existing nodes. This results are shown in figure 3.5.

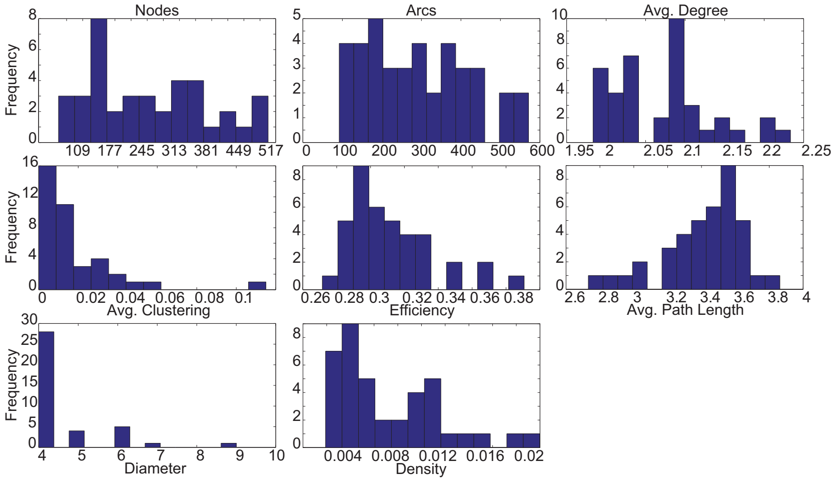


Figure 3.5: From left to right and from top to bottom, the histograms of the frequencies of: number of nodes, number of arcs, average degree, average clustering coefficient, efficiency, average path length, diameter and density of the networks of Experiment 2, part A. The histograms have been created with an ad hoc developed Matlab code derived by using the package Brain Connectivity Toolbox (Rubinov & Sporns, 2010), (http://www.brain-connectivity-toolbox.net/).

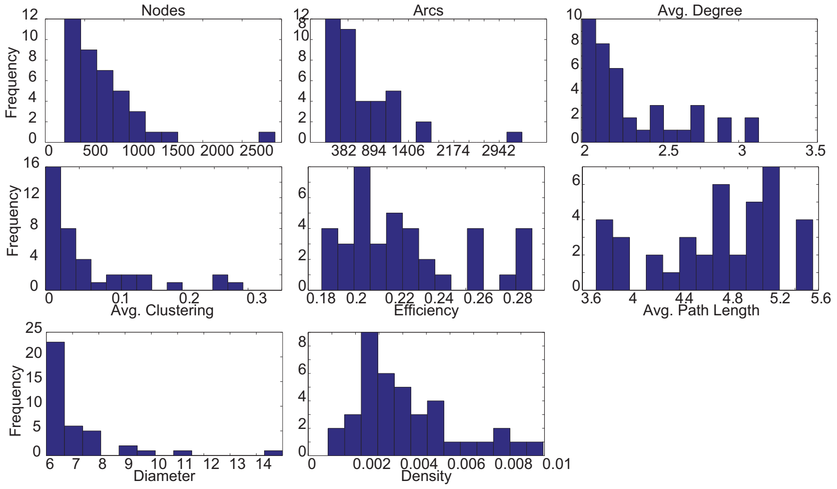


Figure 3.6: From left to right and from top to bottom, the histograms of the frequencies of: number of nodes, number of arcs, average degree, average clustering coefficient, efficiency, average path length, diameter and density of the networks of Experiment 2, part B. The histograms have been created with an ad hoc developed Matlab code derived by using the package Brain Connectivity Toolbox (Rubinov & Sporns, 2010), (http://www.brain-connectivity-toolbox.net/).

Experiment 2 Networks, First Stage Part B: Synonyms of Synonyms

In this second part of the first stage of the second experiment, all the script networks showed a considerable increase in both the number of nodes and edges, with $m > n$ holding for every network. Thus, for all networks, $\bar{k} > 2$. The average clustering coefficient remained always very low (in two networks continued to be zero), but did not decrease, as well as the average shortest path length and average diameter. Here, too, the density decreased. These results are shown in figure 3.6.

Experiment 2 Networks, Second Stage: Hyponyms and Hypernyms

In the last stage of the second experiment, nodes and edges increased in all networks, of course. Instead, average degree, average clustering coefficient, efficiency and average path length of the networks did not follow the same trends. In fact, some networks shown an increase in the values, while others did the opposite. Even for the diameter a unique trend did not exist. There were networks in which such a measure remained unchanged or increased slightly, while some networks shown a decrease. Regarding the density, with the exception of two networks in which there was an increase of the value, an overall decrease for this variable was observed. In figure 3.7 the results of the second stage of experiment 2 are shown.

Macro Level Analysis

Regarding the networks resulting from the union of the networks built from each subject for the two experiments, a monotonic increase in number of nodes, number of arcs and clustering coefficient was detected during the four steps, while the density value decreased monotonically. The average degree decreased in experiment 2, part A, and increased in the next steps. The efficiency decreased until

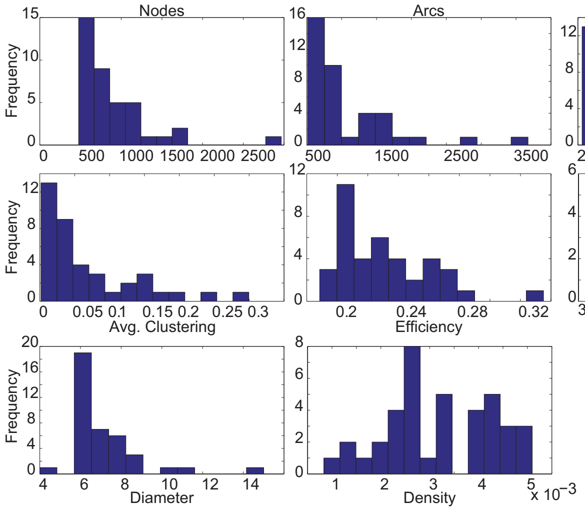
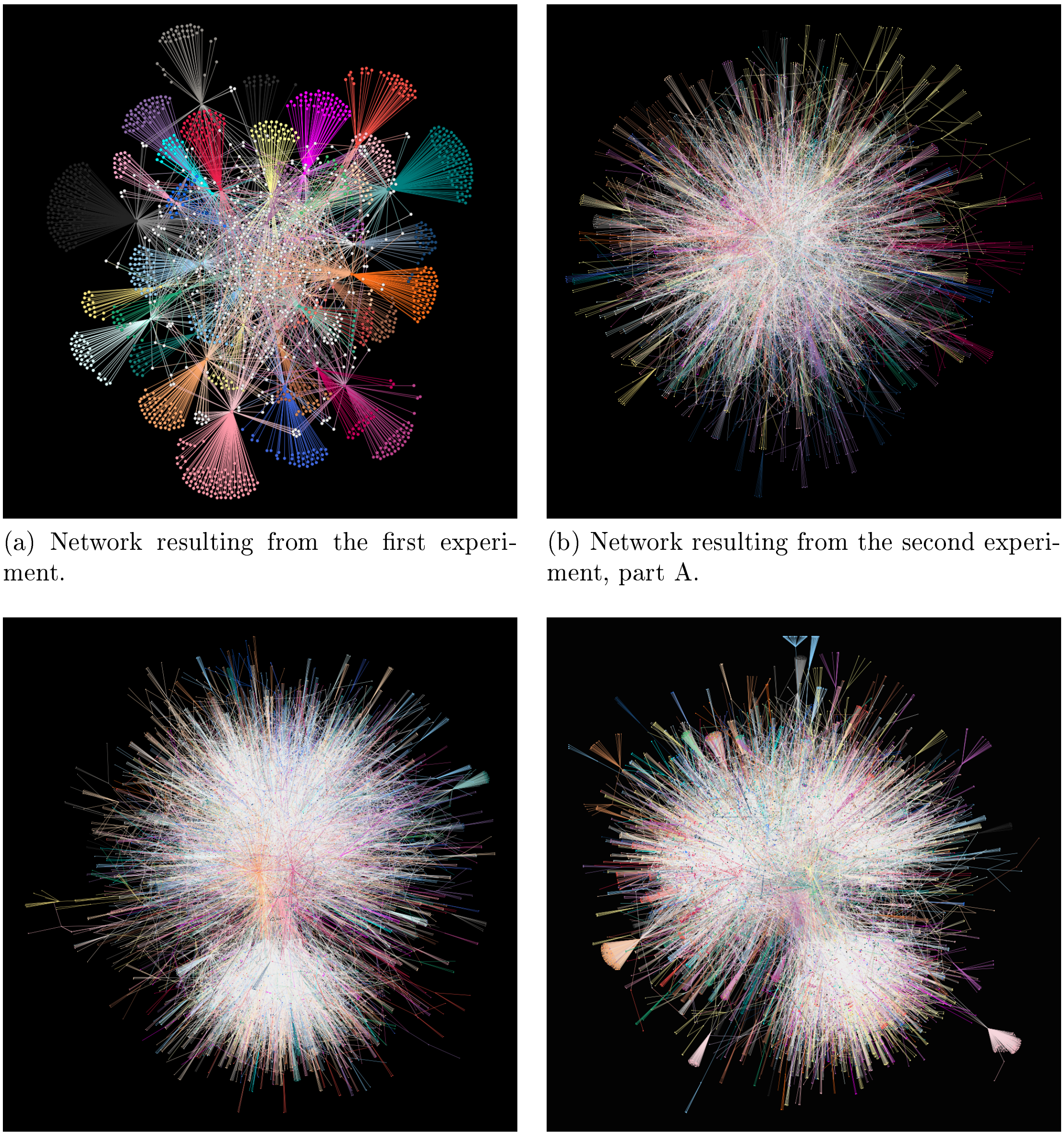


Figure 3.7: From left to right and from top to bottom, the histograms of the frequencies of: number of nodes, number of arcs, average degree, average clustering coefficient, efficiency, average path length, diameter and density for networks of the second stage of experiment 2. The histograms have been created with an ad hoc developed Matlab code derived by using the package Brain Connectivity Toolbox (Rubinov & Sporns, 2010), (http://www.brain-connectivity-toolbox.net/).

Network	n	m	k	C	E	L	d	δ
Experiment1	1770	2750	3.10734	0.00072	0.27085	3.86783	6	0.00176
Experiment2, part A	6595	9919	3.00773	0.05579	0.21648	4.84330	8	0.00046
Experiment2, part B	14505	28784	3.96842	0.11070	0.19435	5.38687	11	0.00027
Experiment2, last stage	17429	34680	3.97917	0.11750	0.19943	5.23480	11	0.00023

Table 3.2: Networks statistics for the networks resulting from the union of the networks built from each subject for the two experiments. Data have been obtained by using Matlab routines derived from Brain Connectivity Toolbox (Rubinov & Sporns, 2010), (http://www.brain-connectivity-toolbox.net/).
n = number of nodes, m = number of edges, k = average degree, C = average clustering coefficient, E = global efficiency, L = average path length, d = diameter, δ = density.

experiment 2, part B, and increased in the last part. The average shortest path length increased until experiment 2, part B, and decreased slightly in the last part. The diameter instead increased until experiment 2, part B, and remained unchanged at the end (see table 3.2). In figure 3.8 four images drawn from the networks resulting from the union of the networks of all the subjects, for the two experiments carried out in this work.



(a) Network resulting from the first experiment.

(b) Network resulting from the second experiment, part A.

(c) Network resulting from the second experiment, part B.

(d) Network resulting from the second experiment, last stage.

Figure 3.8: Networks resulting from the union of the networks of all the subjects, for the different steps of both experiments. Images of the networks was drawn with Gephi 0.8.2 (gephi.github.io)

Network	C_{real}	C_{rand}	L_{real}	L_{rand}	γ	λ	σ
Experiment1	0.0007	0.0991	3.8678	3.2992	0.0073	1.1724	0.0062
Experiment2, part A	0.0558	0.0070	4.8433	4.6473	8.0239	1.0422	7.6993
Experiment2, part B	0.1107	0.0024	5.3869	5.0175	45.3787	1.0736	42.2669
Experiment2, last stage	0.1175	0.0024	5.2348	4.9212	48.1426	1.0637	45.2590

Table 3.3: Small world analysis of the four aggregated networks. Data have been obtained by using Matlab routines derived from Brain Connectivity Toolbox (Rubinov & Sporns, 2010), (http://www.brain-connectivity-toolbox.net/)

Small World and Scale Free

Small world analysis was performed on the four aggregated networks (table 3.3). To perform this task, the four networks were rewired while preserving degree distribution to create equivalent random networks. Average clustering coefficient C and average path length L were calculated for both real and random network. Then the ratios between C_{real} and C_{rand} and between L_{real} and L_{rand}, γ and λ were calculated as well as the ratio between γ and λ, named σ.

The first network does not show small-worldness: not just because of $\sigma < 1$ but because of $C_{real} \ll C_{rand}$ and consequently $\gamma < 1$. The third and fourth networks have a high value for σ: in this case they are small world networks because $\gamma \gg 1$ and $\lambda \approx 1$. The second network also has $\sigma > 1$, $\gamma \gg 1$ and $\lambda \approx 1$, but C_{real} seems very small. In Watts (1999), an electric power grid with 4941 nodes (1654 less than the network analyzed in this work) and $C_{real} = 0.080$ (not so far from the value obtained here) is found, and it is considered a small world network. This brings to the conclusion that the second network shows small world properties, too. The cumulative degree distribution of every network was fitted with three laws: truncated power law, power law and exponential law. The fits were compared by adjusted R^2. The cumulative distribution was used to avoid fluctuation on the tail. Starting from the second experiment, 30 networks were intercepted by a power law, 7 networks by a truncated power law, and 2 by an exponential law. In part B of the second experiment, the trend was reversed in only 15 networks, intercepted by a power law, 22 by a truncated power law. Instead, only 2 two of the networks behaviors were described by an exponential law. In networks of the last stage of

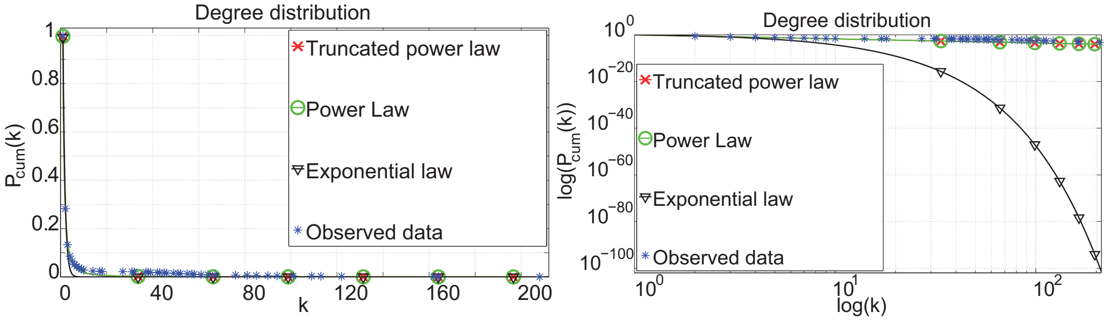


Figure 3.9: Fitting curves for the degree distribution relating to the combined network for experiment 1. The curves have been plotted on Cartesian axes on the left, while on the right are plotted on a logarithmic scale. The power-law and truncated power law are almost overlapped and R^2_{adj} is in favor of the power law. The exponential law was not appropriate to describe these data. The plots were drawn by Matlab after having computed the best fit with its standard routine.

the second experiment, 23 of them were intercepted by a power law, 14 networks by a truncated power law, while only 2 networks followed an exponential law. The final network of 17,429 terms, resulting from the merging of all the networks produced by the subjects was found to be connected, also in the initial stage of experiment 1. The network resulting from the merging of experiment 1 and experiment 2, part A, had a power law degree distribution (Clauset, Shalizi, & Newman, 2009). Between experiment 2, part A, and Experiment 2, part B a transition from scale free to broad scale behavior was detected. In this case, the resulting network was described by a truncated power law (Sjöberg, Albrectsen, & Hjältén, 2000). It has been proven that many biological data (Khanin & Wit, 2006), including the language (Arbesman, Strogatz, & Vitevitch, 2010b), instead of following a power law, do follow a truncated power law. Small variation and the fact that language is a complex system may influence this behavior. The fitting curves for the four stages are shown in figures 3.9, 3.10, 3.11, 3.12. The numerical parameters are reported in table 3.4.

8 categories (adjective, adverb, conjunction, determiner, interjection, noun, verb and preposition), taken from the dictionary of American English language, contained in Wolfram Mathematica, were used to analyze the organizational structure of the resulting lexicon. Unfortunately, Mathematica dictionary does not

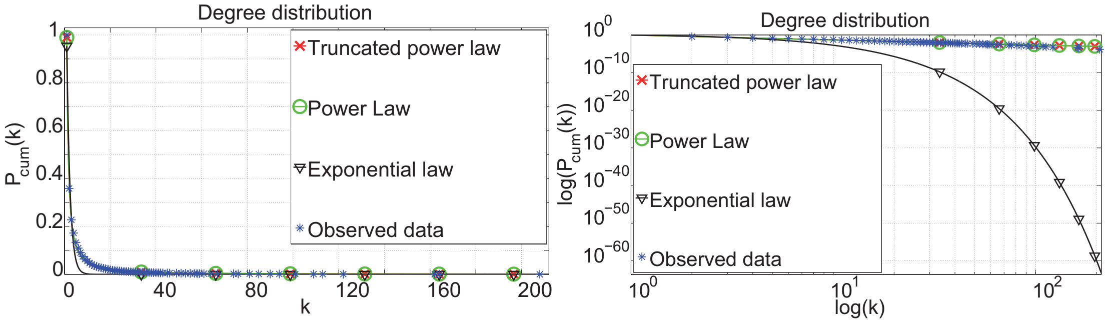


Figure 3.10: Fitting curves for the degree distribution relating to the combined network for experiment 2, part A. The power law and the truncated power law were almost overlapped and R^2_{adj} was in favor of the power law as in the previous network. The plots were drawn by Matlab after having computed the best fit with its standard routine.

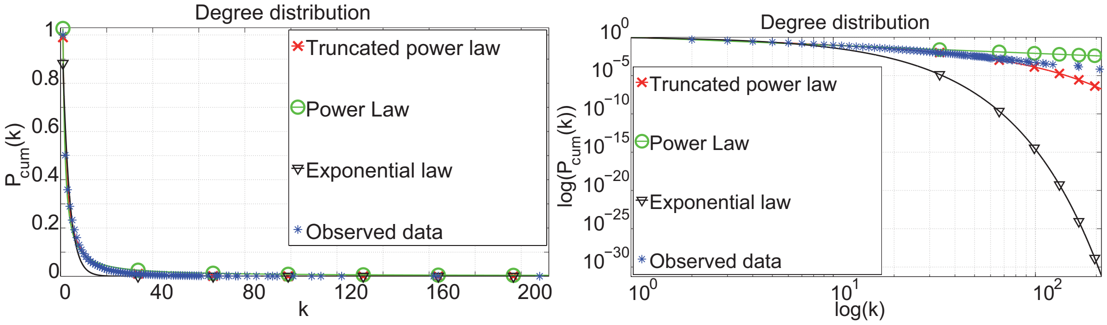


Figure 3.11: Fitting curves for the degree distribution relating to the combined network for experiment 2, part B. The separation between the power law and the truncated power law was clear, especially in the tail of the data distribution. The plots were drawn by Matlab after having computed the best fit with its standard routine.

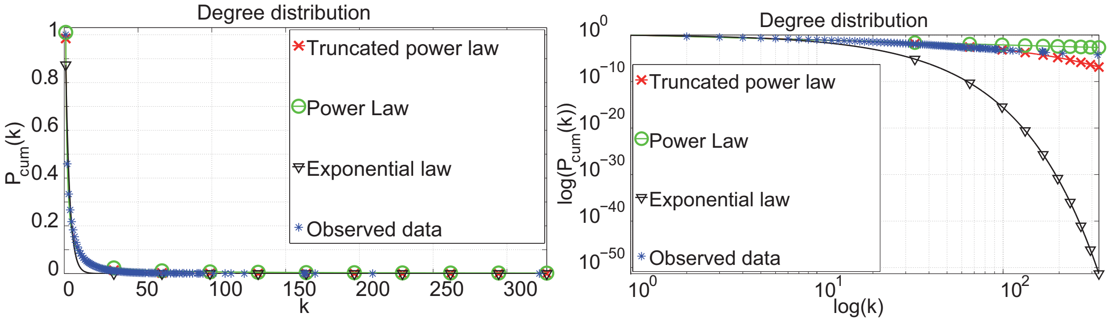
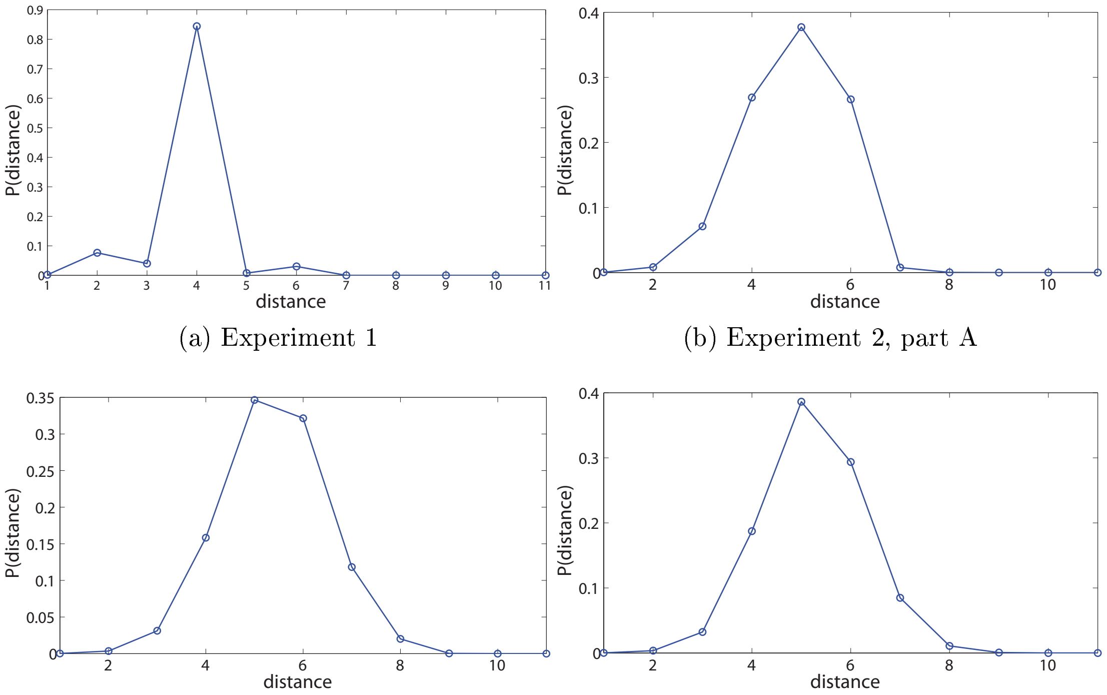


Figure 3.12: Fitting curves for the degree distribution relating to the combined network for experiment 2, last stage. The truncated power law was the one fitting the data the best. As in the previous networks the exponential law is not appropriate to describe these data. The plots were drawn by Matlab after having computed the best fit with its standard routine.

Experiment 1	**a**	$\mathbf{k_c}$	**b**	$\mathbf{R^2_{adj}}$
Truncated power law	0.9957	4811.4	-1.7397	0.9938
Power law	0.9955	///	-1.7400	0.9940
Exponential law	2.9970	0.9033	///	0.9776
Experiment 2, part A	**a**	$\mathbf{k_c}$	**b**	$\mathbf{R^2_{adj}}$
Truncated power law	0.9900	3983.2	-1.3106	0.9975
Power law	0.9899	///	-1.3115	0.9975
Exponential law	1.9016	1.4480	///	0.9470
Experiment 2, part B	**a**	$\mathbf{k_c}$	**b**	$\mathbf{R^2_{adj}}$
Truncated power law	1.0441	18.8741	-0.8115	0.9984
Power law	1.0269	///	-1.0536	0.9861
Exponential law	1.2380	2.9506	///	0.9511
Experiment 2, last stage	**a**	$\mathbf{k_c}$	**b**	$\mathbf{R^2_{adj}}$
Truncated power law	1.0173	30.772	-0.9059	0.9970
Power law	1.0101	///	-1.0662	0.9903
Exponential law	1.2549	2.7672	///	0.9336

Table 3.4: The table shows the values obtained with a Matlab package. Data have been obtained by combining ad hoc scripts with the Matlab "cftool" package. The fit were computed with a power law of the form $P(k) = ak^b$, a truncated power law of the form $P(k) = ak^b e^{-k/k_c}$ and an exponential law of the form $P(k) = ae^{-k/k_c}$. The best fit was chosen comparing the R^2_{adj} relative to the fits.

include terms composed of more than one word. So, from the 17,429 words contained in the final lexicon, the automatic classification of only 7000 terms was possible. The remaining words were classified manually, using an online research on different dictionaries. After the classification, fits were performed category by category. In experiment 1, categories such as conjunctions and propositions did not have enough data, while in experiment 2, they seemed to follow a power law. Neither determiners had enough data in experiment 1, but it seemed to follow an exponential law in experiment 2, and a power law in the last two parts of experiment 2. The category interjections showed a fluctuating behavior, following a power law in experiment 1 and 2, part A, and an exponential law in part B and in the last stage. In the whole lexicon, the category of nouns obeyed to a truncated power law. The category of adverbs registered a power law in experiment 1 and a truncated power law in experiment 2. The last two categories, adjectives and verbs, adopted the trend of the whole network. They presented a scale free behavior in experiment 1 and 2, part A and a broad scale law in the last two steps



(a) Experiment 1

(b) Experiment 2, part A

(c) Experiment 2, part B

(d) Experiment 2, last stage

Figure 3.13: From left to right and from top to bottom, distance distributions for the networks of the two experiments. All the distribution shown a peak that shifted to the right during the evolution of the network. The plots were drawn by Matlab after having computed the the distances between each pair of nodes.

of the experiment.

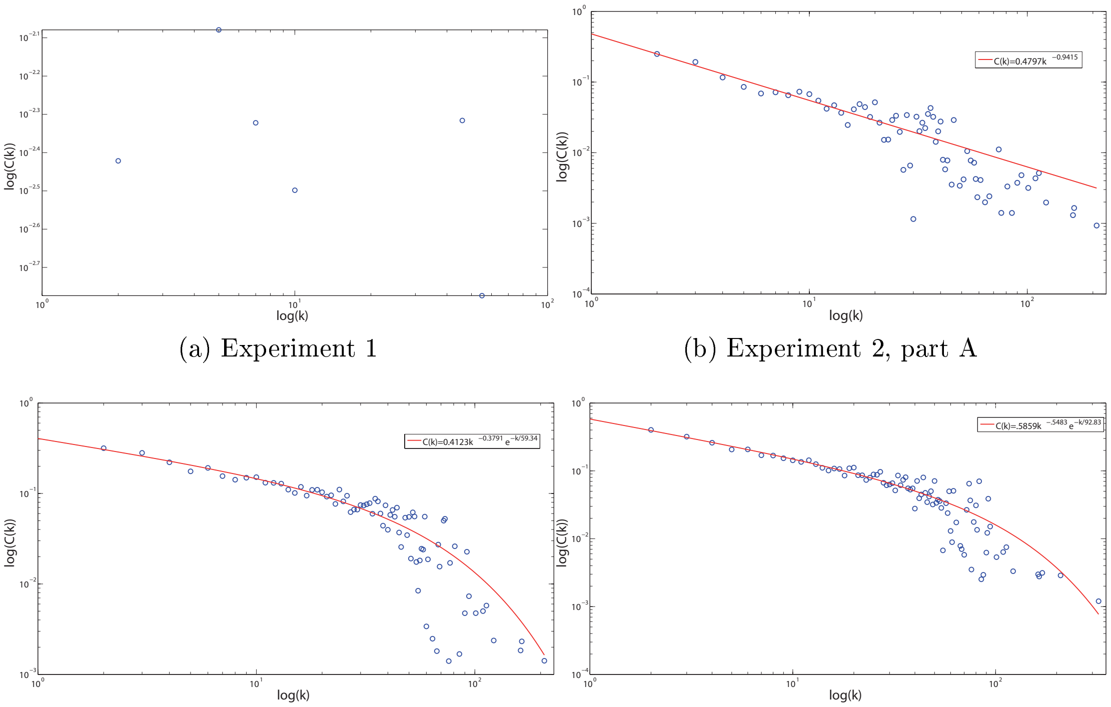
Distance and Degree Correlations

To investigate deeply the semantic network built from the experiments, the distribution of the shortest paths lengths, clustering coefficient as a function of the degree and the K-Nearest-Neighbor were also analyzed. For all the four networks, all the distance between nodes were computed, then their distributions were plotted. As shown in figure 3.13, the majority of nodes in the network from the first experiment are placed at a distance of 4, while this peak shifts slightly to the right with a short tail in the other networks. These networks average path length and diameter resulted to be probably greater than an equivalent syntactic network because of the lack of functional words which are very rare here (Liu, 2009).

The correlation between clustering coefficient and the degree was investigated. The average of the clustering coefficient of every node with degree k was computed.

Thus the function $C(k)$ was analyzed for the four aggregated networks. This function express the presence or absence of a hierarchical organization of the network (Ravasz, Somera, Mongru, Oltvai, & Barabási, 2002). In figure 3.14 is shown that the network from experiment 1 was not complex enough to show this kind of characteristic. The three networks from the second experiment, on the other hand, shown a good correlation, i. e. a strong hierarchical structure among nodes: The second network clustering distribution was well fitted by a power law, while the last two by a truncated power law. The way this networks were built played a fundamental role in this result: the central node of every script generated a star of terms and every one of them did the same in the next step with the additional birth of links among them. Then while every node was responsible for the birth of the nodes in the next step, it linked with other nodes of the same experiment. This increased the clustering coefficient while building a hierarchical structure. Thus the memory system, seen as a growing structure, evolves in a way that maintain this kind of structure to create useful paths when a specific word needs to be retrieved. These paths belong to communities made of low degree nodes connected to a local hub (Pastor-Satorras & Vespignani, 2004). The hierarchical structure is needed to simplify and accelerate this kind of memory processes.

K-Nearest-Neighbor K_{NN} is a correlation measure that gives the assortativity of a network, i. e. how much the nodes with low (high) degree are connected among them. If this happens, the network is said assortative, if not it is disassortative. A simplified way to compute $K_{NN}(k)$ (Caldarelli, 2007) consists in calculating the average degree of the neighbors of every node and then averaging it over nodes with the same degree k. The behavior of the function tells everything on the assortativity of the network: if it is directly proportional to k, the network is assortative; if it is inversely proportional, the network is disassortative; if the function is almost constant there is no correlation among nodes with the same degree. The network of the first experiment shown a bipartition of K_{NN}



(a) Experiment 1

(b) Experiment 2, part A

(c) Experiment 2, part B

(d) Experiment 2, last stage

Figure 3.14: While the first network was not complex enough for this kind of analysis, the second clustering distribution averaged on the degrees $C(k)$ followed a power law and the last two a truncated power law. The networks from the experiment 2 shown a correlation between clustering coefficient and degree, highlighting a hierarchical structure. The plots and the fits were obtained by Matlab.

values: as it was expected, the low degree nodes were connected with the high degree ones, and the data points in the plot were not a curve but two flat lines. It was reasonable to say that this network was disassortative because the links were almost between a high degree node and a low degree one. The network of the experiment 2, part A was slightly disassortative 3.15. The third network plot was almost constant, so there was no correlation among degrees when the network was filled up with the missing synonyms. This behavior was broken again in the last network: the introduction of hyponyms and hypernyms specialized the network. Stars of terms were linked to a node, increasing its degree while decreasing its average neighborhood degree. Excluding these few nodes represented in the tail of the distribution in 3.15, the rest of the plot was flat, so the correlation is not very significant, nor strong. Usually a semantic network does not show strong correlation between K_{NN} and the degree because of the lack of functional words like prepositions: functional words, in syntactic networks, are usually connected to many low degree resulting in a disassortative network. While processing a word retrieval task, it is unneeded to go through functional words, but is needed, instead, to pass from a concept to another. Therefore there is no need for such a network to be strongly disassortative.

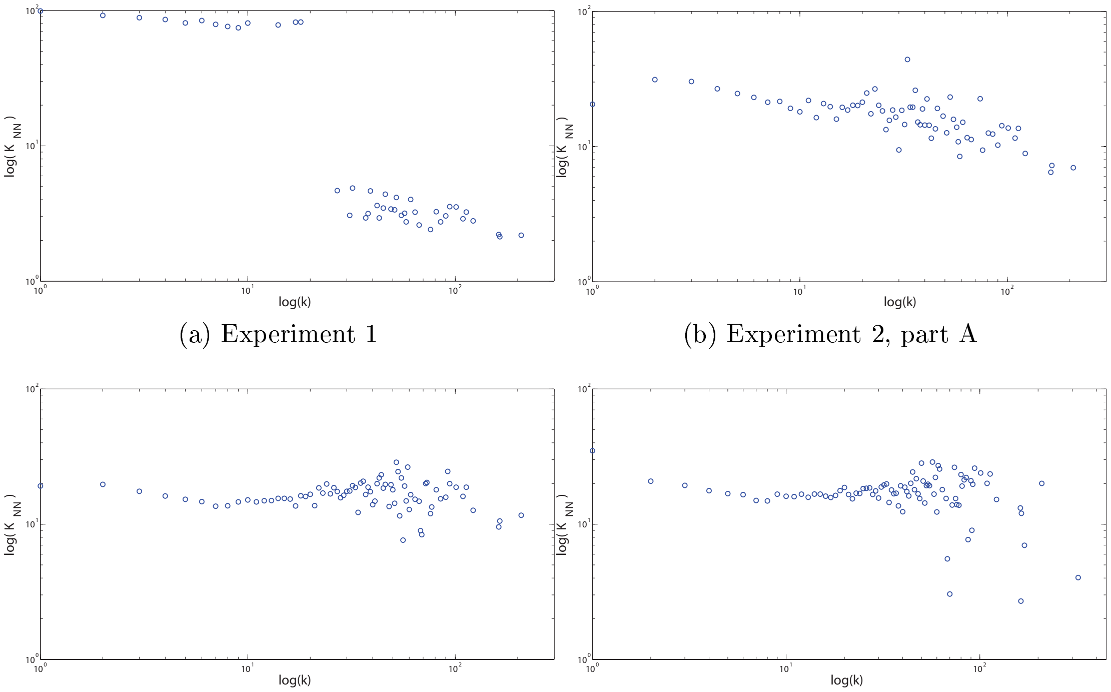
From this kind of performance, it seemed that the construction of the network did not follow a linear path, depending on the degree of the individual terms, resulted by the process of word connections operated by students, between words with different degree. This could lead us to conclude that language networks were different from all other real networks, with a scale free and behavior. Instead, having such discontinuities, this behavior can best be intercepted by a truncated power law, which produced the typical displacements in the fitting of the growth curve

3.6 Conclusions

These results highlighted how concepts are organized in memory and how memory is organized for scripts, but also how they are retrieved.

From the analysis of the networks, meso and macro levels interaction was evident (Borge-Holthoefer & Arenas, 2010).

From the analysis, it was also evident that data coming from the exploration of semantic memory for script, related to the first experiment, were not complex enough to be of interest for the present study, because the networks had very few nodes (just 7 had more than 100) and they were star graphs. So, no network parameters, but the number of nodes, were useful to draw conclusions on the underlying cognitive processes. Some scripts had many nodes, because they were very familiar for everyone (like "Go to a party" or "Go to the park"). They could be expressed by very common words, actions and things. Other scripts presented few nodes as they were unfamiliar situations for the majority of the subjects, such as "Horse riding" or "Give private lessons", or also considered boring activity, like



(a) Experiment 1

(b) Experiment 2, part A

(c) Experiment 2, part B

(d) Experiment 2, last stage

Figure 3.15: The first experiment resulted in a disassortative network: the high degree nodes linked almost only to the low degree ones and vice versa. In the other experiment, only the network from part A has shown a light correlation between K_{NN} and degree, resulting in a light disassortativity. The other two networks had an almost flat distribution. Because of the hyponyms and hypernyms, the last network distribution had some noise in the tail. The log-log plots were obtained by Matlab.

"Clean the room".

On the second experiment, the complexity of the networks increased. Scripts with a low number of synonyms shown a lower average clustering coefficient, while scripts with a high number of synonyms, decreased their density parameter. This could mean that it could be possible to segment for each student, for each script, the amount of terms at the micro level of the networks, turning the networks to be as a dictionary for each student, which could be useful for second language teaching applications.

After the merging of the networks, four graphs have been obtained, sufficiently complex to show interesting results. The introduction of an online dictionary to complete the task, expanded the opportunity to expand the semantic space. In fact, part B of the second experiment collected the highest number of new nodes. This result was confirmed by the increase of the average degree of the related networks: every term had an average of one new term linked to it. Thus, the distances among nodes did not become too high because of the synonyms: they connected different parts of the network, allowing them to be settled in the same neighborhood. Synonyms increased the average clustering coefficient while preserving the average path length, resulting in a small world behavior: there were no distant semantic areas, while the concepts remained well defined. As expected, the efficiency of the network was low. Even having a low average path length, the density was close to zero. This happened because the memory system was not based on shortest paths, but on logic associations: it had longer but more typical or more atypical paths. The truncated power law shown that the new terms tended to link themselves with high degree nodes, but also that the system presented relatively few hubs. Keeping low as possible the number of hubs, as said before, could be a good strategy to keep short distances and a high preference for logic association paths instead of the generic geodesic ones. The memory system shown sign of a hierarchal organization, as suggested by the clustering as function of the degree. The memory grew collecting terms and concepts that were linked to some preexistent ones, so every concept had a new starting point to increase the data into this system, with relations like synonymy, hyponymy and hypernymy. The lack of functional terms in this system, created neighborhoods of terms with different degrees. The terms here are not connected each other because they have the same importance, but because of a semantic and conceptual relationship, and this was clear in the K-Nearest-neighbors analysis.

At the micro level of the network, interesting dynamics of the degree and clustering have been observed. The starting nodes, had a greater degree, while the others have degree one or two. When hyponyms and hypernyms was introduced, many nodes (in particular nouns) increased their degree, but the average degree of the network remained almost the same. As the growth process went on, the nodes forming the base of the script, increased their degree, creating of stable interrelated elements of the scripts in memory. Continuing the growth process, the network did not increase its average degree: not every synonym had new connections. This could mean that going far from the starting node, subjects did not find connections, and the script organization became even less connected. Only in the final part of the second experiment, the average degree of the network increased locally, by connecting semantic areas until then separated. The first task assigned caused the first connections to be inherent to the script, but synonyms corrupted the semantic area acting as a bridge between two different areas. The clustering coefficient followed exactly the behavior of the degree. In the first level, stars of nodes connected very weakly to each other by nodes belonging to more than one script, with a very low clustering coefficient. Synonyms led to a strong increasing for this parameter. Obviously, the great increasing of the number of nodes decreased the density of the network, so the average clustering coefficient was just over 0.1. The network had not only an increasing of bridges, as different meanings of the same word, but also an even more increasing of the nodes sharing synonyms. Thus, subjects chose the most pertinent meanings for each script. The dimension of their vocabulary represented a limitation as well, both for the number of nodes and the clustering: having a wider vocabulary allowed the subjects to bridge these gaps and disregard the situation, widening the meanings of the chosen terms.

When analyzing the meso level, the search for communities were thought to produce good results, but the network topology caused the nodes belonging to more scripts to rule the communities organizations, instead of the nodes belonging to only one script. Groups of nodes belonging to several scripts joined in communities, dragging along pieces of other scripts, thus creating heterogeneous communities which were not suitable to be combined with the typicality scores.

The network's growth was not described by a power law because the subjects did not follow a preferential attachment algorithm (Steyvers & Tenenbaum, 2005), but realizing random connections, regardless of whether or not the node had a certain semantic centrality in the construction of the network. So, the semantic network, representing the subjects' memory system, was built differently from spreading activation model.

Compared to the model of (Collins & Loftus, 1975), this is an alternative model (which shows chaotic dynamics (Bilotta & Pantano, 2010; Bilotta, Pantano, & Stranges, 2007)) of processes that take place not only with spreading activation, but through a construction, reconstruction of meaning and typicality, making always new connections between different semantic levels.

Chapter 4

Mental Lexicon and Language Structures

4.1 Introduction

Network science is a very effective tool at human semantic knowledge level: semantic networks are useful to investigate new words learning and words retrieval in the mental lexicon (Steyvers & Tenenbaum, 2005; Vitevitch, 2008). It proved to be useful as statistical and computational model even in the comparison of different languages semantic networks (Arbesman et al., 2010a). In all these studies, the component analysis highlighted the importance of both large and small well connected components and the importance of the connections between them, as well. In every semantic network, there is also the possibility to find some isolated components, sometimes composed by a single node, which are not connected to the rest of the network, representing a specialized knowledge area. But the main large connected component always exhibits a broad scale (or scale free) and a small world behavior (Barabási & Albert, 1999; i Cancho, 2005; Watts & Strogatz, 1998). All of these works focused on morphology, phonology and semantic or syntactic representation, but none of them used scripts (R. Schank, 1982; R. C. Schank & Abelson, 1977) and network science together.

The script model (R. C. Schank & Abelson, 1977) hypothesize that information retrieval in the human mind is made from a long term memory (Allington, 2005). This retrieval task is performed for memory recall (Light & Anderson, 1983), of course, but also for language or text comprehension (Bower et al., 1979; Cellar & Barrett, 1987; Gernsbacher, 1991; Pollatsek et al., 2012; Zwaan & Radvansky, 1998), or to interact with other people (Abelson, 1976).

The script implies a temporal dimension: the script is like a story told from the beginning to the end. Thus, this model introduces some hidden rules and inferences which raise its complexity Abelson (1981). But one of the aims of this thesis is to represent all of these inferences with a semantic network based on scripts, with all the arcs among the nodes to fully include all the hidden rules.

The network obtained in the experiments described in chapter 3 was used side by side with the Free Association Norms (Nelson et al., 2004), gaining more (directed) connections among terms in the same script.

To detect the relationship between word association and typicality score (script cohesion) and between word association and network topology, random walks were used.

Random walking (Pearson, 1905; Révész, 1990) was the simplest method to model memory for scripts. Since classic random walk could be too simple as a model, biased random walks, using Page Rank (Griffiths et al., 2007) or typicality score were utilized.

4.2 Modularity and Community Detection

The single script networks obtained in the experiments had a low clustering coefficient, so none of the networks had a significant community structure. With the merging of these networks, new communities were formed from parts of different scripts which shared some meaning. These communities were made up of parts from two or three different scripts (only in two of them four scripts contributed to the community). When all the fully evolved networks were joined, a strange result was obtained: with a modularity value of about 0.75, and a number of communities between 32 and 37 (values averaged over several runs of Louvain method (Blondel et al., 2008)), all the communities had not the majority of nodes coming from a single script, but they were determined by nodes shared among several scripts. Thus, in this model of mental lexicon, the memory was not guided from the different scripts but, the nodes used in several scripts acted as bridges grouping together groups of words and connecting parts of different scripts. Just in case, several modifications of the resolution parameter in the Louvain algorithm (Blondel et al., 2008) were made, to confirm that the network has been observed with the right resolution: the only effect was a smaller modularity value. So, the communities and their composition were robust and significant. All these results, with the percentage nodes belonging to several networks are shown in 4.1.

Table 4.1: Modularity, number of communities and shared nodes results for the final network.

Network	Modularity	Number of Communities	shared nodes
Experiment 1	~ 0.63	$17 \div 19$	26%
Experiment 2 part A	~ 0.75	$37 \div 40$	30%
Experiment 2, part B	~ 0.75	$31 \div 36$	38%
Experiment 2, last stage	~ 0.75	$32 \div 37$	35%

The detection of overlapping communities could figure out if these networks can constitute a benchmark for the various methods of community detection or if the initial networks are not significant modules.

4.3 Modeling

Free Association Norms

A network was built starting from the University of South Florida Free Association Norms database (http://web.usf.edu/FreeAssociation/) (Nelson et al., 2004) to strengthen the model of the semantic network built from the experiments. This collection of terms, was built giving some cue words to subjects who had to answer

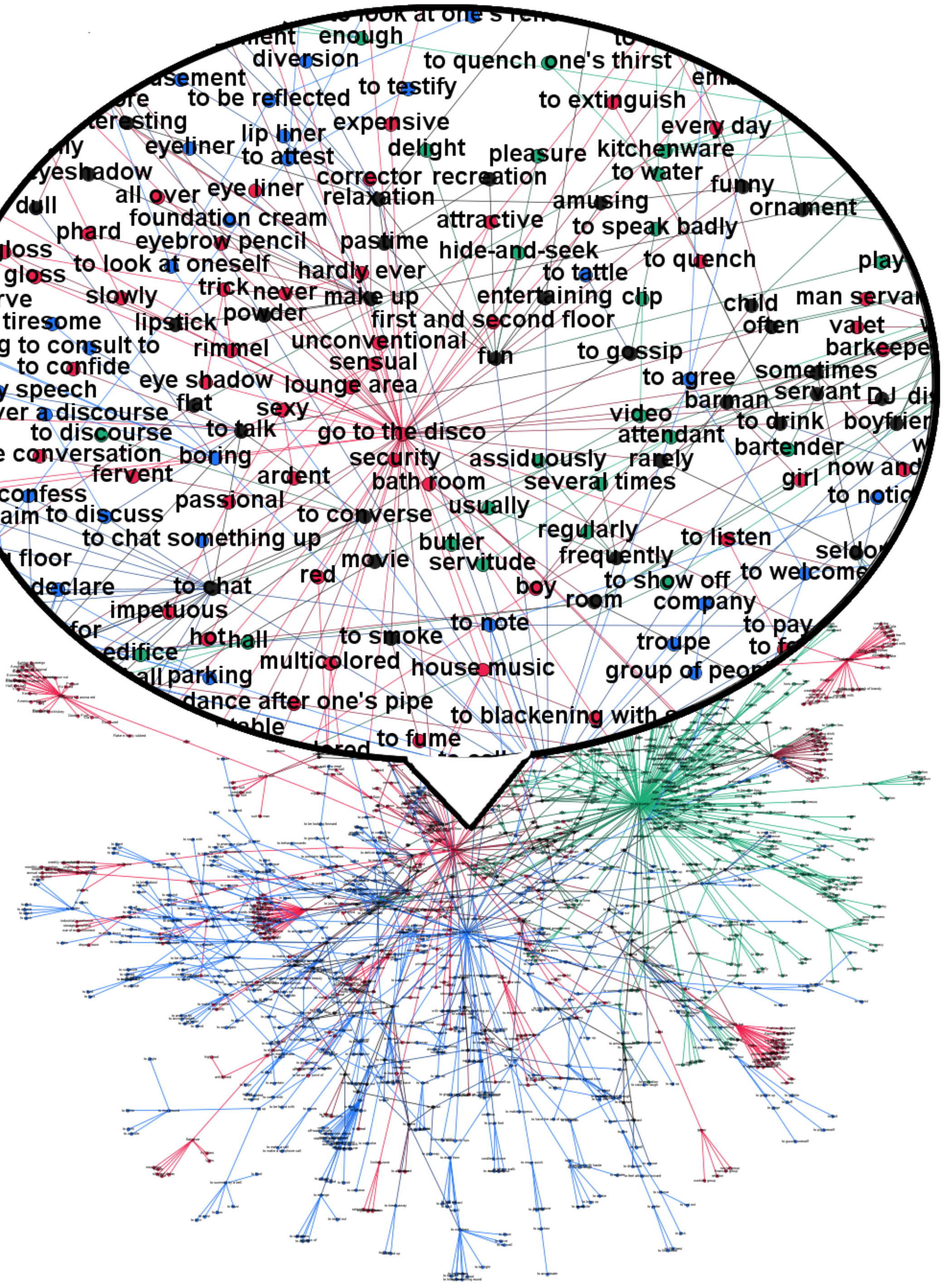


Figure 4.1: Three joined script networks. The black nodes are shared by the three networks and represent their shared cluster.

with a response one, following the first association which came to their mind. Every node in the network represented a different word in the database. If the node n_i represent a cue term which has a response term represented by the node n_j, then a directed arc from n_i to n_j was inserted in the network, making it a directed network. This network was also a weighted one: the strength of the arc going from n_i to n_j, represented the fraction of subjects who answered with the term represented by n_j to the cue term represented by n_i. Obviously, there are some terms which appear only among response ones, so these nodes had only incoming arcs, while the cue terms which were not found in the response ones had only outgoing arcs. The final network, called $\mathcal{N}$ had 10617 nodes and 72176 weighted and directed arcs. Thus $\mathcal{N}$ and $\mathcal{R}$, the final network from the experiments, were used in our model.

Random Walk Model

A random walk (Pearson, 1905) is a sequence of steps in which each step is taken in a random direction among every possible one. A discrete random walk on a network with adjacency matrix $A = \{A_{ij}\}$(Révész, 1990), can be described as follows: putting a random walker on a node n_i of the network, it will walk to another node, choosing randomly among one of the outgoing arcs of n_i, reaching one of its neighbors n_j. Here, the random walk is used to represent the search for terms useful to write down a script.

Calling k_i^{out} the outdegree and s_i^{out} the outstrength of the node n_i, and w_l, $l = 1, \ldots, k_i^{out}$ the weights of the arcs outgoing from n_i, the probability for the random walker to choose the arc e_l with weight w_l and reach the node n_j is $P(n_j|n_i) = \frac{w_l}{s_i^{out}}$.

This definition of probability holds perfectly for both $\mathcal{N}$ and $\mathcal{R}$. $\mathcal{N}$ is directed and weighted, so there is no possible misunderstanding on the meaning of s_i^{out} and w. For $\mathcal{R}$, which is an undirected and unweighted network, $k_i = s_i^{out} \, \forall i$, because an unweighted network can be seen as a weighted one with unitary weights, and an undirected network is a directed one with the outdegree equal to the indegree for every node. Thus, for $\mathcal{R}$, the same probability defined before can be written as $P(n_j|n_i) = \frac{1}{k_i}$

In $\mathcal{N}$, for a terminal node n_i, holds $k_i^{out} = 0$ and is an attractor: once the walker reach that node, it is trapped and has no more steps to take. After m steps, a random walker covered a path C, with $|C| \leq m$. This path C is a subgraph of the whole network.

Let $C_y^x(i)$ be the subgraph composed by the paths walked by x random walkers, starting from the node n_i performing y steps. By definition, $C_y^x(r)$ is the semantic field of node n_i. So, it is possible to define a semantic similarity:

Definition 4.1. In a semantic network, two nodes n_i and n_j are semantically similar if $C_y^x(u) \cap C_y^x(w) \neq \emptyset$. The cardinality of the intersection measures how much the nodes are semantically similar.

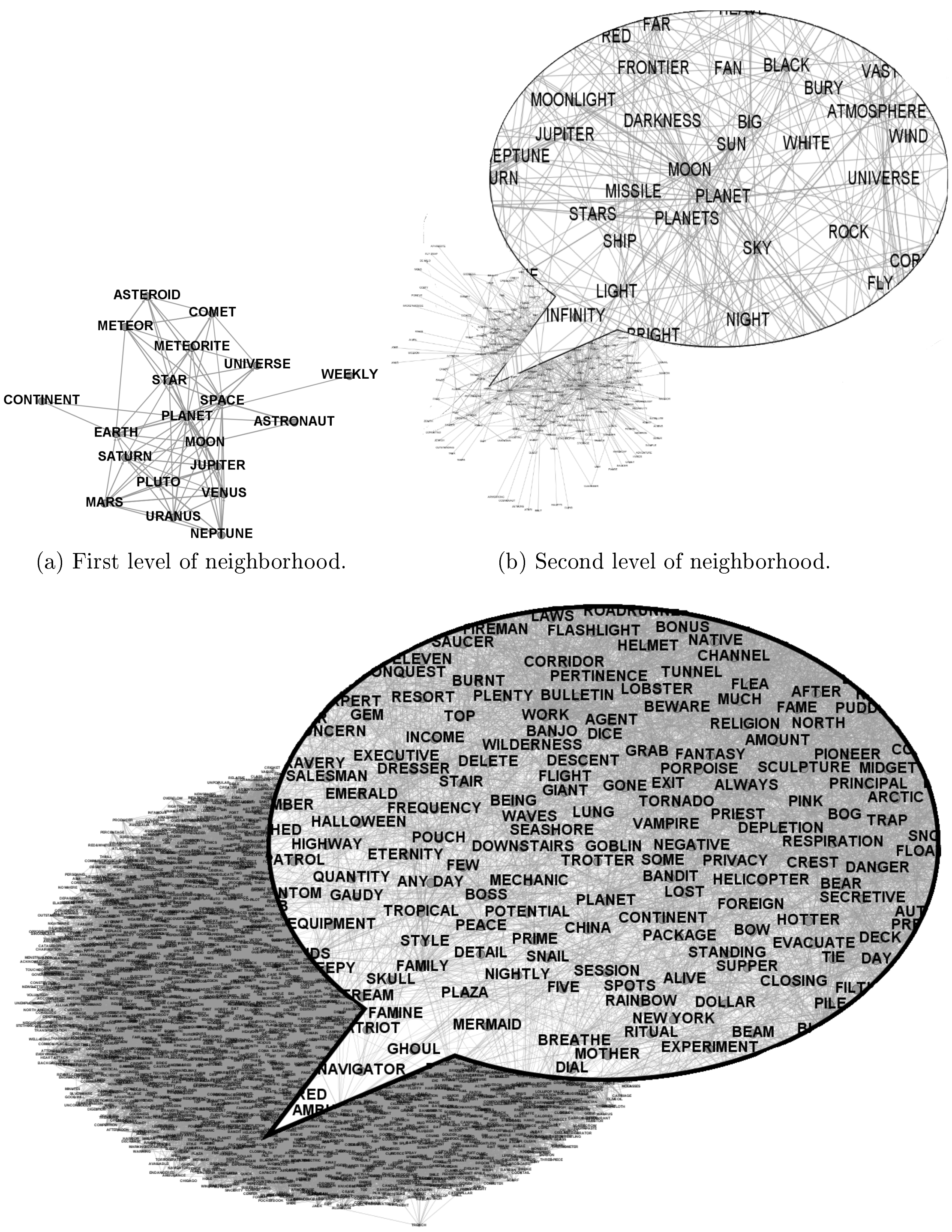
Random walk model, after several simulations on the network representing subjects' mental lexicon, proved to be a valid model to describe the subjects cognitive organization.

Neighborhoods

A Matlab code (neighborhood_FAN.m) was developed to extract the neighborhood of a node in $\mathcal{N}$, then the neighborhood of these neighbors, and finally, the neighbors of the latter. The input passed to the program, were: the weighted adjacency matrix, the nodes' labels and grammatical category, the label of the starting node. An additional function was added: for every level it is possible to choose the neighborhood should be extracted as is, or if from the neighborhood should be extracted only the nodes belonging to a fixed grammatical category, with the possibility to choose nouns and verbs together.

The software's output consists of three pajek files containing the subgraphs corresponding to the three neighborhood levels.

In figure 4.2, the three neighborhood levels for the word PLANET are shown.



(a) First level of neighborhood.

(b) Second level of neighborhood.

(c) Third level of neighborhood.

Figure 4.2: The 3 neighborhood levels for the word PLANET.

In order to have a better view of these three levels of neighborhood, another Matlab function (kettle.m) was developed. This code plots the label of the starting node together with three subplots representing the labels of the neighbors at every level (see figure 4.3). With this code, a dynamic visualization of the neighborhoods was developed. Even though the x axis has no meaning, being only necessary for a more comprehensible view, the system visualizes the neighborhoods' growth. On the y axis of the first subplot, the weight of the arc linking the starting node with every neighbor is measured. At each step of the simulation, in the second subplot, one neighbor for each of the nodes appeared in the first one is shown, with the y axis measuring the sum of the weights of the arcs going from the nodes in the previous subplot to the nodes in the second one. The third subplot follow the second one, but it shows the neighbors of the nodes in the second subplot. The number of nodes in the neighborhoods is usually increasing going from the first to the third level, but many of the neighbors shows a little instrength value: the visualization of these terms is poor because they result overlapped in the plot.

Another of visualization was implemented, following JTRACE (Strauss, Harris, & Magnuson, 2007) idea: the cumulative strength of the arcs between the word "PLANET" and the words belonging to its first neighborhood versus the number of nodes in the third level neighborhood (figure 4.4).

Semantic Fields Features

To analyze the features of the word's semantic field in the whole network, simulations with a high number of random walkers, for a high number of time steps and with several starting nodes were run (figure 4.5). All these simulations gave quantitative measurements on how the words are linked to various parts of the network through different paths with different lengths, but gave no information about semantic fields organizational features or about subjects' mental lexicon and cognitive processes.

For example, the semantic field $C_{50}^{50}(WATER)$ (figure 4.6) was obtained col-

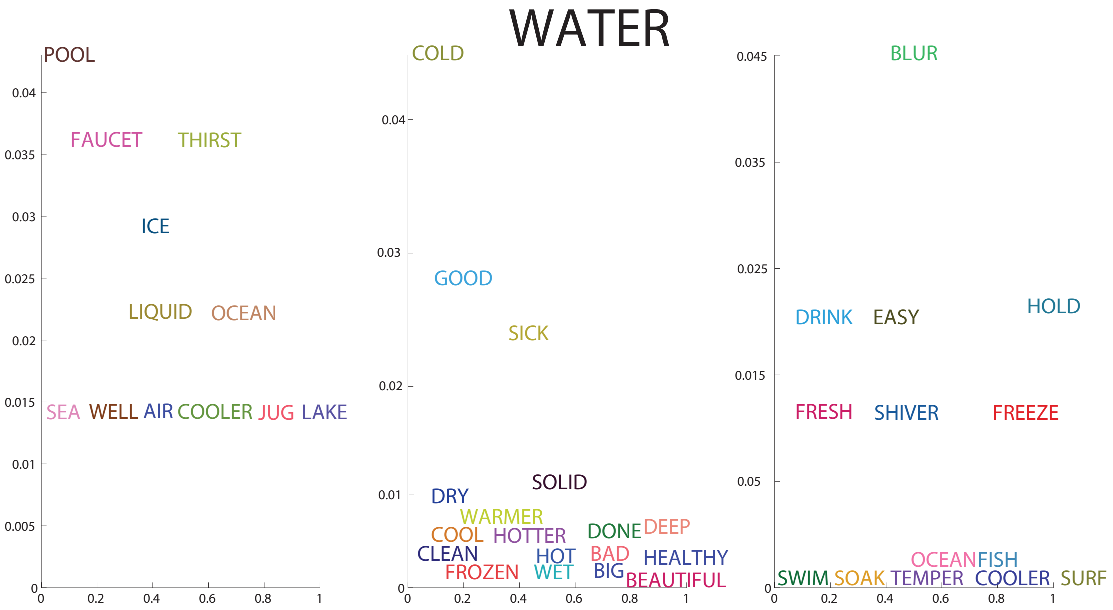


Figure 4.3: Representation of the implemented system. From left to right, the nouns in the first, the second and the third neighborhoods of the word "WATER". The higher a word is, the higher is the strength of the path leading to it from the word "WATER" ("POOL" in the first neighborhood, "COLD" in the second, "BLUR" in the third one). The second and the third neighborhood are not entirely visualized because of the high number of words with a low strength value: the simulation was stopped after a few steps to allow a good visualization of this example.

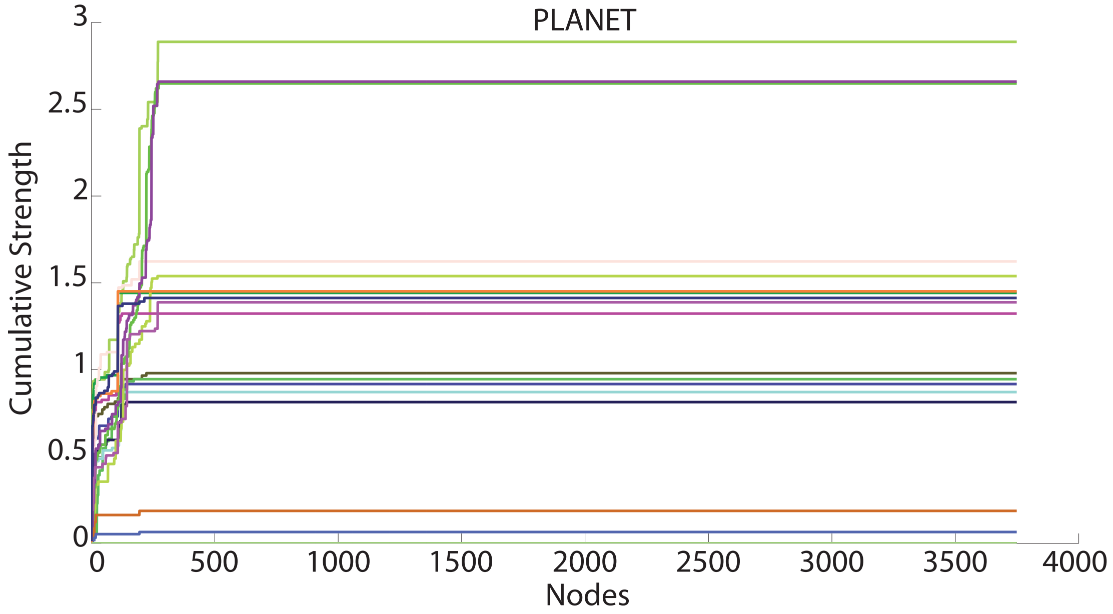


Figure 4.4: Cumulative strength of the first neighbors versus number of nodes in the third level neighborhood of the word PLANET.

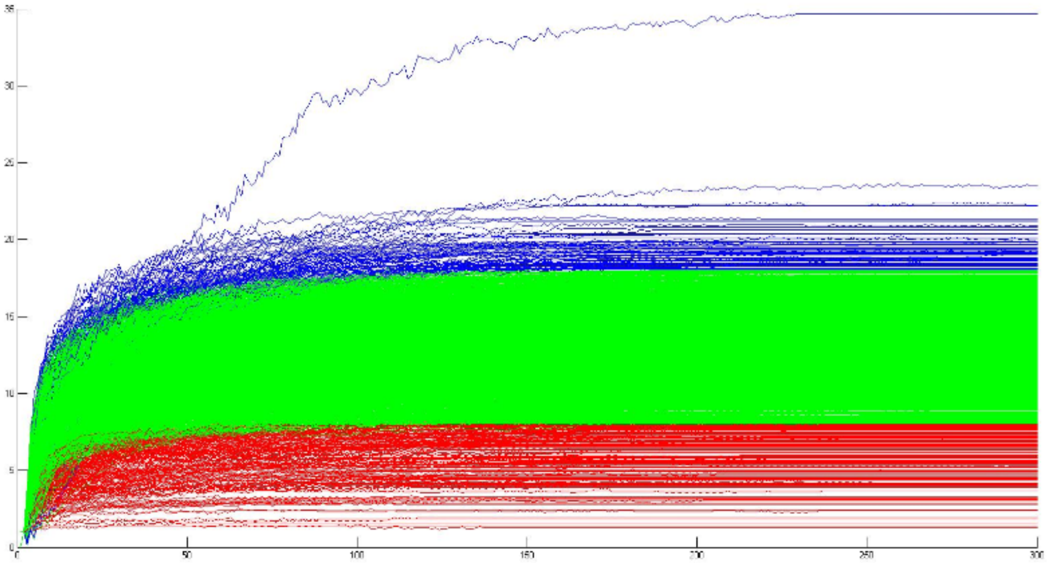


Figure 4.5: 100 random walkers, travelling in the network for 300 steps. On the x axis the steps are measured, while on the y axis the mean square distance from the starting node. The paths the walkers took, can be divided into three big clusters, according to the distance reached from the starting node.

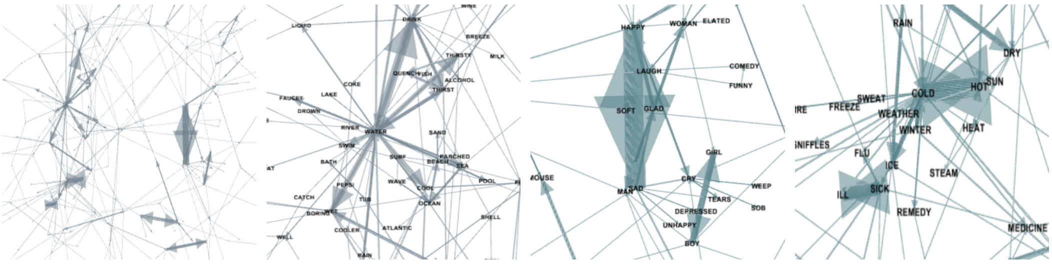


Figure 4.6: $C_{50}^{50}(WATER)$ network (first panel), composed of 429 and 719 arcs, with some significant details (second, third and fourth panels).

lecting all the paths of the 50 walkers during 50 steps and weighting the arcs with the number of times it was walked through and assigning the nuber of visits to each node. This network had 429 nodes and 719 directed arcs, with $\bar{k} = 1.767$, $d = 37$ and $C = 0.081$ (very low). The arcs with a high weight, were present in both directions.

Several communities are found in this network (figure 4.7).

In the presented model, the network extracted by the random walkers are not deterministic (figure 4.8), but are useful to statistically analyze the results.

The most visited nodes and the most traveled arcs belongs to the intersection. Thus, a core of terms and relationships, exists for every term, independently on

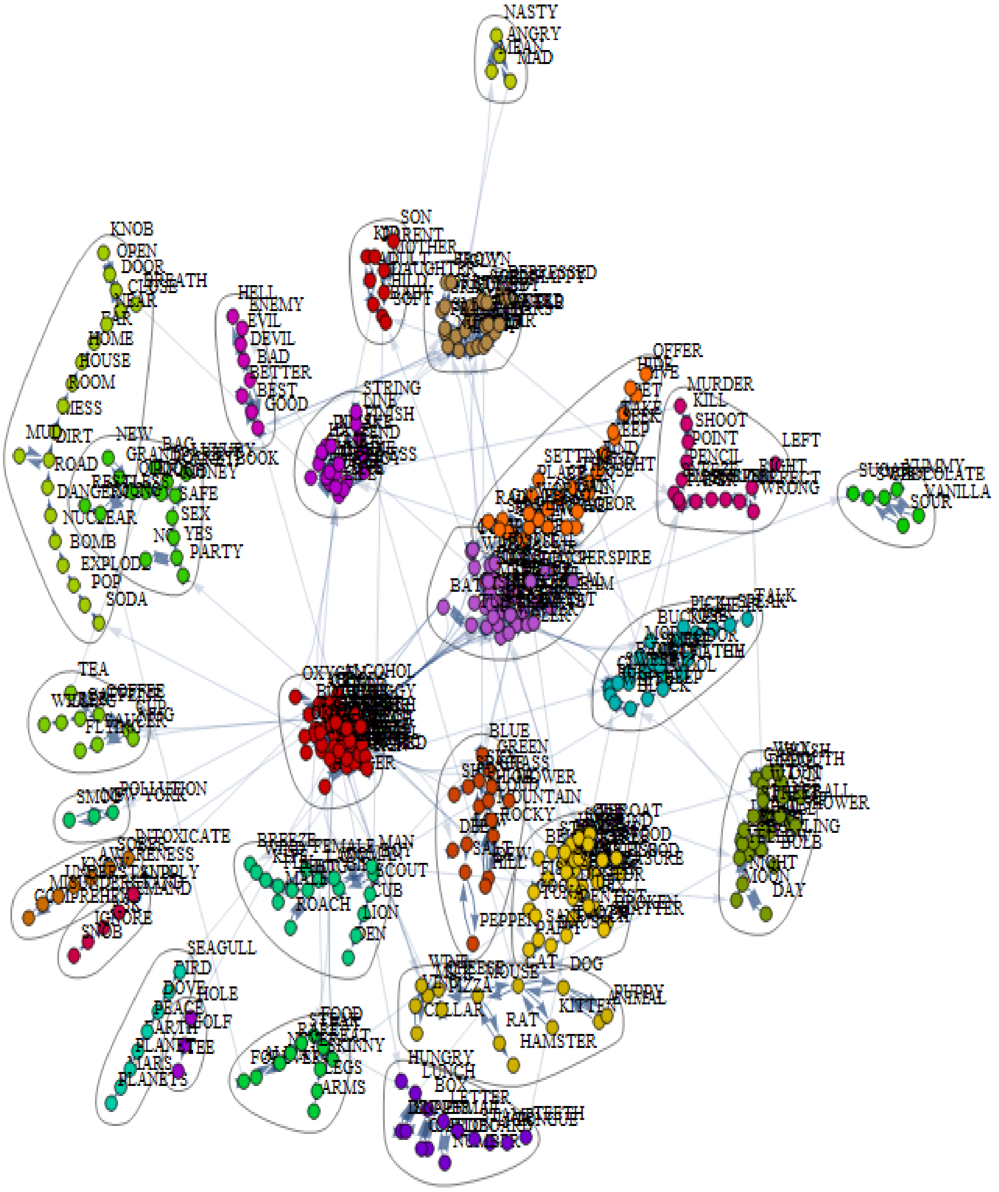


Figure 4.7: $C_{50}^{50}(WATER)$ communities. The largest community, in red, contains the starting node: WATER.

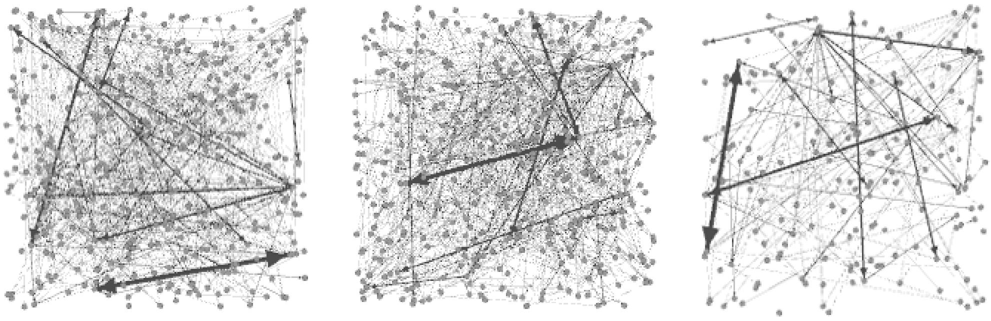


Figure 4.8: In the first two panels, the network $C_{50}^{50}(WATER)$ extracted in two different runs. The firs one has 429 nodes and 719 arcs, while the second 448 nodes and 779 arcs. In the third panel, their intersection, with 192 nodes and 201 arcs, is shown.

the path chose by the walkers.

Analysis

Random walks allowed the extraction of a set of words reachable by a fixed number of steps took in the network along the arcs. A deeper analysis highlighted the presence of communities and subcommunities.

Simulation runs with a low number of random walkers gave interesting results, from a cognitive perspective. As a matter of fact, increasing the number of steps and decreasing the number of walkers, the extracted semantic fields changed. It was possible to change semantic field, just by tuning these two parameters. As expected, the gradualism or organization concept, according to a Wittgensteinian approach, was not realistic due to the possibility of the subjects to explore and navigate different semantic fields, even with a script constraint.

They analysis of the experimental data of the network $\mathcal{N}$ brought light to three main phenomena.

Words association with a slow gradualism (Wittgenstein hypothesis), was rejected. The data shown an easy transition between different semantic fields. Further investigations are required, but gradualism appears to be valid only at a narrow local scale.

The data structure extracted by the random walkers, was fractal-like, with communities and subcommunities. This fractal structure provides cognitive markers like in the tip of the tongue phenomenon (Aitchison, 2012): when a needed word is not found, the recall process continues to wander around a set of words which recall the needed one.

During random walks runs, cognitive organization grows as a global structure, changing semantic fields during the evolution. The different semantic fields, forming communities, bring forth a trace of a possible small world structure which allows to easily access distant and different areas in the network.

4.4 Different Random Walk Models

To better analyze the network obtained from experiments and from Free Association Norms database, three different random walk models were developed. Starting from the results of the previous simulation runs and on the results of the analysis presented in chapter 3, in order to model organization, accessing and retrieval cognitive processes, different starting nodes and different networks ($\mathcal{N}$ and $\mathcal{R}$) were selected. As said before, random walk is a possible, simple and general model to reproduce searching in human mental lexicon.

The three random walk models, vary only in the bias. An unbiased random walk, is the one presented in subsection 4.3. A biased random walk, take into consideration a measure (bias) to prefer a path instead of another. The difference is that the probability to choose a specific node is not uniform on all the reachable nodes, but it is proportional to the bias.

The three models are: Unbiased Random Walk (URW), Page Rank Biased Random Walk (PRBRW), and Typicality Score Biased Random Walk (TSBRW).

URW was already defined before and is based on the network topology only. Thus, selected a starting node, the process goes on selecting one node. Every time a node is selected as the next one, the process can be seen as another instance of a URW process starting from the newly selected node.

PRBRW is biased on Page Rank (Page et al., 1998), computed in advance for every node of the networks. The basic behavior of the walker is the same: from a starting node, the walker will move to one of its neighbors as the first step and so on. The probability to choose a fixed node among the neighbors is the difference from the previous model. This probability is no longer uniform but proportional to the Page Rank of the neighbors: $P(n_j|n_i) = PR(n_j)/\sum_k PR(n_k)$, where $PR(n_j)$ is the Page Rank of the node n_j and the sum runs over all the neighbors of the node n_i. Obviously, if the network is directed, the neighborhood of n_i is considered composed by the nodes with an incoming arc from n_i.

TSBRW is similar to PRBRW, but the probability to choose the next node is proportional to the typicality score of the nodes collected during the experiment. In this model, the walker should be able to distinguish the borders of a script in the network during the simulation run. Additionally, when the walker steps on a typical node of one of the scripts created during the experiment 1, the scripts to which the node belongs to is recognized, and an inverse spreading algorithm is started: every node in the script pointing to the node on which the walker is, receive a fraction of its score and every one of them pass a fraction of their score to every other node pointing to them. Every node which received an additional score cannot receive additional scores anymore. In this way, the additional score is inversely proportional to the distance from the current node. Thus, the random walker recognize the boundaries of the scripts it is traveling through and tend to choose nodes belonging to these scripts.

Simulation Runs

400 nodes were randomly selected for the simulations. Starting from every selected node, 100 random walkers, for every one of the three models, run for 3000 steps. For each of the 400 starting nodes, the squared distance from the starting node, averaged over the 100 walkers, was computed. After a few hundred steps, the average squared distance, stabilized around an equilibrium point. Thus 3000 steps were too much, and were reduced to 500, while the number of starting nodes was increased: all of the nodes were selected as a starting node.

Results

a. Results concerning the first computational model showed that, after a transition period of just about a hundred steps, the trajectories of the random walkers settled down around a certain distance, with small oscillations. Trends of distances in time were represented in figure 4.9. Typical random walkers behavior, colored in blue, was related to almost all the nodes. A small number of green curves (FURTIVELY, HIGH RISK AREAS, HOUSE-

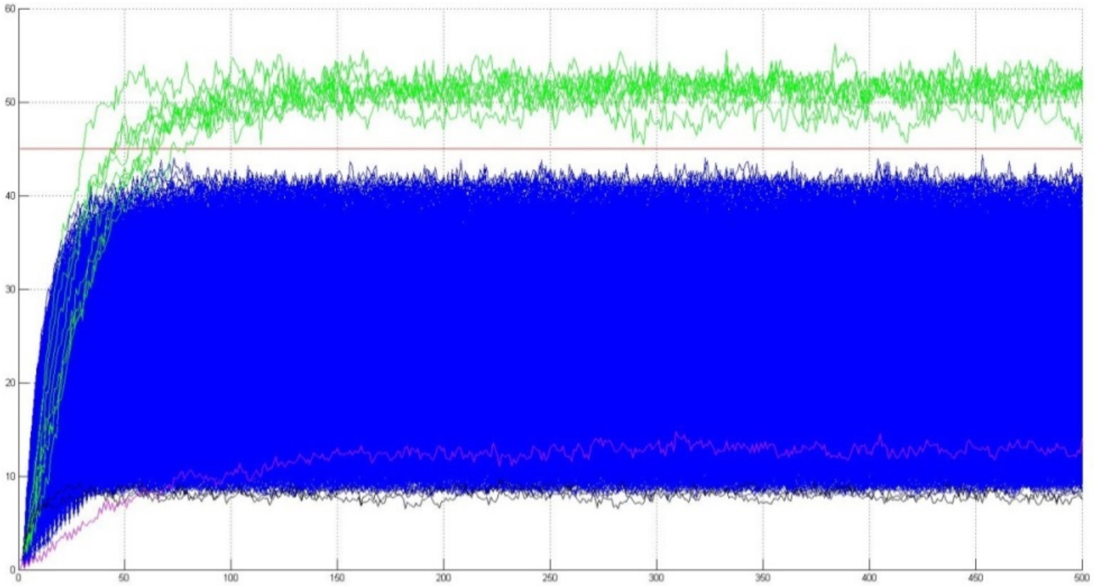


Figure 4.9: Quadratic distances curves traveled by random walkers, averaged on the number of walkers. The blue curves identified the most typical behavior of the random walkers. The x axis represented the time steps, while the y axis represented the quadratic distance.

WIFE, LADYS MAID, PRIVATE AREAS, SEISMIC ZONES, RAISE THE VOICE, WORKWOMAN), showed a remarkable speed of the random walkers in leaving the initial node to reach a long distance from it. Black curves (PAY , GOING TO THE ZOO), faintly visible at the bottom of the diagram, represented the trajectories of random walkers who remained closest to the starting node. The magenta curve (HORSE) showed the longest period of transition, as it traveled along a huge number of hyponyms. In order to understand degree-based network, paths were represented as tree diagrams (See figure 4.10). In this structure, each node was connected with its descendants by bidirectional arcs, since the network was undirected. Branch lines were nearly all regular, except in cases where a node, visited not less than two steps before, is visited again. In this case, the arc was represented with a coming back line (like from JACKET to DRESS in figure 4.10). As said before, the random walker starting from the node "HORSE" developed an interesting tree-diagram in the space of the possible connections (See figure 4.11), due to the large number of hyponyms, who constituted the majority

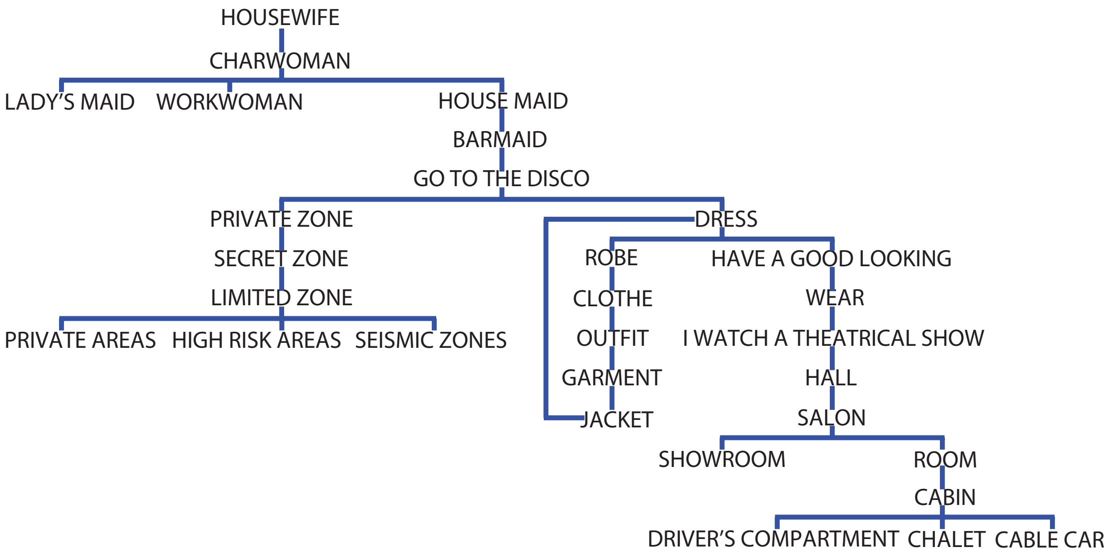


Figure 4.10: Diagram of the first 100 steps of a random walk started from node HOUSEWIFE. As it is possible to see, the node DRESS presents both outgoing and coming back lines.

of its neighbors, which made difficult for the system to leave the starting node. Interestingly, while the other random walkers collected nouns and verbs, random walkers started from the node PAY (figure 4.12), collected a huge number of adverbs in the first part of its path.

b. Page Rank system results produced trajectories with the same trends of figure 4.9. Also for this algorithm, simulation runs were stopped at 500 steps, this number being considered useful to represent available paths inside the network. The analysis on Page Rank simulations gave similar results (figure 4.13): a dense blue band of trajectories, mixed with each other, with 8 green trajectories that are at a greater distance without continuity. Surprisingly these 8 trajectories differed from others found with degree-based random walkers, even if they started from the same nodes. At the bottom of figure 4.13, the four trajectories in black, starting from the nodes BOOK, GOING TO THE ZOO, PAY, TAKE A TRIP ON BICYCLE WITH MY FRIENDS), presented trajectories whose length were lower than in the previous simulation. Page Rank tree-diagrams (figures 4.14, 4.15, 4.16) showed interesting patterns. In fact, if compared to the diagram of figure 4.10, ob-

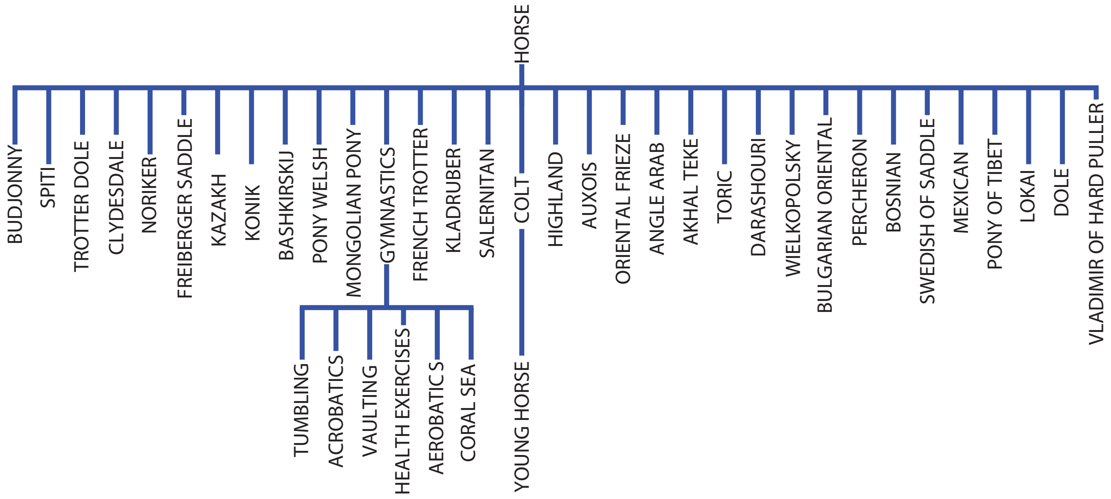


Figure 4.11: The HORSE node presented many hyponyms with degree equal to 1. This allowed the walker to visit a collection of nodes very close to the starting node.

tained by the degree-based simulation run, the random walker starting from the node "HOUSEWIFE" (figure 4.14), tough producing long trajectories, returned back several times. This confirmed that Page Rank based curves reached shorter average quadratic distances than those with degree-based simulation. The random walker starting from the node "PAY" produced a short distance trajectory, (figure 4.15) as well, whereas Page Rank system browsed a higher portion of the network, and many nodes were visited more than once. Page-rank-based random walk changes in the tree diagrams were much more evident (see for example figure 4.16), with slower and longer transients and a chain of words not strictly connected with one semantic field, but with a huge number of highly page-ranked words. Loops are also present in great number.

c. Results based on typicality scores random walking were qualitatively and quantitatively different from the first two runs (figure 4.17). The aim was to understand if a random walking that took into account the nodes individually visited, could collect the script-based organization, moving on related nodes in the same script. 2327 trajectories, with high rate of returns to the starting node, and very slow in leaving its neighbor, represented less

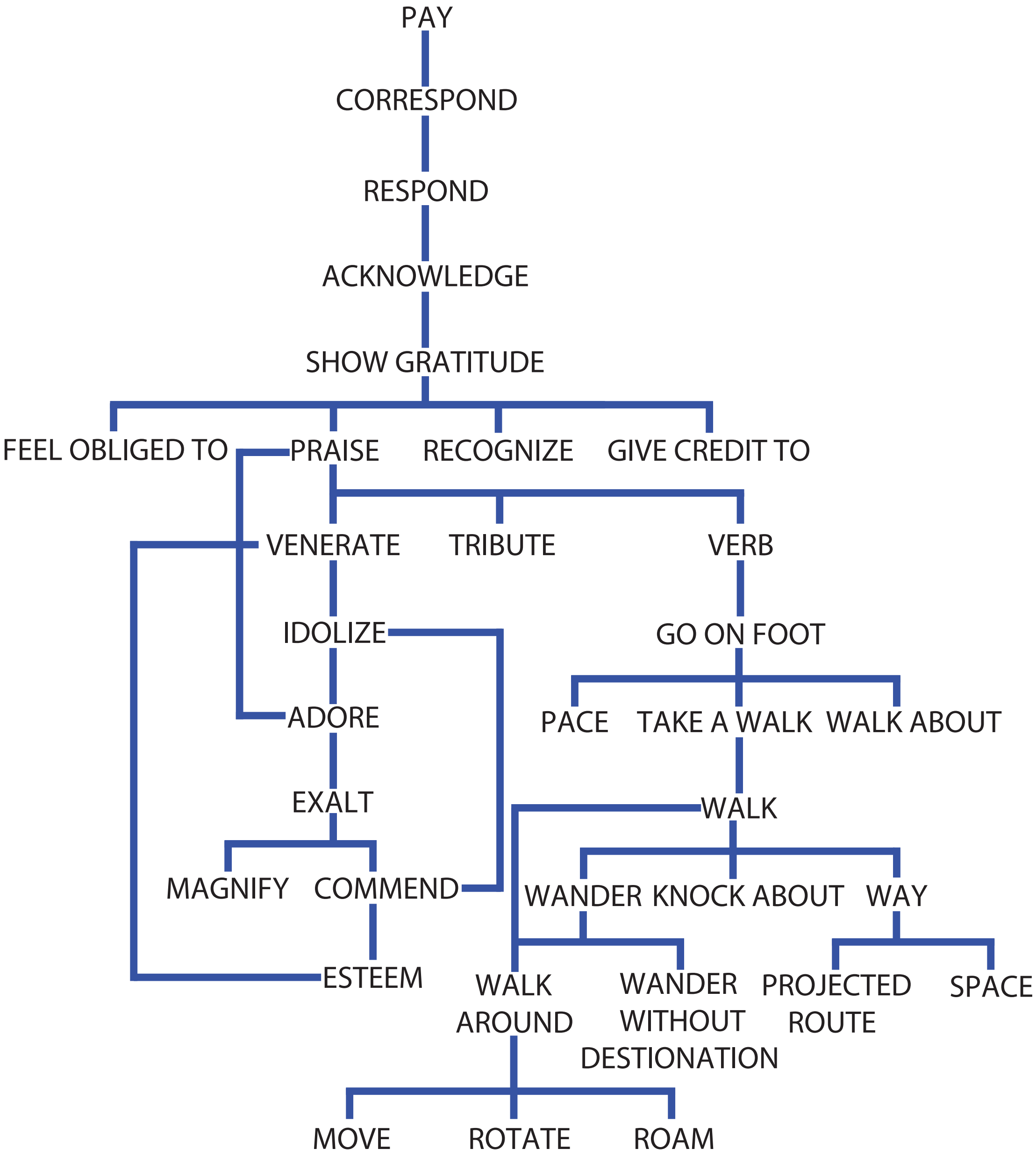


Figure 4.12: Diagram of the random walker started from node PAY.

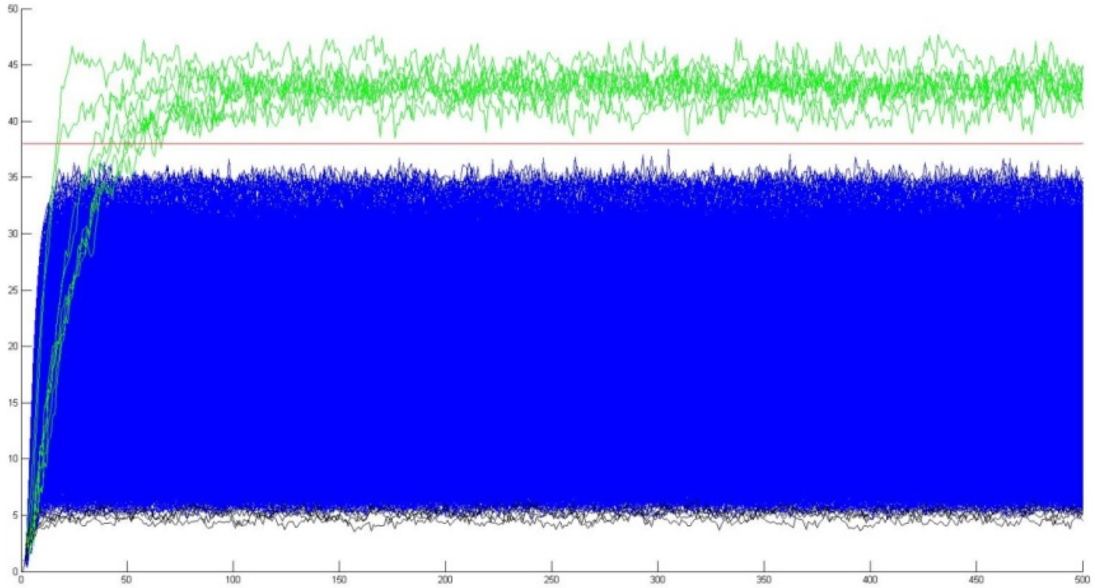


Figure 4.13: Quadratic distances curves traveled by random walkers, averaged on the number of walkers, with the algorithm based on Page Rank. The 8 curves in green were separated from the rest of the sample, clearly showing a sudden and oscillating trend. Instead, black curves on the bottom are related to paths that were closest to the starting node.

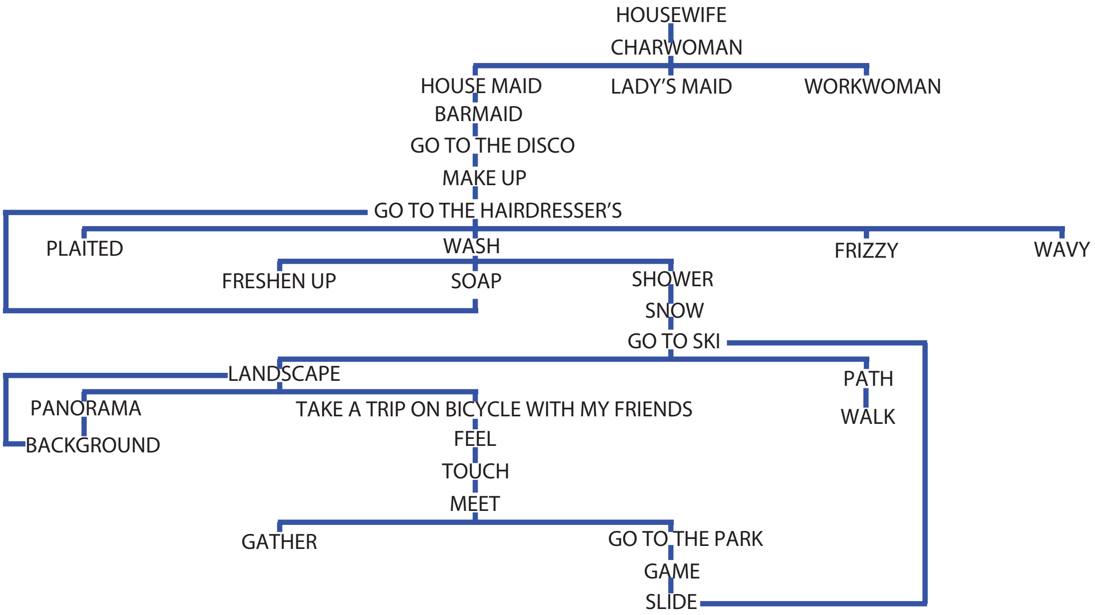


Figure 4.14: HOUSEWIFE tree-diagram based on Page Rank model presented a more complex structure.

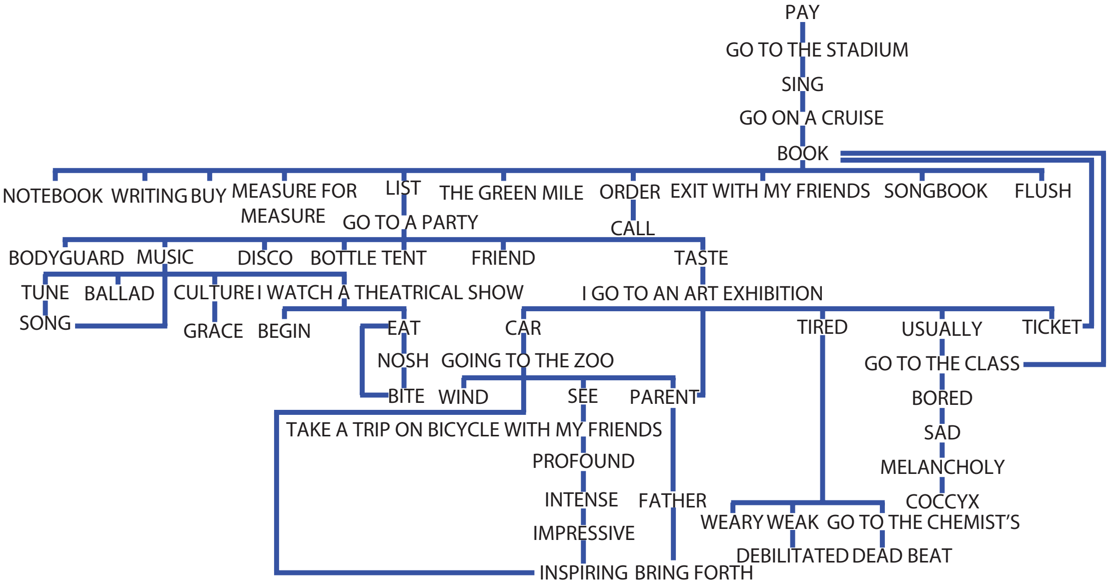


Figure 4.15: Diagram of a walker started from node PAY with the algorithm based on Page Rank.

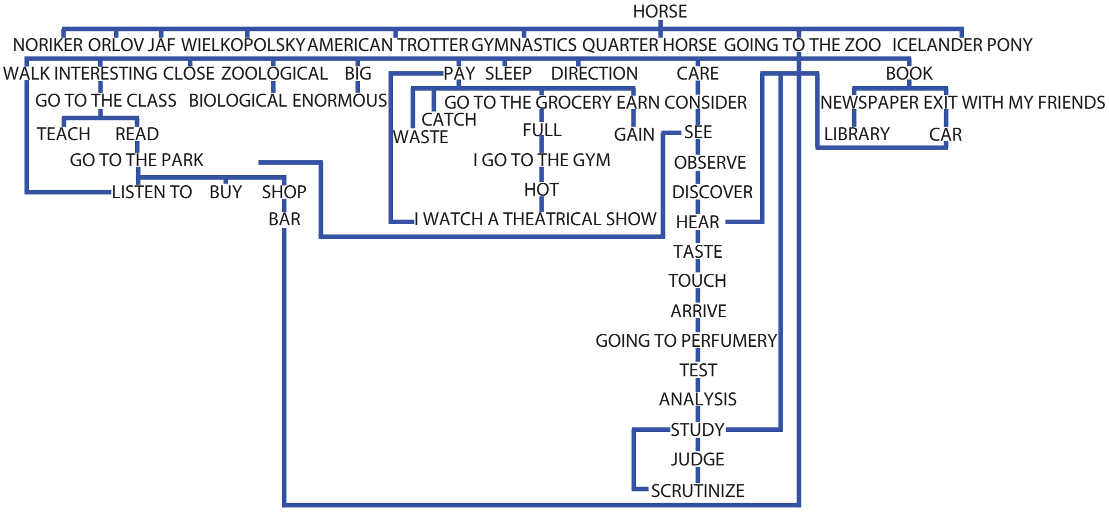


Figure 4.16: Diagram of a walker started from node HORSE following Page Rank based algorithm.

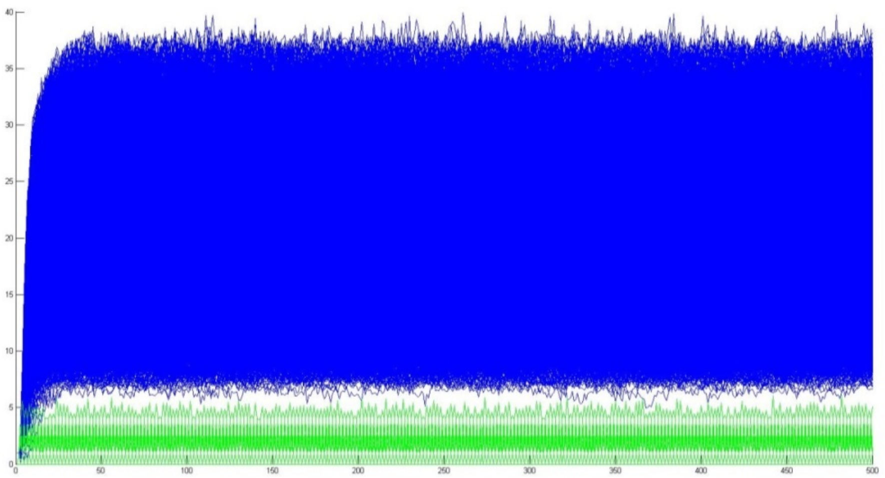


Figure 4.17: Quadratic distances curves traveled by random walkers, averaged on the number of walkers, with the typicality scores based algorithm. The trajectory organization was very different, according to different behaviors in this simulation run.

than 15% of the total trajectories (green color). The other trajectories were condensed in the blue band, with characteristic behavior very similar to the previous two runs, with a rapid initial increase and then a short oscillating goings-on. Diagrams presented loops and adherence to the scripts the initial node belonged to. Figures 4.18, 4.19 and 4.20 showed the diagrams of three different simulations, with this last method. As it is possible to see comparing figure 4.12 with figure 4.5 not only the method is important. In fact, the networks explored were different but the method to simulate word retrieval was the same, but the results were completely different. This brings to the conclusion that the topology, the structure and the organization of the mental lexicon is important in cognitive processes as well. For example, the random walk method starting from a node recognized a certain subnetwork in which terms were collected and, again, this subnetwork strongly depends on the topology of the global network.

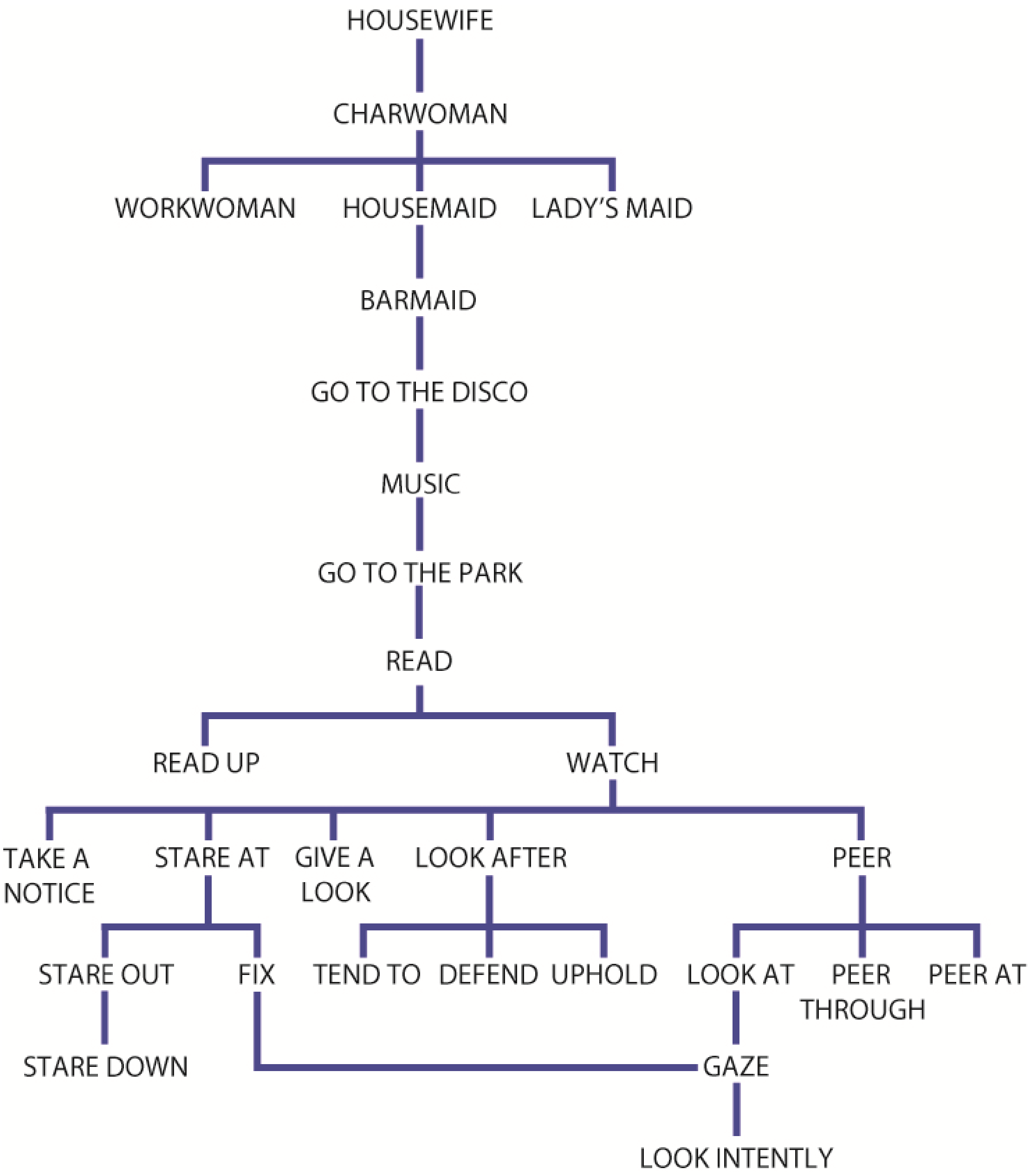


Figure 4.18: Diagram of a 100 steps simulation of a random walker started from the node HOUSEWIFE with the typicality and inverse spreading based algorithm. In this case the walker struggled to stay on a script, because he found nodes that acted as a bridge among the initial nodes of two different scripts: The node were connecting GO TO THE DISCO MUSIC and GO TO THE PARK.

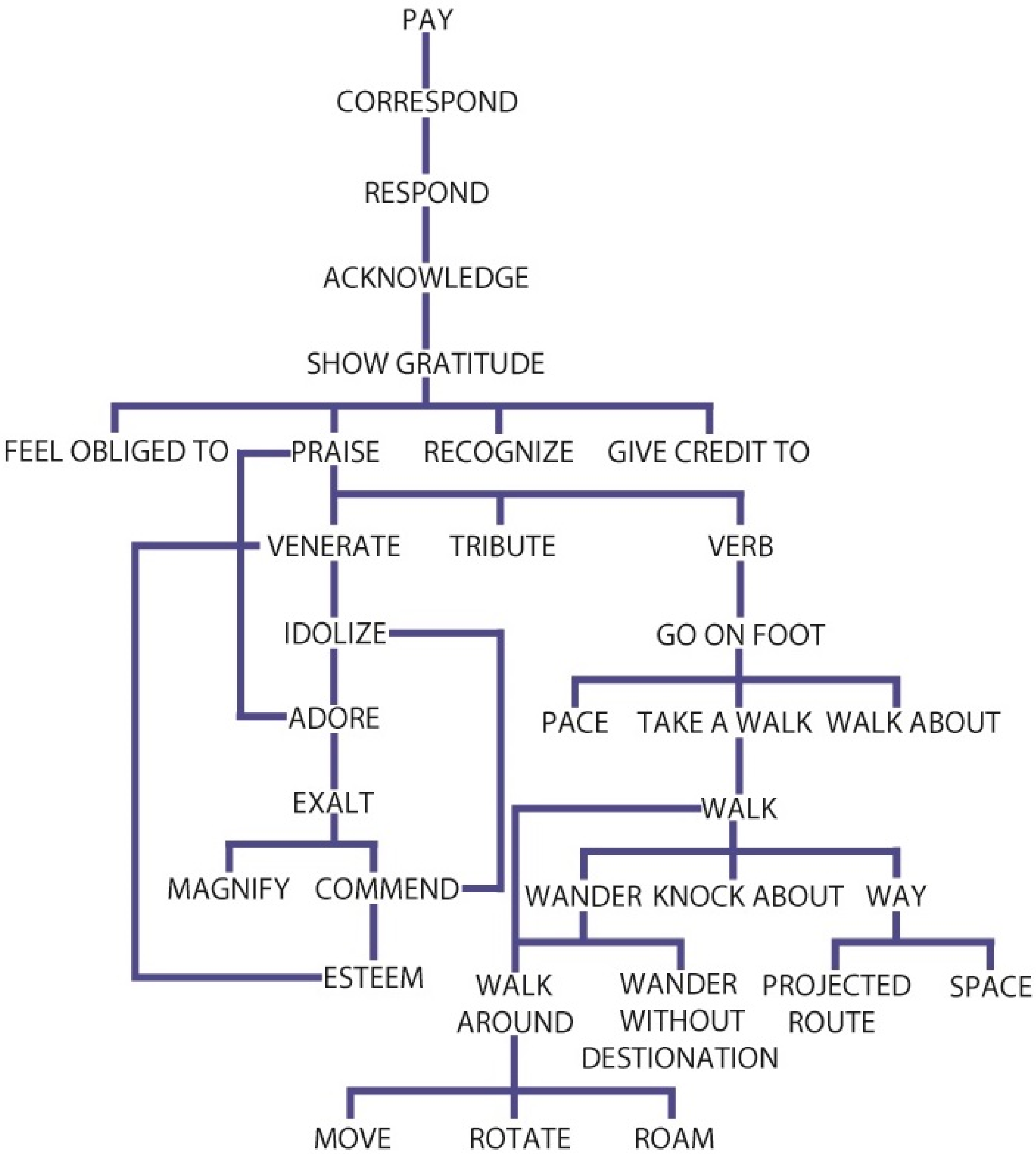


Figure 4.19: Diagram of a 100 steps simulation of a random walker started from the node PAY, with the typicality and inverse spreading based algorithm. The random walk found himself in an attractive area in which there was the verb PRAISE; then it got out and entered that of the verb WALK. The random walk remained over again on both verbs.

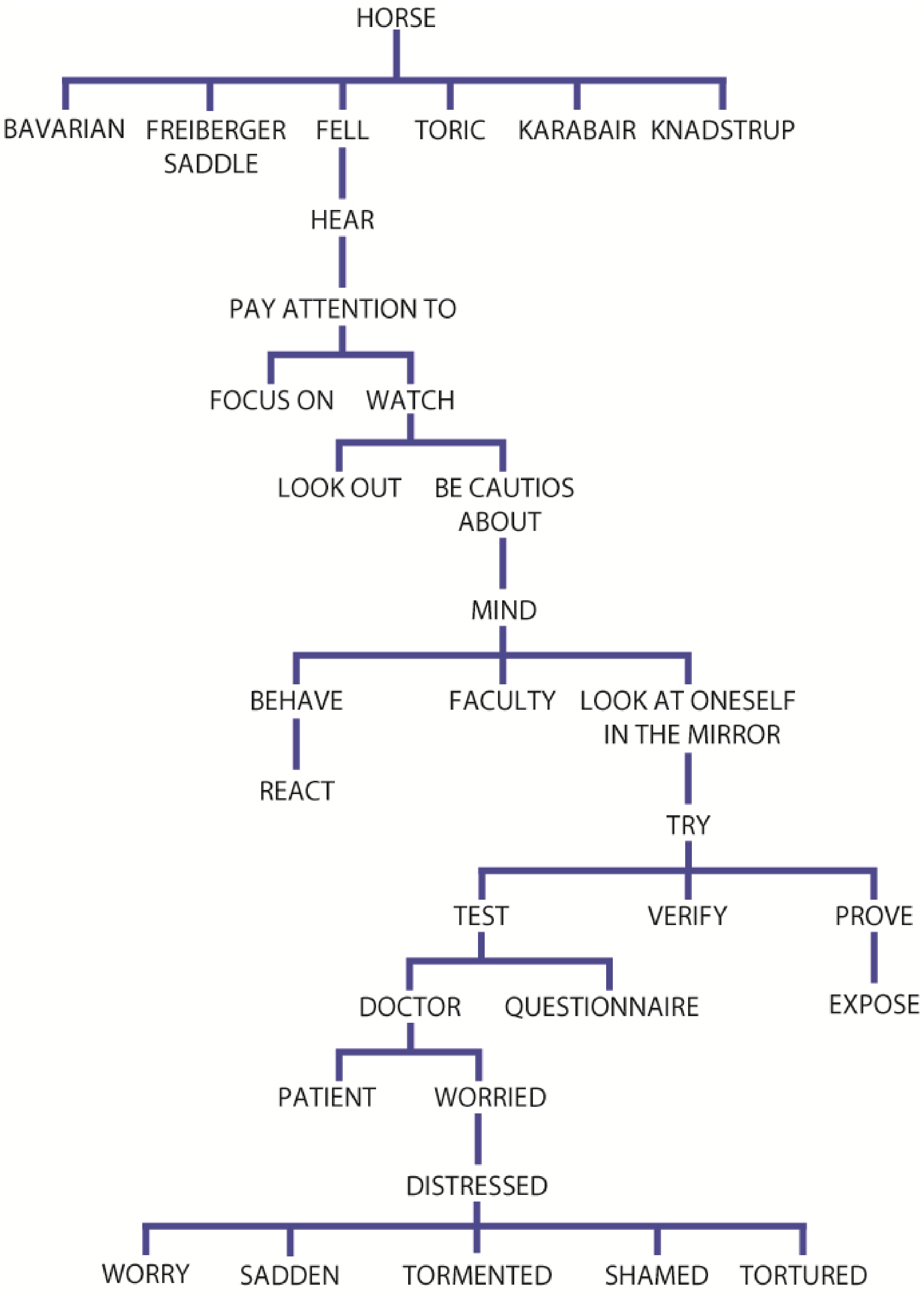
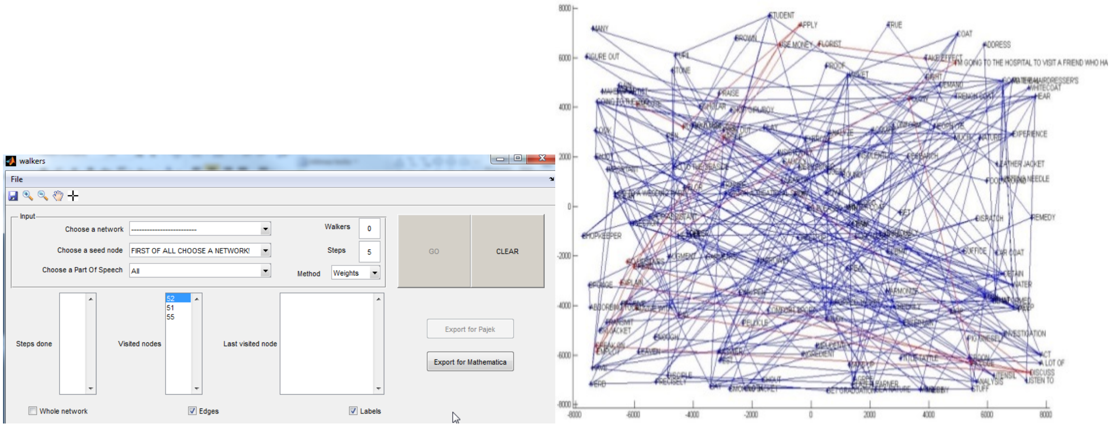


Figure 4.20: Diagram of a 100 steps simulation of a random walker started from the node HORSE, with the typicality and inverse spreading based algorithm. As in the simulations with other algorithms, the initial node had many hypernyms, but the random walker came out of that group, quickly getting away from the initial node.

The computational system

To be able to interactively simulate the developed random walk methods on language networks a Matlab interface was created (figure 4.21). The system allows for multiple runs of simulation and for a collection of the data in a graph (figure 4.22). In this display, nodes were represented as spheres of different color. Red and blue colors were used to represent the nodes of the two networks, while green nodes were the nodes in common between the two networks. The 3D graphic had nodes arranged on the radius of a circle, equal to the distance from the starting node, having a height and an intensity of color, directly proportional to the number of times they had been visited by the random walkers. The system also allows to create text files with a syntax well-suited for Mathematica, or Pajek files. The first drop-down menu allows the uploading of any kind of network, in Pajek format, or to choose from 4 presented networks. In particular, the Free Association Norms network, consisting of about 10,000 nodes; the experiment network, consisting of about 17,000 nodes; the two possible intersection networks between previous two networks. The second drop-down menu allows to choose the node, from which starting the simulation run. The third drop-down menu allows to select which grammatical category to be considered. The number of random walkers,



walkers

Input
Choose a network: -------------------------
Choose a seed node: FIRST OF ALL CHOOSE A NETWORK!
Choose a Part Of Speech: All
Walkers: 0
Steps: 5
Method: Weights

Steps done
Visited nodes: 52 51 55
Last visited node

□ Whole network
☑ Edges
☑ Labels

(a) (b)

Figure 4.21: A Matlab program has been developed for the interactive exploration of a network using random walk. On the a), it is possible to see the interface of the system, on b), the result of a simulation run.

the number of steps of the simulation, one of the developed models (Weight, Page rank and Typicality based Random Walk) can also be defined at each run of the simulation.

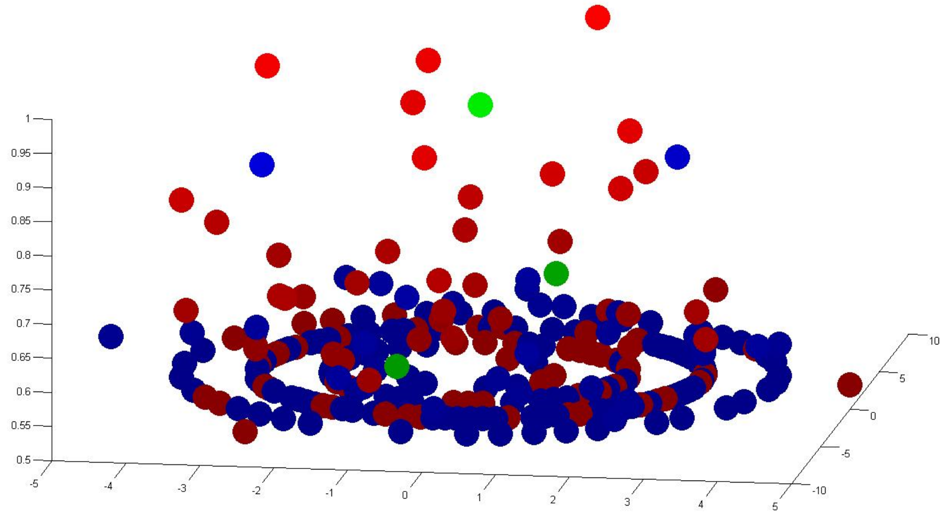


Figure 4.22: Graph for comparison of paths on different networks with nodes in common.

4.5 Conclusions

In this chapter methods to investigate semantic networks at an intermediate level were presented. After the building of networks from experiments, using a random walk approach neighborhoods were extracted and analyzed. A modularity maximization algorithm to search for communities was applied, having a different result from what was expected. At the end 3 different random walk based methods were proposed to explore the networks and analyze interaction among nodes at the meso level.

At the meso level, the neighborhoods investigation and the analysis of the nodes up to three levels of depth allowed to investigate a wider neighborhood for each node. Being the average minimum path of the global network little more than 5, with 3 steps, following all of the arcs, a distance was achieved which did not include the whole network, but an intermediate level of the network centered on the selected node. Selecting the central node of a script the result will thus be that specific script (up to its third level of growth) plus pieces of other scripts dragged in by the synonyms in common, as predictable. When choosing a degree 1 node, as a hyponym or a hypernym, the result will be less than a script, given the small amount of nodes recollected. The first neighborhood level consisted in one single node, the second level included neighbors of the node (leaving out neighbors of the neighbors). The third neighborhood level was more complex in terms of dimension and portion of the explored network: most of the times it only included one script. The analysis of the neighborhoods of the scripts central nodes (instead of the initial ones) proved to be a more interesting method, together with the analysis of the nodes belonging to different scripts. As foreseen, the neighborhoods significantly changed depending on the node chosen as a start: these changes included the number of nodes, the number of arcs and the number of scripts explored. An empiric analysis on a selected sample of nodes demonstrated that the number of nodes, arcs and visited scripts rose when choosing a higher degree node or a higher clustering coefficient one. Consequently, choosing nodes belonging to more than one script and having a clustering coefficient above the average it was possible to obtain networks which are wider and included more parts of different scripts. Hence, presuming that the Mental Lexicon is consistent with the model of the network created, this study suggests that, when dealing with a specific term (such as a hyponym or a hypernym), people tend to stay inside the neighborhood of that term. This neighborhood includes many semantic areas, if the starting term has different meanings or is used in different contexts.

The various kinds of random walks used to navigate the network gave different results, but sharing some traits. The mean square distance covered by the walkers was the measure taken into consideration. It was noticed that, in every procedure, the walkers rapidly went away from the initial node and then stayed at a certain oscillating distance. After around 100 steps, these average values started to oscillate around a value depending on the starting node, covering a wide range of values. These characteristics are common to all the algorithms on which the random walks were based.

The simulation using the classic random walk on a graph and the one using pagerank values distribution of the nodes showed another common feature: the existence of 8 nodes that, if chosen as starting point, pushed the walkers to greater distances. This happened because pagerank was a centrality measure of the nodes, based exclusively on the topology of the network, the same as the degree distribution of the nodes, at the base of classic random walks. All of these 8 nodes were less central, had a higher mean distance from the other nodes and had degree 1 (as HOUSEWIFE, see figure 4.13). On the contrary, nodes with a high centrality or linked to many nodes with degree 1 (as HORSE, which had many hyponyms), tended to keep the walker closer, trapping it inside the exploration of a great amount of nodes. Once finished the exploration, the walker's only alternative was to return to the starting node, thus slowing down its wandering.

Unlike these two algorithms, the one based on typicality as a weight for the network produced many trajectories which repeatedly and frequently returned on the initial node. These attractor nodes showed a very high typicality score compared to the other nodes close to them. The choice of the next node to be visited was influenced by the neighbor nodes' typicality, thus giving the walker a very low possibility to move away, instead, the curves arriving to average distances were the ones inside neighborhood of nodes with high or average typicality: starting the walk near the initial node of a script it is more probable to encounter nodes with a similar level of typicality, giving the walker a wide range of possible paths. The walkers moving farthermost were those starting from areas of the network with low typicality. In these cases, the walker based its choices only on the degree of the nodes until it arrived to areas with higher typicality, where it was forcedly attracted and consequently blocked, especially when walking inside smaller areas with a higher typicality than the neighborhood. Being impossible to make a thorough analysis with the algorithm based on typicality scoring in order to activate a process of inverted spreading, analysis on various samples of nodes were realized. These simulations provided interesting information for the evaluation of the model. For most of the simulations started from non-peripheral nodes, few steps were needed (around 50) for the walkers to meet a node and allow the inverted spreading algorithm to recognize one or more scripts. In this way, the walkers followed paths staying inside one script. When the walkers started from a borderland between various scripts, (especially from nodes which were typical for many scripts), the number of steps needed to recognize a unique script varied a lot in the simulations, and the walker was not able to immediately settle in one area, but wandered through different scripts. Instead, when starting from a node which was typical for only one script and also topologically central inside the subnet composed by that script, the walker immediately settled in that networks, having few possibilities to move away. At the moment the walkers recognized a script and found themselves in a high typicality area of that particular network, the distance from the initial node started to stabilize. The study also showed that the temporal sequence of the typical actions of a script (such as: ENTER before EXIT) was not followed and strongly depended on the starting node and the casualty of the walk, even if limited by typicality: in this last model based on typicality, the walker was attracted by very typical nodes which were located close to it and separated by other very typical nodes. In conclusion, the last model among the ones implemented, proved to be the best to simulate the ability of a subject to identify the semantic area, the context and the topic of the term represented by the selected starting node. All these methods, having the characteristic to explore the meso level and the interaction among the terms, highlighting the bonds more than the nodes of the network, gave us a knowledge on human mental lexicon deeper than how it was before.

Chapter 5

Conclusions

The analysis of the networks built with the data collected during word production experiments validate all the existing theoretical models and add something more. In fact, word production modelled by this kind of network exhibits the same behaviour supposed by spreading activation (Collins & Loftus, 1975), while the process to build this network does not follow the spreading activation paradigm. The network created confirm all the other properties (small world, scale free, etc...) observed on linguistics and semantic networks in literature. Thus, this model of human mental lexicon development represents an alternative to the classical ones and have the merit to show clearly, not only the micro and macro levels, but also the meso level, which shows very important features about the processes behind the language production and development.

While the classical random walk on a network is a naive and, maybe, the simplest possible model of language production when a word collection is given, it paved the way to another model: the typicality based random walk. When biased by the typicality score of the terms in the network, the random walker gains a great knowledge of the network's meaning. It reacts faster than the other models analysed (unbiased and pagerank based) when it is asked to identify the main topic or to choose the successive words according to the situation. This better behaviour of the model proposed is due to the switch of influence from topologically central nodes to the semantically central ones derived by the introduction of the typicality as a score. Even if this model is still unrefined, it gives a prototype for an automatic language production process: the sentences created are not perfect but the walker succeed to identify the right semantic area without exiting from it.

Chapter 6

Appendix: Script tables and degree distribution fits

Table 6.1: Word generation for the 39 scripted activities. AJ=adjectives, AD=Adverbs, Conj=Conjunctions, Det=Determiners, N=Nouns, V=Verbs. The table reports the number of terms, belonging to each grammatical category, collected during the first experiment, as well as the ratio between the number of nouns and the number of verbs.

Script Name	AJ	AD	Conj	Det	N	V	N/V
Go to the bookshop	33	14	1	0	76	95	0.8
Go to the airport	62	39	3	1	119	110	1.081818182
Go to the beach	22	16	0	0	68	79	0.860759494
Go to the shopping center	51	23	2	0	100	88	1.136363636
Go to the stadium	95	23	0	0	288	267	1.078651685
Go out with my dog	35	12	0	0	57	81	0.703703704
Go to ski	28	6	0	0	217	147	1.476190476
Make a constructive experience abroad	122	41	1	0	177	268	0.660447761
Go to the hospital to visit a friend with a broken arm	28	20	0	0	55	72	0.763888889
Go to an art exhibition	122	59	0	0	183	121	1.512396694

Go to the disco	46	15	0	0	70	87	0.804597701
Go to the zoo	52	15	0	1	240	200	1.2
Go to the chemist's	18	11	0	0	43	61	0.704918033
Exit with my friends	21	13	0	0	5	124	0.403225806
Go to the library	60	10	0	0	252	212	1.188679245
Go to work	28	30	0	0	145	117	1.239316239
To cook	16	18	0	0	52	109	0.47706422
Go to a party	26	18	0	1	193	99	1.949494949
Go to the doctor	67	27	0	1	75	111	0.675675676
Go to the hairdresser's	88	21	0	0	223	251	0.888446215
Go to the museum	34	15	0	0	71	104	0.682692308
Go to the class	127	65	0	2	161	156	1.032051282
Give private lessons	50	9	0	0	61	84	0.726190476
Go to the beautician	106	41	0	1	165	150	1.1
Horse riding	15	6	0	0	28	75	0.373333333
Go to the gym	200	24	0	0	181	176	1.028409091
Go to a wedding party	109	44	0	0	194	303	0.640264026
Clean the room	70	59	3	0	89	127	0.700787402
Go to the perfumery	65	30	0	2	138	135	1.022222222
Go to the grocery	143	106	4	0	153	145	1.055172414
Go on a cruise	40	7	0	0	182	189	0.962962963
Go to the theater	70	43	1	1	117	205	0.570731707
Trip on a bicycle with friends	46	25	0	1	114	208	0.548076923
Go to the seaside	8	0	0	0	126	85	1.482352941
Go to the dentist	97	15	0	0	166	190	0.873684211
Go to the birthday party	52	15	0	0	49	129	0.379844961

Go to the jeweler's shop	134	98	4	0	100	94	1.063829787
Go to the park	134	118	8	13	231	246	0.93902439
Go to the university	47	35	0	1	83	80	1.0375
TOTAL	2567	1186	27	25	5047	5580	0.9045

Table 6.2: Presentation of the mean scores on the typicality of items for every script. AJ=adjectives, AD=Adverbs, Conj=Conjunctions, Det=Determiners, N=Nouns, V=Verbs. Data have been obtained by averaging the sum of the typicality score of all the term in each grammatical category for each script and averaging over their number.

Script Name	AJ	AD	Conj	Det	N	V	Total
Go to the bookshop	2.788	2.5	2	0	2.776	2.75	2.880
Go to the airport	3.226	3.667	3.333	6	2.866	2.964	2.892
Go to the beach	3.409	3.688	0	0	3.191	3.487	3.419
Go to the shopping center	2.804	3	1	0	2.72	2.693	2.363
Go to the stadium	3.105	2.348	0	0	2.774	2.801	2.725
Go out with my dog	2.543	2.667	0	0	2.544	2.506	2.577
Go to ski	3.714	4.667	0	0	3.714	3.571	3.722
Make a constructive experience abroad	2.738	2.976	6	0	2.65	2.575	2.378
Go to the hospital to visit a friend with a broken arm	3.179	3.15	0	0	2.891	2.903	2.488
Go to an art exhibition	2.828	3.339	0	0	2.596	2.843	2.517
Go to the disco	2.87	3.8	0	0	2.743	2.736	2.361
Go to the zoo	2.654	3.067	0	1	2.579	2.66	2.507
Go to the chemist's	4	3.636	0	0	3.721	3.738	3.187
Exit with my friends	3.286	3.692	0	0	3.16	3.185	3.206
Go to the library	3.197	2.1	0	0	2.709	2.822	2.549

Go to work	2.536	2.567	0	0	2.596	2.615	2.518
To cook	3.438	3.611	0	0	3.654	3.541	3.494
Go to a party	3.5	3.333	0	4	3.021	3.242	2.889
Go to the doctor	2.851	2.889	0	3	2.84	2.874	2.726
Go to the hairdresser's	3.273	3.286	0	0	3.117	3.12	2.889
Go to the museum	3	3.533	0	0	3.113	3.096	2.864
Go to the class	3.039	2.892	0	1	3.068	3.128	3.194
Give private lessons	3.2	3.667	0	0	3.148	3.024	3.294
Go to the beautician	2.708	2.634	0	5	2.655	2.613	2.482
Horse riding	3.933	4.333	0	0	3.857	2.867	2.891
Go to the gym	2.743	3.542	0	0	2.785	2.773	2.791
Go to a wedding party	3.11	3.273	0	0	3.134	2.805	2.446
Clean the room	2.8	3.017	2.333	0	2.674	2.504	2.672
Go to the perfumery	3.369	3.433	0	3.5	3.246	3.267	2.769
Go to the grocery	2.105	2.142	1.25	0	2.072	2.097	2.078
Go on a cruise	3.05	2.857	0	0	2.863	2.857	2.680
Go to the theater	2.886	2.86	6	6	2.709	2.937	2.609
Trip on a bicycle with friends	2.457	2.44	0	1	2.539	2.356	2.386
Go to the seaside	3.875	0	0	0	3.317	3.024	3.342
Go to the dentist	3.01	3.733	0	0	3	3	2.798
Go to the birthday party	3	2.733	0	0	3	2.736	2.539
Go to the jeweler's shop	1.843	1.867	1.5	0	1.941	1.894	1.956
Go to the park	1.881	1.89	1.625	1.462	2.095	2.057	2.069
Go to the university	2.894	2.771	0	1	2.964	2.913	3.048
TOTAL	2.81	2.793	2.111	2.2	2.834	2.8183	2.819

Table 6.3: Hyponym and hypernym generation results. AJ=adjectives, AD=Adverbs, Conj=Conjunctions, Det=Determiners, N=Nouns, V=Verbs. The table reports the number of terms, belonging to each grammatical category, collected during the last part of the second experiment, as well as the ratio between the number of nouns and the number of verbs.

Script Name	AJ	AD	Conj	Det	N	V	N/V
Go to the bookshop	0	0	0	0	239	1	239
Go to the airport	2	1	0	0	38	10	3.8
Go to the beach	42	5	0	0	218	41	5.3171
Go to the shopping center	6	3	0	0	101	28	3.6071
Go to the stadium	48	3	1	1	572	169	3.3846
Go out with my dog	49	11	0	0	180	128	1.4063
Go to ski	6	1	0	0	49	26	1.8846
Make a constructive experience abroad	8	1	0	0	76	35	2.1714
Go to the hospital to visit a friend with a broken arm	4	1	0	0	52	39	1.3333
Go to an art exhibition	23	1	0	0	190	52	3.6538
Go to the disco	20	3	0	0	249	23	10.8261
Go to the zoo	216	41	2	0	741	725	1.0221
Go to the chemist's	7	0	0	0	31	10	3.1
Exit with my friends	1	1	0	0	32	260	0.1231
Go to the library	5	0	0	0	54	7	7.7143
Go to work	0	0	0	0	206	3	68.6667
To cook	2	0	0	0	58	24	2.4167
Go to a party	10	1	0	0	77	29	2.6552
Go to the doctor	1	0	0	0	33	7	4.7143
Go to the hairdresser's	2	1	0	0	54	3	18

Go to the museum	4	0	0	0	55	7	7.8571
Go to the class	2	0	0	0	75	11	6.8182
Give private lessons	1	1	0	0	44	4	11
Go to the beautician	2	0	0	0	66	15	4.4
Horse riding	16	0	0	0	256	3	85.3333
Go to the gym	1	0	0	0	33	8	4.1250
Go to a wedding party	9	1	0	0	194	19	10.2105
Clean the room	3	0	0	0	20	2	10
Go to the perfumery	6	0	0	0	119	9	13.2222
Go to the grocery	18	4	0	0	47	24	1.9583
Go on a cruise	7	1	0	0	105	18	5.8333
Go to the theater	11	2	0	0	92	29	3.1724
Trip on a bicycle with friends	3	1	0	0	61	14	4.3571
Go to the seaside	14	0	0	0	113	50	2.26
Go to the dentist	9	1	0	0	65	15	4.3333
Go to the birthday party	3	0	0	0	65	12	5.4167
Go to the jeweler's shop	6	2	0	0	41	19	2.1579
Go to the park	17	1	0	0	124	33	3.7576
Go to the university	27	6	0	0	80	78	1.0256
TOTAL	611	94	3	1	4905	1990	2.465

Table 6.4: Synonym generation results. AJ=adjectives, AD=Adverbs, Conj=Conjunctions, Det=Determiners, N=Nouns, V=Verbs. The table reports the number of terms, belonging to each grammatical category, collected during parts A and B of the second experiment, as well as the ratio between the number of nouns and the number of verbs.

Script Name	AJ	AD	Conj	Det	N	V	N/V
Go to the bookshop	38	17	1	0	106	119	0.8908

Go to the airport	202	82	3	2	232	218	1.0642
Go to the beach	51	23	0	0	101	101	1
Go to the shopping center	169	76	3	0	340	260	1.3077
Go to the stadium	235	42	1	1	638	598	1.0669
Go out with my dog	89	23	0	0	161	176	0.9148
Go to ski	85	15	0	1	608	351	1.7322
Make a constructive experience abroad	683	185	4	0	926	1589	0.5828
Go to the hospital to visit a friend with a broken arm	150	66	3	0	285	287	0.993
Go to an art exhibition	178	79	0	0	295	187	1.5775
Go to the disco	59	18	0	0	75	102	0.7353
Go to the zoo	54	11	0	1	206	174	1.1839
Go to the chemist's	150	59	1	0	423	488	0.8668
Exit with my friends	17	8	0	0	39	246	0.1585
Go to the library	97	14	0	0	433	311	1.3923
Go to work	25	30	0	0	170	141	1.2057
To cook	36	59	2	1	75	262	0.2863
Go to a party	41	27	1	2	178	95	1.8737
Go to the doctor	150	26	0	1	136	221	0.6154
Go to the hairdresser's	108	25	0	1	286	292	0.9795
Go to the museum	91	45	1	0	199	311	0.6399
Go to the class	368	117	1	3	447	437	1.0229
Give private lessons	125	6	0	1	190	252	0.754
Go to the beautician	374	103	0	1	506	508	0.9961
Horse riding	44	6	1	0	38	148	0.2568

Go to the gym	379	71	0	0	610	457	1.3348
Go to a wedding party	112	36	0	0	180	311	0.5788
Clean the room	236	158	6	1	334	435	0.7678
Go to the perfumery	67	28	0	2	73	163	0.4479
Go to the grocery	440	223	5	2	450	426	1.0563
Go on a cruise	182	12	1	0	494	590	0.8373
Go to the theater	354	205	3	0	510	815	0.6258
Trip on a bicycle with friends	111	42	0	2	195	450	0.4333
Go to the seaside	25	3	0	0	265	129	2.0543
Go to the dentist	165	13	0	0	270	266	1.015
Go to the birthday party	159	11	0	0	122	439	0.2779
Go to the jeweler's shop	456	212	4	2	313	310	1.0097
Go to the park	163	138	9	13	350	343	1.0204
Go to the university	94	34	0	1	143	166	0.8614
TOTAL	6562	2348	50	38	11402	13174	0.8655

References

Abelson, R. P. (1976). Script processing in attitude formation and decision making.

Abelson, R. P. (1981). Psychological status of the script concept. *American psychologist*, *36*(7), 715.

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon.* John Wiley & Sons.

Allington, D. (2005). Re-reading the script: a discursive appraisal of the use of the'schema'in cognitive poetics. *Working With English: Medieval and Modern Language, Literature and Drama*, *2*, 1–9.

Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, *97*(21), 11149–11152.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, *22*(3), 261–295.

Anderson, J. R. (2005). *Cognitive psychology and its implications.* Macmillan.

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010a). Comparative analysis of networks of phonologically similar words in english and spanish. *Entropy*, *12*(3), 327–337.

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010b). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, *20*(03), 679–685.

Atkinson, R. C., & Juola, J. F. (1974). *Search and decision processes in recognition* *memory.* WH Freeman.

Barabási, A.-L. (2011). The network takeover. *Nature Physics*, *8*(1), 14.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, *286*(5439), 509–512.

Barabási, A.-L., et al. (2009). Scale-free networks: a decade and beyond. *science*, *325*(5939), 412.

Bilotta, E., & Pantano, P. (2010). *Cellular automata and complex systems: Methods for modeling biological phenomena: Methods for modeling biological phenomena.* IGI Global.

Bilotta, E., Pantano, P., & Stranges, F. (2007). A gallery of chua attractors: Part i. *International journal of Bifurcation and chaos*, *17*(01), 1–60.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, 1170–1182.

Bonacich, P. (1991). Simultaneous group and individual centralities. *Social Networks*, *13*(2), 155–168.

Bonacich, P., & Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social networks*, *23*(3), 191–201.

Borge-Holthoefer, J., & Arenas, A. (2010). Semantic networks: structure and dynamics. *Entropy*, *12*(5), 1264–1302.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive psychology*, *11*(2), 177–220.

Caldarelli, G. (2007). Scale-free networks: complex webs in nature and technology. *OUP Catalogue*.

Cellar, D. F., & Barrett, G. V. (1987). Script processing and intrinsic motivation: The cognitive sets underlying cognitive labels. *Organizational Behavior and Human Decision Processes*, *40*(1), 115–135.

Charniak, E. (1972). *Toward a model of children's story comprehension* (Tech. Rep.). DTIC Document.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, *51*(4), 661–703.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2), 240–247.

Costa, L. d. F., Oliveira Jr, O. N., Travieso, G., Rodrigues, F. A., Villas Boas, P. R., Antiqueira, L., ... Correa Rocha, L. E. (2011). Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, *60*(3), 329–412.

Cramer, A. O., Waldorp, L. J., van der Maas, H. L., & Borsboom, D. (2010). Comorbidity: a network perspective. *Behavioral and Brain Sciences*, *33*(2-3), 137–150.

Da Fontoura Costa, L. (2004). What's in a name? *International Journal of Modern Physics C*, *15*(3), 371-379. Retrieved from `http://www.scopus.com/inward/record.url?eid=2-s2.0-13844289260&partnerID=40&md5=533d4252fbe0910c7a3912e4a9f3ed30` (cited By 15)

Davidson, D., & Hoe, S. (1993). Children's recall and recognition memory for typical and atypical actions in script-based stories. *Journal of Experimental Child Psychology*, *55*(1), 104–126.

De Deyne, S. L. S., Navarro, D. J., Perfors, A. F., & Storms, G. (2012). Strong structure in weak semantic similarity: A graph based account. In *Annual meeting of the cognitive science society (34th: 2012: Sapporo, japan) cogsci 2012.*

de Jesus Holanda, A., Pisa, I. T., Kinouchi, O., Martinez, A. S., & Ruiz, E. E. S. (2004). Thesaurus as a complex network. *Physica A: Statistical Mechanics* *and its Applications*, *344*(3), 530–536.

Erdös, P., & Rényi, A. (1959). On random graphs, i. *Publicationes Mathematicae (Debrecen)*, *6*, 290–297.

Ferreira, A., Corso, G., Piuvezam, G., & Alves, M. (2006). A scale-free network of evoked words. *Brazilian Journal of Physics*, *36*(3A), 755–758.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215–239.

Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current opinion in neurobiology*, *23*(2), 250–254.

Gernsbacher, M. A. (1991). Cognitive processes and mechanisms in language comprehension: The structure building framework. *Psychology of Learning and Motivation*, *27*, 217–263.

Graesser, A. C., Gordon, S. E., & Sawyer, J. D. (1979). Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer+ tag hypothesis. *Journal of Verbal Learning and Verbal Behavior*, *18*(3), 319–332.

Graesser, A. C., Woll, S. B., Kowalski, D. J., & Smith, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 503.

Grafman, J., Thompson, K., Weingartner, H., Martinez, R., Lawlor, B. A., & Sunderland, T. (1991). Script generation as an indicator of knowledge representation in patients with alzheimer's disease. *Brain and language*, *40*(3), 344–358.

Gravino, P., Servedio, V. D., Barrat, A., & Loreto, V. (2012). Complex structures and semantics in free word association. *Advances in Complex Systems*, *15*(03n04).

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind predicting fluency with pagerank. *Psychological Science*, *18*(12), 1069–1076.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of memory and language*, *63*(3), 259–273.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–739.

Hubbell, C. H. (1965). An input-output approach to clique identification. *Sociometry*, 377–399.

Hudson, J. A. (1988). Children's memory for atypical actions in script-based stories: Evidence for a disruption effect. *Journal of Experimental Child Psychology*, *46*(2), 159–173.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210–1224.

i Cancho, R. F. (2005). The structure of syntactic dependency networks: insights from recent advances in network theory. *Problems of quantitative linguistics*, 60–75.

Jasny, B. R., Zahn, L. M., Marshall, E., & Cho, A. (2009). Complex systems and networks. *Science*, *325*, 405.

Jurafsky, D., & Martin, J. H. (2000). *Speech & language processing.* Pearson Education India.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), 39–43.

Khanin, R., & Wit, E. (2006). How scale-free are biological networks. *Journal of computational biology*, *13*(3), 810–818.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge university press.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, *46*(5), 604–632.

Krylov, N. S. (2014). *Works on the foundations of statistical physics.* Princeton

University Press.

Lamb, S. M. (1970). Linguistic and cognitive networks. In *In (p. garvin, ed.): Cognition: A multiple view* (pp. 195–222). New York: Spartan Books.

Light, L. L., & Anderson, P. A. (1983). Memory for scripts in young and older adults. *Memory & Cognition*, *11*(5), 435–444.

Liu, H. (2009). Statistical properties of chinese semantic networks. *Chinese Science Bulletin*, *54*(16), 2781–2785.

MacKay, D. G. (1992). Errors, ambiguity, and awareness in language perception and production. In *Experimental slips and human error* (pp. 39–69). Springer.

Makaruk, H. E., & Owczarek, R. (2008). Hubs in languages: Scale free networks of synonyms. *arXiv preprint arXiv:0802.4112*.

Mandler, J. M., & Johnson, N. S. (1980). On throwing out the baby with the bathwater: A reply to black and wilensky's evaluation of story grammars*. *Cognitive Science*, *4*(3), 305–312.

Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and instruction*, *4*(2), 91–115.

Milgram, S. (1967). The small world problem. *Psychology today*, *2*(1), 60–67.

Minksy, M. (1975). A framework for representing knowledge. *The psychology of computer vision*, 211–277.

Mittal, K., & Jain, A. (2015). Word sense disambiguation method using semantic similarity measures and owa operator. *Corpus*, *5*(02).

Motter, A. E., de Moura, A. P., Lai, Y.-C., & Dasgupta, P. (2002). Topology of the conceptual network of language. *Physical Review E*, *65*(6), 065102.

Neal, Z. P. (2012). *The connected city: How networks are shaping the modern metropolis.* Routledge.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, *69*(2), 026113.

Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica*, *124*(1), 106–128.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The pagerank citation ranking: Bringing order to the web. *Stanford Digital Libraries Working Paper*.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on association for computational linguistics* (p. 271).

Pastor-Satorras, R., & Vespignani, A. (2004). *Structure and evolution of the internet: A statistical physics approach.* Cambridge University Press, Cambridge.

Pearson, K. (1905). The problem of the random walk. *Nature*, *72*(1865), 294.

Polepalli Ramesh, B., Sethi, R., & Yu, H. (2015). Figure-associated text summarization and evaluation. *PloS one*, *10*(2), e0115671.

Pollatsek, A., Ashby, J., & Clifton Jr, C. (2012). *Psychology of reading.* Psychology Press.

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral science*, *12*(5), 410–430.

Rabinowitz, J. C., Mandler, G., & Patterson, K. E. (1977). Determinants of recognition and recall: Accessibility and generation. *Journal of Experimental Psychology: General*, *106*(3), 302.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551–1555.

Révész, P. (1990). Random walk in random and non-random environments. *World Scienfitic, Singapore*.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*(3), 1059–1069.

Rumelhart, D., & Ortony, A. (1977). *The representation of knowledge in memory. in rc anderson, rj spiro, & we montague (eds.) schooling and the acquisition of knowledge.* Hillsdale, NJ: Erlbaum.

Schank, R. (1982). *Dynamic memory, 1982.* Cambridge University Press, New York.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: an inquiry into human knowledge structures. hillsdale, nj: L.* Erlbaum.

Sigman, M., & Cecchi, G. A. (2002). Global organization of the wordnet lexicon. *Proceedings of the National Academy of Sciences*, *99*(3), 1742–1747.

Sjöberg, M., Albrectsen, B., & Hjältén, J. (2000). Truncated power laws: a tool for understanding aggregation patterns in animals? *Ecology letters*, *3*(2), 90–94.

Smith, D. A., & Graesser, A. C. (1981). Memory for actions in scripted activities as a function of typicality, retention interval, and retrieval task. *Memory & Cognition*, *9*(6), 550–559.

Solé, R. V., Corominas-Murtra, B., Valverde, S., & Steels, L. (2005). Language networks: their structure, function, and evolution.

Sporns, O. (2010). Connectome. *Scholarpedia*, *5*(2), 5584.

Sporns, O. (2011). *Networks of the brain.* MIT press.

Stevenson, M., & Wilks, Y. (2003). Word sense disambiguation. *The Oxford Handbook of Comp. Linguistics*, 249–265.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41–78.

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jtrace: A reimplementation and extension of the trace model of speech perception and spoken word recognition. *Behavior Research Methods*, *39*(1), 19–30.

Strori, D., Bombaci, A., & Bingol, H. (2007). Cross comparison of synonym graphs in a multi linguistic context. In *Computer and information sciences, 2007. iscis 2007. 22nd international symposium on* (pp. 1–7).

van den Heuvel, M. P., Mandl, R. C., Stam, C. J., Kahn, R. S., & Pol, H. E. H. (2010). Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. *The Journal of Neuroscience*, *30*(47), 15915–15926.

Van Der Hofstad, R. (2009). Random graphs and complex networks. *Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf*, 11.

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research*, *51*(2), 408–422.

Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon 1. *American Journal of Sociology*, *105*(2), 493–527.

Watts, D. J. (2004). The" new" science of networks. *Annual review of sociology*, 243–270.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *nature*, *393*(6684), 440–442.

Wechsler, D. (2008). Wechsler adult intelligence scale–fourth edition (wais–iv). *San Antonio, TX: NCS Pearson.*

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162.