COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.

Javier Borge Holthoefer

ISBN: 978-84-694-2176-5
Dipòsit Legal: T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Javier Borge Holthoefer

SEMANTIC NETWORKS AND COGNITIVE DYNAMICS

TESI DOCTORAL
dirigida pel Dr. Alexandre Arenas Moreno

Departament
d'Enginyeria Informàtica i Matemàtiques

UNIVERSITAT ROVIRA I VIRGILI

Tarragona
2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Departament d'Enginyeria
[DΣIM]
Informàtica i
Matemàtiques

Av. Països Catalans, 26
43007 Tarragona

FAIG CONSTAR que aquest treball, titulat *Semantic networks and cognitive dynamics*, que presenta Javier Borge Holthoefer per a optar al títol de Doctor, ha estat realitzat sota la meva direcció al Departament d'Enginyeria Informàtica i Matemàtiques d'aquesta universitat.

Tarragona, 20 de gener de 2011

El director de la tesi doctoral

Dr. Alexandre Arenas Moreno

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

A la Silvia:
Jo només sé dos o tres coses,
ella ja en sap quatre o cinc...
i ens ha costat Déu i ajuda
arribar fins aquí.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Agraïments

A la Silvia, l'amor de la meva vida. Sense ella res tindria cap sentit.
A l'Àlex Arenas: és una sort i un plaer haver-lo conegut, haver-hi treballat, poder-hi conviure. Sense ell no hauria arribat fins aquí. I just al costat: Yamir Moreno, Sergio Gómez, Jose Cuesta i Víctor Eguíluz: d'ells espero encara aprendre molt. També un record pels amics Toni Garcia, Gorka Zamora, Sergi Lozano, Sandro Meloni, Pau Erola.
Agraïments a José E. García-Albea, Joan Ferré, Manel Sanromà i Francesc Giralt: els he anat trobant pel camí, i sempre han tingut una paraula (o dos) que han acabat sent lliçons.
Last but not least, al Hamlet.

Javier Borge Holthoefer
Tarragona
Finals de 2010

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Contents

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Resum

Introducció

A la dècada de 1960s s'iniciava amb forta embranzida un nou paradigma científic, la Ciència Cognitiva, que es posava com horitzó una definició integral dels mecanismes cognitius –la caixa negra– (humans i animals) coherents amb, explicatius de la conducta que exhibim els éssers vius –els *outputs* externs, mesurables. No és poca cosa. Confiaven, no obstant, en una autèntica sinergia de corrents i disciplines que, complementant-se, poguessin assolir l'objectiu marcat. S'hi compten, entre aquestes disciplines, la filosofia (Fodor), la psicologia (Miller, Gardner), les ciències de la computació –sota la forma de la llavors naixent intel·ligència artificial (Newell, Simon, Minsky)–, la cibernètica (Wiener), la teoria de la informació (Shannon i Weaver), la lingüística (Chomsky), l'antropologia, la neurociència...

És en aquest context que, l'any 1969, Collins i Quillian van plantejar-se una pregunta ben complicada: com s'organitzen els conceptes a la nostra ment? No eren pas els primers a preguntar-s'ho, però sí ho van ser a intentar construir un model computacional que fós alhora experimentalment contrastable. En concret, proposen una estructura rígida i jeràrquica en forma d'arbre: els conceptes s'organitzen per nivells; els nodes "fulla" són exemplars concrets; els nodes "pare" representen categories. El model incorpora també una dinàmica, la *spreading activation* segons la qual l'activació d'un node inicia un procés difusiu sobre els enllaços de l'arbre, que s'atenua en el temps (a mesura que el senyal s'allunya del node on s'ha originat). Les prediccions del model es veieren confirmades només parcialment quan es provaven en humans. Més enllà de l'èxit de la seva proposta, Collins i Quillian (i, en treballs successius, Collins i Loftus) llegaren les bases per l'estudi empíric de l'organització dels *building blocks* de la nostra vida mental, de la seva estructura a gran escala i dels mecanismes que hi operen. Aquesta és la primera llavor de la present memòria,

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

però en la seva forma original contempla una rigidesa i una escassedat de dades que fa inviable qualsevol avenç.

Han de passar 30 anys abans no trobem un segon ingredient que superi, almenys quantitativament, el primer escull. En els darrerrs anys la revolució dels ordinadors i d'Internet ens ha permès el que abans només era un desig: capacitat d'emmagatzemament suficient, velocitat de processament i disponibilitat de dades. Mentre Collins i Quillian manegaven dades d'uns centenars de conceptes/nodes, nosaltres podem ja implementar i analitzar estructures amb milers, i de vegades milions de vèrtex.

Falta encara una tercera llavor. Fins ara, tenim que els conceptes i les paraules estan relacionats, formant estructures a gran escala; sabem que determinades dinàmiques sobre aquesta estructura possibiliten l'activitat mental (almenys la lingüística). Tenim a més disponibilitat de dades empíriques i força computacional per estudiar-les. Pel camí, hem deixat enrera un arbre per passar a una xarxa. La teoria de grafs ens proporciona tècniques i conceptes per comprendre el que tenim entre mans. Però el canvi d'escala que suposa passar d'estructures analitzables a cop d'ull a estructures desordenades i amb milers d'elements ens obliga a alguna cosa més. Cal un canvi en l'enfocament analític. Aquest enfocament ha de ser preeminentment estadístic, per poder respondre la pregunta, "Quin aspecte té aquesta xarxa, si de fet és impossible visualitzar-la?". És davant aquest panorama que sorgeixen amb força les eines de la física estadística, i particularment la disciplina dedicada a sistemes complexes. Ara sí: als conceptes de grau, *path length*, partició, etc. de la teoria de grafs, hi podem afegir els de auto-organització, criticalitat, emergència, transició de fase, etc. que ens proporciona la llarga tradició iniciada amb Maxwell i Boltzmann.

A vol d'ocell

Han quedat esboçades les tres grans linies mestres d'aquest treball. En ell ens proposem desenvolupar els resultats més significatius de l'aplicació dels mètodes esmentats a les dades de les que hem anat disposant, provinents sobretot del treball experimental en psicolingüística i neuropsicologia.

Més enllà de la Introducció, el Capítol 2 es dedica a la revisió dels conceptes elementals (i altres més elaborats) que conformen la teoria de xarxes complexes moderna. Les eines exposades poden classificar-se, *grosso modo*, en tres grans grups, segons el nivell d'anàlisi al què estan dedicades:

- Descriptors a nivell *micro*, els quals ens ajuden a caracteritzar a nivell individual les unitats que composen el sistema. En aquest nivell hi són pertinents

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

el grau k, el *clustering coefficient* o la *vertex betweenness* dels nodes considerats individualment.

- Descriptors a nivell *macro*, que ens informen de les propietats estadístiques globals –de la xarxa sencera: distribució de grau $P(k)$, *clustering coefficient* promig, correlacions de grau $P(k|k')$, etc.

- Descriptors a nivell *meso*. Entre els dos extrems anteriors existeix un tercer nivell d'anàlisi, el mesoscòpic. Aquí pertany la descripció dels subconjunts de nodes que estan més densament connectats entre ells que amb l'exterior. En altres paraules, en aquest nivell s'estudia l'existència i caracetrització de l'estructura de comunitats (o modular) del sistema.

Així doncs, en aquest capítol s'estableixen les bases metodològiques que han servit per construir la resta del treball.

El Capítol 3 es dedica a aclarir dos aspectes importants: en primer lloc, quina mena de dades s'han emprat (i s'empren) en l'estudi a gran escala del llenguatge i els fenòmens cognitius que l'envolten. A banda d'una mera exposició, s'intenta una valoració crítica sobre la validesa i l'abast d'aquestes dades: en l'aproximació *complex networks* al llenguatge, no poden posar-se al mateix calaix estudis basats en *corpus*, *thesauri* o diccionaris, que aquells on les dades s'han obtingut en entorns controlats, experimentalment validats i amb vistes a l'estudi psicolingüístic.

En segon lloc, es revisen els treballs més destacables (*to our best knowledge*) que s'han fet fins al moment actual al voltant del llenguatge i la cognició. Per claredat, distingim els treballs que estudien el llenguatge *per se*, merament com objecte d'estudi; d'aquells que intenten anar més enllà de la pura descripció per cercar prediccions cognitivament plausibles i dinàmiques cognitives que eixamplin la nostra comprensió de les mateixes.

Al Capítol 4 s'introdueix el *Random Inheritance Model*, que representa un intent per comprendre com emergeixen la similitud semàntica entre paraules i les categories semàntiques. Més encara, s'intenta formalitzar el fenomen de categorització semàntica en el context de processos markovians i els estats intermitjos d'aquests processos, i.e. estats metaestables.

Els resultats es comparen amb altres tècniques de *information extraction* reconegudes a la literatura, i amb dades empíriques basades en les respostes de subjectes humans.

Al marge del funcionament *normal* de les nostres estructures lingüístiques, ens interessa també comprendre les seves lesions i malalties. Al Capítol 5 presentem un model de *degradació semàntica* que emula processos neurodegeneratius i prediu acuradament, a nivell qualitatiu, les observacions experimentals amb malalts d'Alzheimer que s'han fet en l'àmbit de la neuropsicologia.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

En aquests processos degeneratius convergeixen interessos multidisciplinars, que van de la mateixa cognició al fenomen de percolació en física estadística, passant per la comprensió de la relació existent entre la *performance* lingüística i les seves arrels neurològiques.

El Capítol 6 queda finalment dedicat a una reflexió global d'aquesta memòria, tancant el treball amb les linies de recerca que l'autor, si les circumstàncies ho permeten, preten promoure en el futur.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Chapter 1

Introduction

The origins of Artificial Intelligence (AI) can be traced in Philosophy for at least 2000 years. Aristotle or Llull represent important steps in that direction regarding logic, the formal theory of deductive reasoning. Descartes seems to have been interested in "mechanical man", although more as a metaphor than as a possibility. Leibniz, on the other hand, foresaw the possibility of mechanical reasoning devices using rules of logic to settle disputes. Both Leibniz and Pascal designed calculating machines that mechanized arithmetic, but they never made the claim that the devices could think. But only in the last half century have we had computational devices and programming languages powerful enough to build experimental tests of ideas about what intelligence is. What previously could have only been a formalism or a *gedankenexperiment* in the History of Thought; or mere fiction in Literature, is now possible. It is in this scenario that AI emerged as an independent discipline.

AI in its formative years was influenced by ideas from many disciplines. These came from people working in engineering (such as Wiener's work on cybernetics, which includes feedback and control), biology (for example, Ashby, McCulloch and Pitts's work on neural networks), experimental psychology (Newell and Simon), communication theory (Shannon's theoretical work), game theory (notably by Von Neumann and Morgenstern), mathematics and statistics (Good), logic and philosophy (Turing, Church, Hempel), and linguistics (Chomsky's work on grammar). These lines of work made their mark and continue to be felt. However, just as any mature paradigm, it also began a process of specialization: in most of current AI, those influences have been increasingly abandoned.

Following this line, there are at least two complementary views of AI: one as an engineering discipline concerned with the creation of intelligent machines, the other as an empirical science concerned with the computational modeling of human intelligence. When the field was young, these two views were seldom distinguished. Since then, a substantial divide has opened up, with the former view dominating modern AI and the latter view characterizing much of modern cognitive science.

We might call the first view *weak AI* [133, 152, 153], which intends to develop techniques that end up with a system or an artifact with intelligent behavior. The system is intelligent in the sense it can make correct predictions, take good decisions,

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

provide valid inferences, whatever may be the mechanisms or algorithms that make such behavior possible. Achieving human or close-to-human level of performance is the point. Take for example the efforts in automatic text analysis which seek to extract relevant information, which need not be inspired in any cognitive strategy to do so (see, for example, LSA proposal in [99, 100, 101]). In this case, we can say that the artificial system *performs* like a human, rather than *thinks* like a human.

On the other hand, *strong AI* is of course also interested in artifacts showing intelligent behavior; however, it pays much attention on how humans act intelligent, and try to reproduce it. As Herbert Simon put it, "The purpose of AI is to use a computer's artificial intelligence to understand how humans think. In a humanoid way. If you test your programs not merely by what they can accomplish, but how they accomplish it, then you're really doing cognitive science; you're using AI to understand the human mind"[1]. Underlying this rather daring perspective is the principle by which cognition is a form of computation: fundamental work by Alan Turing [178] in the 1930s formalized the notion of universal computation; the Church-Turing thesis [38, 178] proposed that all sufficiently powerful computing devices were essentially identical in the sense that any one device could emulate the operations of any other. From here it was a small step to the bold hypothesis that human cognition was a form of computation in exactly this sense, and could therefore be emulated by computers[2].

AI is, in this strong sense, embarked on a long-term objective, along with other disciplines, which is the understanding of human cognition. In the end, the study of actual cognition means research about what kind of mechanisms, processes and constraints are actually going on internally (in the minds and brains of real subjects), such that they end up with intelligent external performance (behavior). In this line of work one can find, for instance, part of the neural networks tradition [90, 109, 146, 147] which intended neurological realism; or the work at the Sony Computer Science Laboratory in Paris [167, 168, 169, 181] with multi-agent systems studying the emergence of language conventions; or the family of probabilistic approaches to language [19, 79, 81].

It is along the lines of AI that this work is to be placed, and more generally in the broader discipline of cognitive science. For the last 40 years, cognitive science has increasingly settled as a stable paradigm [69]. Remarkably, an interdisciplinary nature lies at the kernel of the cognitive approach. Accordingly, the work presented in this thesis intends to capture such multidisciplinary spirit that characterizes the cognitive effort: first, the core of the thesis, the object under study is language (the

[1] excerpt from Doug Stewart: *Interview with Herbert Simon*, June 1994. Omni Magazine.

[2] Paradoxically enough, the foundational idea of the *weak* AI line is based also on an idea by Turing, i.e. the Turing Test exposed in his work "Computing Machinery and Intelligence" [179].

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

specification of its structure in memory, the study of how such structure facilitates performance, the emergence of categories and meaning, etc.). Secondly, this research project presents a methodology strongly based on computer science (using computation both as an *instrument*, to measure and calculate quantities; and as an object of study itself, to simulate dynamics which mimic cognitive processes, resting on the basic assumption that mental activity is, from a formal point of view, computational –strong AI). Finally, the methodological principles and modeling techniques of the whole research are to be found in statistical physics and complex systems science. In the recent years there has been a trend towards applications of statistical physics to interdisciplinary fields as diverse as biology, medicine, information technology, computer science or, as the next chapters try to illustrate, cognitive science. As we shall see, this general framework stands as a proper methodology to gain insight on real cognition.

Cognition, computer science and statistical physics: these main lines fetch up to a specific object, which is *complex networks*. Originally developed in the sociology environment, network analysis methods have been used to model many complex real-world phenomena. Examples are numerous. The Internet is a network of routers or domains. The World Wide Web (WWW) is a network of websites. An organization is a network of people. Global economy is a network of national economies, which are themselves networks of markets; and markets are themselves networks of interacting producers and consumers. Food webs and metabolic pathways can all be represented by networks. Moreover, diseases are transmitted through social networks; and computer viruses occasionally spread through the Internet. Energy is distributed through transportation networks, both in living organisms, man-made infrastructures, and in many physical systems such as the power grids. All of them stand as examples of research under a network modeling approach. What about cognition? Complex networks are suitable models at different levels of explanation of cognitive phenomena. On one side, the brain is a network of neurons. It probably stands as the most complex challenge in complex systems science. On the other, language fits perfectly as an object of study under the shape of a network (be it at the level of phone-phone, word-word, word-document relationships).

Marr's levels might be illustrative of the structure of this work. David Marr advocated and exemplified an approach to brain modeling that is based on computational sophistication together with a thorough knowledge of the biological facts. In his 1982 book *Vision* [107], he envisioned a research program for the field of vision research, using a distinction between three complementary levels at which information processing systems may be described:

1. The computational level, at which a system is described (mental represenations).

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

2. The algorithmic level, at which a system's method is described (cognitive processes).

3. The implementational level, at which a system's means are described (neural structures).

This distinction may be applied to the cognitive system. Marr's point was that the levels of description should all be taken equally seriously, to arrive eventually at a comprehensive theory consisting of three complementary descriptions which, together, explain "how the goal is reached with a method that is allowed by the means".

Marr's distinction between the three levels of description has stimulated integrative theoretical research. It is useful to specify the position of scientific findings in the total field of cognitive (neuro)science.

This three-wise perspective clarifies the role of complex networks in this work. Language has been frequently modeled as a network in the psycholinguistic area (lexical and semantic networks) [13, 40, 41, 91]. Then complex network theory, understood now as a toolbox, can provide novel insight at the first level regarding topology, organization, modular (community, category) structure, etc. To do so, it comprehends many descriptors ranging from the local to the large-scale level (see Chapter 2 for a review on network theory as a set of descriptors; and most of Chapter 3 for a state-of-the-art review, i.e. how network theory as a toolbox has brought interesting insight to language).

The algorithmic level deals with the study of cognitively plausible dynamics (algorithms) operating on top of the structural level (lexical and/or semantic networks). We know from complex systems literature that one fundamental aspect concerning the analysis of complex systems is the evidence of mutual influence between dynamical behavior and topological structure. The importance of structure to understand (cognitive) dynamics again justifies a network theory approach. Such approach sheds light on the mechanisms and strategies that humans put at work when processing language, either to search and retrieve words, or to form emerging semantic categories, among others. Work along this level of description can be found partially in Chapters 3 (reviewing other author's proposals), 4 and 5.

Real as it may appear to us (fortunately), language is not a tangible object. Although we hear/read or produce/write meaningful utterances almost any time while awake (some of us even asleep), messages and ideas, words and relations among them originate in an abstract object (mind) and remain abstract (concepts, mental representations). Then, levels 1 and 2 do not suffice for a complete understanding of linguistic phenomena. The third level fills the gap by studying the neurological roots of language. Unfortunately, this is not a work in the order of mag-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

nitude of chapters; it is rather in the order of a lifetime. Nonetheless, Chapter 6, devoted to Conclusions, tackles this and other issues about the research currently developed and, hopefully, in the future.

This document contains more material than the one just outlined above. If we are going to say that a given program thinks like a human, we must have some way of determining how humans think. We need to get inside the actual workings of human minds. There are three ways to do this: through introspection –trying to catch our own thoughts as they go by; through psychological experiments –observing a person in action; and through brain imaging –observing the brain in action. Introspection has been rejected as a plausible theory-building method for a long time; beyond the first two decades of the 20th century it became an outsider from scientific studies. The latter two –psychological experimentation and brain imaging– stand as our only viable way to get some insight to cognition. In Chapter 3 we describe the main datasets that are used along the rest of chapters, along with a brief description of how they were acquired. Also, whenever necessary, some psycholinguistic well-known empirical methods are outlined (such as the semantic priming paradigm in Chapter 5). The current manuscript doesn't get into neuroimaging or brain mapping for the reasons exposed above.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Chapter 2

Review on Network Theory

Historically, the study of networks has been mainly the domain of a branch of discrete mathematics known as *graph theory*. Since its birth in 1736, when the Swiss mathematician Leonhard Euler published the solution to the Königsberg bridge problem (consisting in finding a round trip that traversed each of the bridges of the prussian city of Königsberg exactly once), graph theory has witnessed many exciting developments and has provided answers to a series of practical questions. In addition to the developments in mathematical graph theory, the study of networks has seen important achievements in some specialized contexts, as for instance in the social sciences. Social networks analysis started to develop in the early 1920s and focuses on relationships among social entities, as communication between members of a group, trades among nations, or economic transactions between corporations.

Although the concept of small-world was already well known by sociologists [114, 176], it was in 1998 when Watts and Strogatz introduced the model of "small world" network [183], which eventually became the seed for the modern theory of complex networks. Soon it turned out that the nature of many interaction patterns observed both in natural and artificial scenarios (for instance, the World-Wide-Web, metabolic networks or scientific collaboration networks) was even more complex than the small world model. As Strogatz points out [171], networks are inherently difficult to understand, as the following list of possible complications illustrates.

1. Structural complexity: the wiring diagram could be an intricate tangle.

2. Network evolution: the wiring diagram could change over time. On the World-Wide Web, pages and links are created and lost every minute.

3. Connection diversity: the links between nodes could have different weights, directions and signs. Synapses in the nervous system can be strong or weak, inhibitory or excitatory.

4. Dynamical complexity: the nodes could be nonlinear dynamical systems. In a gene network or a Josephson junction array, the state of each node can vary in time in complicated ways.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

5. Node diversity: there could be many different kinds of nodes. The biochemical network that controls cell division in mammals consists of a bewildering variety of substrates and enzymes.

6. Meta-complication: the various complications can influence each other. For example, the present layout of a power grid depends on how it has grown over the years –a case where network evolution (2) affects topology (1). When coupled neurons fire together repeatedly, the connection between them is strengthened; this is the basis of memory and learning. Here nodal dynamics (4) affect connection weights (3).

In the past decade we have witnessed the evolution of the field of complex networks, and language has not been left out of this process: these advances have made it possible to address the previous questions from a statistical physics point of view, and to characterize the structure of language, comparing such characterizations for different languages (even for different domains), setting up growth models for them, simulating dynamics on the structures, etc.

Thus, we need a summary of this last decade's methodological advances for a full understanding of the remaining chapters. There exist many excellent reviews and books in the literature about the structure and dynamics of complex networks [3, 16, 21, 32, 39, 37, 51, 123, 128, 145]. Here we overview only those minimal requirements of the theory that will be mentioned along the current work.

2.1 Terminology in complex networks

Let us start by introducing the objects under discussion. A network is a graph with N nodes and L links. If the network is *directed* links are then named *arcs*, and account for the directionality of the connections. Otherwise the network is *undirected*, and we refer to links or edges indistinctly. Besides direction, the links can also be valued. A *weighted network* associates a label (weight) to every edge in the network. Two vertices i and j are *adjacent*, or neighbors, if they have an edge connecting them. Notice that, in a directed network, i being adjacent to j does not entail j being adjacent to i. Networks with multiple links (multigraphs) are not considered.

A *path* in a network is a sequence of vertices $i_1, i_2, \ldots i_n$ such that from each of its vertices there is an edge to the next vertex in the sequence. The first vertex is called the *start* vertex and the last vertex is called the *end* vertex. The *length* of the path or *distance* between i_1 and i_n is then the number of edges of the path, which is $n-1$ in unweighted networks. For weighted networks, the length is the addition of each weight in every edge. When i_1 and i_n are identical, their distance is 0. When i_1 and i_n are unreachable from each other, their distance is defined to be infinity (∞).

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

A *connected network* is an undirected network such that there exists a path between all pairs of vertices. If the network is directed, and there exists a path from each vertex to every other vertex, then it is a *strongly connected network*. A network is considered to be a *complete network* if all vertices are connected to one another by one edge. We denote the complete network on n vertices K_n. A *clique* in a network is a set of pairwise adjacent vertices. Since any subnetwork induced by a clique is a complete subnetwork, the two terms and their notations are usually used interchangeably. A k-clique is a clique of order k. A *maximal clique* is a clique that is not a subset of any other clique.

2.2 Complex network descriptors

2.2.1 Degree and Degree Distribution

The simplest and the most intensively studied one vertex characteristic is degree. Degree, k, of a vertex is the total number of its connections. If we are dealing with a directed graph, in-degree, k_{in}, is the number of incoming arcs of a vertex. Out-degree, k_{out} is the number of its outgoing arcs. Degree is actually the number of nearest neighbors of a vertex. Total distributions of vertex degrees of an entire network, $p(k)$, $p(k_{in})$ (the in-degree distribution), and $p(k_{out})$ (the out-degree distribution) are its basic statistical characteristics. We define $p(k)$ to be the fraction of vertices in the network that have degree k. Equivalently, $p(k)$ is the probability that a vertex chosen uniformly at random has degree k. Most of the work in network theory deals with cumulative degree distributions, $P(k)$. A plot of $P(k)$ for any given network is built through a cumulative histogram of the degrees of vertices, and this is the type of plot used throughout this article (and often referred to just as "degree distribution"). Although the degree of a vertex is a local quantity, we shall see that a cumulative degree distribution often determines some important global characteristics of networks.

From $P(k)$ we can calculate the moments of the distribution. The n-moment of $P(k)$ is defined as

$$\langle k^n \rangle = \sum_{k}^{N} k^n p(k) \tag{2.1}$$

The first moment $\langle k \rangle$ is the mean degree of the network.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

2.2.2 Strength Distribution

In weighted networks the concept of degree of a node i (k_i) can be complemented with the notion of strength of that node, $s_i = \sum_{j \in \Gamma_i} \omega_{ij}$, i.e. the sum over the nodes j in the of i, of weights from node i towards each of the nodes j in its neighborhood Γ_i. In this type of network it is possible to measure the average strength $\langle s \rangle$ with a slight modification of Equation 2.1. On the other hand, it is also possible to plot the cumulative strength distribution $P(s)$, but it is important to make a good choice in the number of bins of the histogram (this depends on the particular distribution of weights for each network).

2.2.3 Shortest Path and Diameter

For each pair of vertices i and j connected by at least one path, one can introduce the shortest path length, the so-called *intervertex distance* d_{ij}, the corresponding number of edges in the shortest path. Then one can define the distribution of the shortest-path lengths between pairs of vertices of a network and the average shortest-path length L of a network. The average here is over all pairs of vertices between which a path exists and over all realizations of a network. It determines the effective "linear size" of a network, the average separation of pairs of vertices. In a fully connected network, $d = 1$. Recall that shortest paths can also be measured in weighted networks, then the path's cost equals the sum of the weights. One can also introduce the maximal intervertex distance over all the pairs of vertices between which a path exists. This descriptor determines the maximal extent of a network; the maximal shortest path is also referred to as the *diameter* (D) of the network.

2.2.4 Clustering Coefficient

The presence of connections between the nearest neighbors of a vertex i is described by its clustering coefficient. Suppose that a node (or vertex) i in the network has k_i edges and they connect this node to k_i other nodes. These nodes are all neighbors of node i. Clearly, at most

$$\binom{k_i}{2} = \frac{k_i(k_i - 1)}{2} \tag{2.2}$$

edges can exist among them, and this occurs when every neighbor of node i connected to every other neighbor of node i (number of loops of length 3 attached to vertex i). The clustering coefficient C_i of node i is then defined as the ratio between the number E_i of edges that actually exist among these k_i nodes and the total possi-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

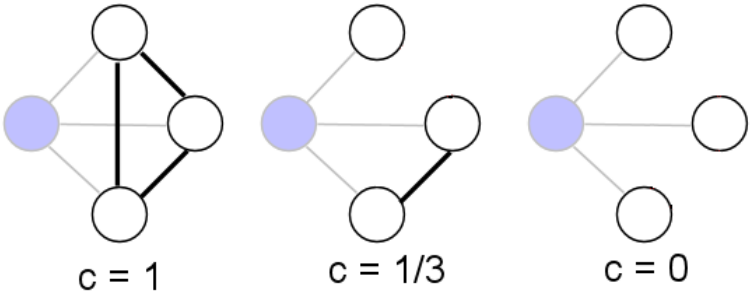

Figure 2.1: An illustration of the concept of clustering C, calculated on the gray node. In the left figure, every neighbor of the mentioned node is connected to each other; therefore, clustering coefficient is 1. In the middle picture, only two of the gray node neighbors' are connected, yielding a clustering coefficient of $1/3$; finally, in the last illustration none of the gray node's neighbors are linked to each other, which yields a clustering coefficient of 0. From Wikipedia Commons.

ble number:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{2.3}$$

Equivalently, the clustering coefficient of a node i can be defined as the proportion of 3-cliques in which i participates. The clustering coefficient C of the whole network is the average of C_i over all i, see Figure 2.1. Clearly, $C \leq 1$; and $C = 1$ if and only if the network is globally coupled, which means that every node in the network connects to every other node. By definition, *trees* are graphs without loops, i.e. $C = 0$.

The clustering coefficient of the network reflects the transitivity of the mean closest neighborhood of a network vertex, that is, the extent to which the nearest neighbors of a vertex are the nearest neighbors of each other [183]. The notion of clustering was much earlier introduced in sociology [182].

2.2.5 Centrality Measures

Centrality measures are some of the most fundamental and frequently used measures of network structure. Centrality measures address the question,"Which is the most important or central node in this network?", that is, the question whether nodes should all be considered equal in significance or not (whether exists some kind of hierarchy or not in the system). The existence of such hierarchy would then imply that certain vertices in the network are more *central* than others. There are many answers to this question, depending on what we mean by important. In this

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Section we briefly explore two centrality indexes (*betweenness* and *eigenvector centrality*) that are widely used in the network literature. Note however that betweenness or eigenvector centrality are not the only method to classify nodes' importance. Within graph theory and network analysis, there are various measures of the centrality of a vertex. For instance, besides betweenness or eigenvector centrality, there are two other main centrality measures that are widely used in network analysis: *degree centrality* and *closeness*. The first, and simplest, is degree centrality, which assumes that the larger is the degree of a node, the more central it is. The closeness centrality of a vertex measures how easily other vertices can be reached from it (or the other way: how easily it can be reached from the other vertices). It is defined as the number of vertices minus one divided by the sum of the lengths of all geodesics from/to the given vertex.

a. Betweenness One of the first significant attempts to solve the question of node centrality is Freeman's proposal (originally posed from a social point of view): *betweenness* as a centrality measure [67]. As Freeman points out, a node in a network is central to the extent that it falls on the shortest path between pairs of other nodes. In his own words, "suppose that in order for node i to contact node j, node k must be used as an intermediate station. Node k in such a context has a certain "responsibility" to nodes i and j. If we count all the minimum paths that pass through node k, then we have a measure of the "stress" which node k must undergo during the activity of the network. A vector giving this number for each node of the network would give us a good idea of stress conditions throughout the system" [67]. Computationally, betweenness is measured according to the next equation:

$$C_B(i) = \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \tag{2.4}$$

with σ_{jk} as the number of shortest paths from j to k, and $\sigma_{jk}(i)$ the number of shortest paths from j to k that pass through vertex i. Note that shortest paths can be measured in a weighted and/or directed network, thus it is possible to calculate this descriptor for any network [29]. Commonly, betweenness is normalized by dividing through by the number of pairs of vertices not including v, which is $(n-1)(n-2)$. By means of normalization it is possible to compare the betweenness of nodes from different networks.

b. Eigenvector centrality A more sophisticated version of the degree centrality is the so-called eigenvector centrality [24]. Where degree centrality gives a simple count of the number of connections a vertex has, eigenvector centrality acknowledges that not all connections are equal. In general, connections to people who are

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

themselves influential will lend a person more influence than connections to less influential people. If we denote the centrality of vertex i by x_i , then we can allow for this effect by making x_i proportional to the average of the centralities of is network neighbors:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} A_{ij} x_j \tag{2.5}$$

where λ is a constant. Defining the vector of centralities $x = (x_1, x_2, \dots)$, we can rewrite this equation in matrix form as

$$\lambda x = Ax \tag{2.6}$$

and hence we see that x is an eigenvector of the adjacency matrix with eigenvalue λ. Assuming that we wish the centralities to be non-negative, it can be shown (using the Perron-Frobenius theorem) that λ must be the largest eigenvalue of the adjacency matrix and x the corresponding eigenvector. The eigenvector centrality defined in this way accords each vertex a centrality that depends both on the number and the quality of its connections: having a large number of connections still counts for something, but a vertex with a smaller number of high-quality contacts may outrank one with a larger number of mediocre contacts. In other words, eigenvector centrality assigns relative scores to all nodes in the network based on the principle that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

Eigenvector centrality turns out to be a revealing measure in many situations. For example, a variant of eigenvector centrality is employed by the well-known Web search engine Google to rank Web pages, and works well in that context. Specifically, from an abstract point of view, the World Wide Web forms a directed graph, in which nodes are Web pages and the edges between them are hyperlinks [2]. The goal of an Internet search engine is to retrieve an ordered list of pages that are relevant to a particular query. Typically, this is done by identifying all pages that contain the words that appear in the query, then ordering those pages using a measure of their importance based on their link structure. Although the details of the algorithms used by commercial search engines are proprietary, the basic principles behind the PageRank algorithm (part of Google search engine) are public knowledge [130], and such algorithm relies on the concept of eigenvector centrality. Despite the usefulness of centrality measures, hierarchy detection and node's role determination is not a closed issue. For this reason, other classifying techniques will be explored in subsequent Sections.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

2.2.6 Degree-Degree correlation: assortativity

It is often interesting to check for correlations between the degrees of different vertices, which have been found to play an important role in many structural and dynamical network properties. The most natural approach is to consider the correlations between two vertices connected by an edge. A way to determine the degree correlation is by considering the Pearson correlation coefficient of the degrees at both ends of the edges [126, 127]

$$r = \frac{\frac{1}{N}\sum_{j>i} k_i k_j a_{ij} - [\frac{1}{N}\sum_{j>i}\frac{1}{2}(k_i+k_k)a_{ij}]^2}{\frac{1}{N}\sum_{j>i}\frac{1}{2}(k_i^2+k_j^2)a_{ij} - [\frac{1}{N}\sum_{j>i}\frac{1}{2}(k_i+k_k)a_{ij}]^2} \tag{2.7}$$

where N is the total number of edges. If $r > 0$ the network is assortative; if $r < 0$, the network is disassortative; for $r = 0$ there are no correlation between vertex degrees.

Degree correlations can be used to characterize networks and to validate the ability of network models to represent real network topologies. Newman computed the Pearson correlation coefficient for some real and model networks and discovered that, although the models reproduce especific topological features such as the power law degree distribution or the small-world property, most of them (e.g., the Erdös–Rényi and Barabási–Albert models) fail to reproduce the assortative mixing ($r = 0$ for the mentioned models) [126, 127]. Further, it was found that the assortativity depends on the type of network. While social networks tend to be assortative, biological and technological networks are often disassortative. The latter property is undesirable for practical purposes, because assortative networks are known to be resilient to simple target attack, at the least.

There exist alternative definitions of degree-degree relations. Whereas correlation functions measure linear relations, information-based approaches measure the general dependence between two variables [160]. Specially interesting is *mutual information* provided by the expression

$$I(q) = H(q) - H_c(q|q') = \sum_{k=1}^{N}\sum_{k'=1}^{N} q_c(k,k') log \frac{q_c(k,k')}{q(k)q(k')} \tag{2.8}$$

See the work by Solé and Valverde [160] for details.

2.3 Network models

2.3.1 Regular Graphs

Although regular graphs do not fall under the definition of complex networks (they are actually quite far from being complex, thus their name), they play an important

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

role in the understanding of the concept of "small world", see below. For this reason we offer a brief comment on them.

In graph theory, a regular graph is a graph where each vertex has the same number of neighbors, i.e. every vertex has the same degree. A regular graph with vertices of degree k is called a k-regular graph or regular graph of degree k [141].

2.3.2 Random Graphs

Before the burst of attention on complex networks in the decade of 1990s, a particularly rich source of ideas has been the study of random graphs, graphs in which the edges are distributed randomly. Networks with a complex topology and unknown organizing principles often appear random; thus random-graph theory is regularly used in the study of complex networks. The theory of random graphs was introduced by Paul Erdös and Alfréd Rényi [54, 55, 56] after Erdös discovered that probabilistic methods were often useful in tackling problems in graph theory. A detailed review of the field is available in the classic book of Bollobás [23]. Here we briefly describe the most important results of random graph theory, focusing on the aspects that are of direct relevance to complex networks.

a. The Erdös–Rényi Model In their classic first article on random graphs, Erdös and Rényi define a random graph as N labeled nodes connected by n edges, which are chosen randomly from the $N(N-1)/2$ possible edges [54].

In a random graph with connection probability p the degree k_i of a node i follows a binomial distribution with parameters $N-1$ and p:

$$P(k_i = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{2.9}$$

This probability represents the number of ways in which k edges can be drawn from a certain node. To find the degree distribution of the graph, we need to study the number of nodes with degree k, N_k . Our main goal is to determine the probability that N_k takes on a given value, $P(N_k = r)$. According to Equation 2.9, the expectation value of the number of nodes with degree k is

$$E(N_k) = NP(k_i = k) = \lambda_k \tag{2.10}$$

with

$$\lambda_k = N\binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{2.11}$$

The distribution of the N_k values, $P(N_k = r)$, approaches a Poisson distribution,

$$P(N_k = r) = \frac{\lambda_k^r}{r!} e^{-\lambda_k} \tag{2.12}$$

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Thus the number of nodes with degree k follows a Poisson distribution with mean value λ_k.

Although random graph theory is elegant and simple, and Erdös and other authors in the social sciences, like Rapoport [136, 137, 138, 139], believed it corresponded fundamental truth, reality interpreted as a network by current science is not aleatory. The established links between the nodes of various domains of reality follow fundamental natural laws. Despite some edges might be randomly set up, and they might play a non-negligible role, randomness is not the main feature in real networks. Therefore, the development of new models to capture real-life systems' features other than randomness has motivated novel developments. Specially, two of these new models occupy a prominent place in contemporary thinking about complex networks. Here we define and briefly discuss them.

b. Watts–Strogatz small-world network In simple terms, the small-world concept describes the fact that despite their often large size, in most networks there is a relatively short path between any two nodes. The distance between two nodes is defined as the number of edges along the shortest path connecting them. The most popular manifestation of small worlds is the "six degrees of separation" concept, uncovered by the social psychologist Stanley Milgram [114, 176], who concluded that there was a path of acquaintances with a typical length of about six between most pairs of people in the United States. This feature (short path lengths) is also present in random graphs. However, in a random graph, since the edges are distributed randomly, the clustering coefficient is considerably small. Instead, in most, if not all, real networks the clustering coefficient is typically much larger than it is in a comparable random network (i.e., same number of nodes and edges as the real network). Beyond Milgram's experiment, it was not until 1998 that Watts and Strogatz' work [183] stimulated the study of such phenomena. Their main discovery was the distinctive combination of high clustering with short characteristic path length, which is typical in real-world networks (either social, biological or technological) that cannot be captured by traditional approximations such as those based on regular lattices or random graphs. From a computational point of view, Watts and Strogatz proposed a one-parameter model that interpolates between an ordered finite dimensional lattice and a random graph. The algorithm behind the model is the following [183]:

- *Start with order*: Start with a ring lattice with N nodes in which every node is connected to its first k neighbors ($k/2$ on either side). In order to have a sparse but connected network at all times, consider $N \gg k \gg ln(N) \gg 1$.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

- *Randomize*: Randomly rewire each edge of the lattice with probability p such that self-connections and duplicate edges are excluded. This process introduces $pNK/2$ long-range edges which connect nodes that otherwise would be part of different neighborhoods. By varying p one can closely monitor the transition between order (p=0) and randomness (p=1).

The simple but interesting result when applying the algorithm was the following. Even for a small probability of rewiring, when the local properties of the network are still nearly the same as for the original regular lattice and the average clustering coefficient does not differ essentially from its initial value, the average shortest-path length is already of the order of the one for classical random graphs (see Figure 2.2).

As discussed in [184], the origin of the rapid drop in the average path length L is the appearance of shortcuts between nodes. Every shortcut, created at random, is likely to connect widely separated parts of the graph, and thus has a significant impact on the characteristic path length of the entire graph. Even a relatively low fraction of shortcuts is sufficient to drastically decrease the average path length, yet locally the network remains highly ordered. In addition to a short average path length, small-world networks have a relatively high clustering coefficient. The Watts–Strogatz model (SW) displays this duality for a wide range of the rewiring probabilities p. In a regular lattice the clustering coefficient does not depend on the size of the lattice but only on its topology. As the edges of the network are randomized, the clustering coefficient remains close to $C(0)$ up to relatively large values of p.

2.3.3 Scale-Free Networks

Certainly, the SW model initiated a revival of network modeling in the past few years. However, there are some real-world phenomena that small-world networks can't capture, the most relevant one being evolution. In 1999, Barabási and Albert presented some data and formal work that has led to the construction of various scale-free models that, by focusing on the network dynamics, aim to offer a universal theory of network evolution [14].

Several empirical results demonstrate that many large networks are scale free, that is, their degree distribution follows a power law for large k. The important question is then: what is the mechanism responsible for the emergence of scale-free networks? Answering this question requires a shift from modeling network topology to modeling the network assembly and evolution. While the goal of the former models is to construct a graph with correct topological features, the modeling of scale-free networks will put the emphasis on capturing the network dynamics.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

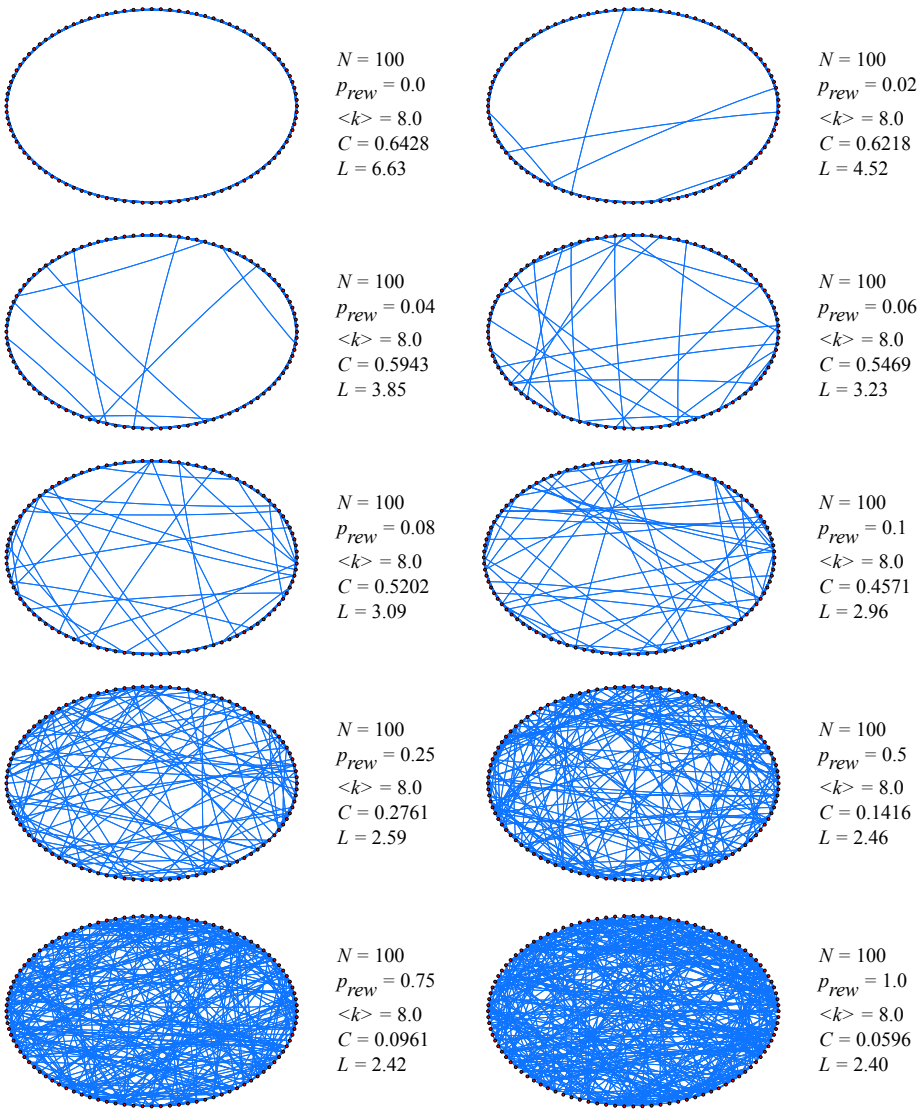

Figure 2.2: From regularity to randomness: note the changes in average path length and clustering coefficient as a function of the rewiring probability $L(p)$, $C(p)$ for the family of randomly rewired graphs. For low rewiring probabilities the clustering is still close to its initial value, whereas the average path length has already decreased significantly. For high probabilities, the clustering has dropped to an order of 10^{-2}. This figure illustrates the fact that small-world is not *a* network, but a family of networks.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

In the first place, the network models discussed up to now (random and small-world) assume that graphs start with a fixed number N of vertices that are then randomly connected or rewired, without modifying N. In contrast, most real-world networks describe open systems that *grow* by the continuous addition of new nodes. Starting from a small nucleus of nodes, the number of nodes increases throughout the lifetime of the network by the subsequent addition of new nodes. For example, the World Wide Web grows exponentially in time by the addition of new web pages.

Second, network models discussed so far assume that the probability that two nodes are connected (or their connection is rewired) is independent of the nodes degree, i.e., new edges are placed randomly. Most real networks, however, exhibit *preferential attachment*, such that the likelihood of connecting to a node depends on the nodes degree. For example, a web page will more likely include hyperlinks to popular documents with already high degrees, because such highly connected documents are easy to find and thus well known.

a. The Barabási–Albert model These two ingredients, growth and preferential attachment, inspired the introduction of the Barabási–Albert model (BA), which led for the first time to a network with a power-law degree distribution. The algorithm of the BA model is the following:

1. *Growth*: Starting with a small number (m_0) of nodes, at every time step, we add a new node with $m(\leq m_0)$ edges that link the new node to m different nodes already present in the system.

2. *Preferential attachment*: When choosing the nodes to which the new node connects, we assume that the probability Π_i that a new node will be connected to node i depends on the degree k_i, such that

$$\Pi_i = \frac{k_i}{\sum_j k_j} \tag{2.13}$$

It is specially in step (1) of the algorithm that the scale-free model captures the dynamics of a system. The power-law scaling in the BA model indicates that growth and preferential attachment play important roles in network development. However, some question arise when considering step (2): admitting that new nodes' attachment might be preferential, is there only one equation (specifically, the one mentioned here) that grasps such preference across different networks (social, technological, etc.)? Can preferential attachment be expressed otherwise?

In the limit $t \to \infty$ (network with infinite size), the BA model produces a degree distribution $P(k) \approx k^{-\gamma}$, with an exponent $\gamma = 3$, see Figure 2.3.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

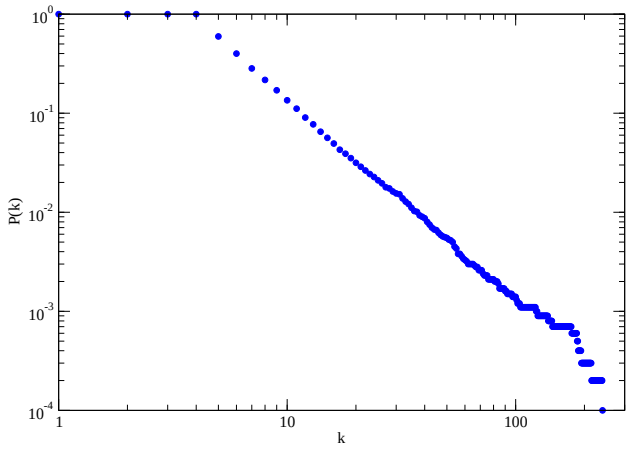

Figure 2.3: Cumulative degree distribution for a SF with $N = 10000$, constructed according to the BA model. For each node entering the network, 3 new edges are placed. The horizontal axis is vertex degree k and the vertical axis is the cumulative probability distribution of degrees, i.e., the fraction of vertices that have degree greater than or equal to k.

The average distance in the BA model is smaller than in a ER-random graph with same N, and increases logarithmically with N. Analytical results predict a double logarithmic correction to the logarithmic dependence $L \sim \frac{logN}{log(logN)}$. The clustering coefficient vanishes with the system size as $C \sim N^{-0.75}$. This is a slower decay than that observed for random graphs, $C \sim \langle k \rangle N^{-1}$, but it is still different from the behavior in small-world models, where C is independent of N.

b. Other SF models The BA model has attracted an exceptional amount of attention in the literature. In addition to analytic and numerical studies of the model itself, many authors have proposed modifications and generalizations to make the model a more realistic representation of real networks. Various generalizations, such as models with nonlinear preferential attachment, with dynamic edge rewiring, fitness models and hierarchically and deterministically growing models, can be found in the literature. Such models yield a more flexible value of the exponent γ which is restricted to $\gamma = 3$ in the original BA construction. Furthermore, modifications to reinforce the clustering property, which the BA model lacks, have also been consid-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

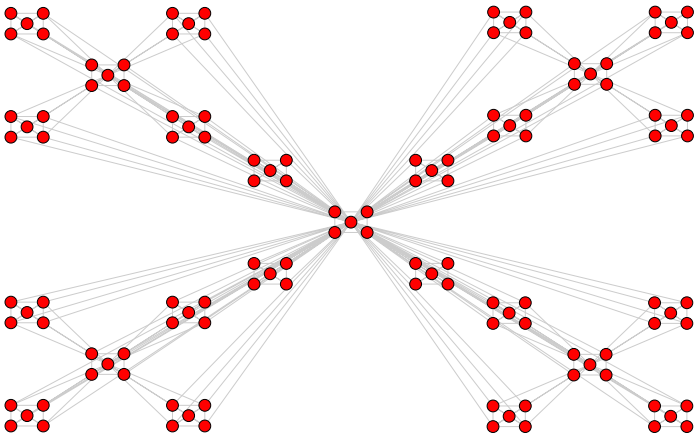

Figure 2.4: The RB model yields a hierarchical network, that combines the scale-free property with a high degree of clustering. The starting point is a small cluster of five densely linked nodes; next, four replicas of this hypothetical module are generated. The four external nodes of the replicated clusters are connected to the central node of the old cluster, obtaining a large 25-node module. This replication and connection can be repeated recursively, thus obtaining networks of size 25, 125, etc.

ered.

Among these alternative models we can find the *Dorogovtsev–Mendes–Samukhin* (DMS) model, which considers a linear preferential attachment; or the *Ravasz–Barabási* (RB) model, which aims at reproducing the hierarchical organization observed in some real systems (this makes it useful as an appropriate benchmark for multi-resolution community detection algorithms, see next Section and Figure 2.4).

The *Klemm–Eguiluz* (KE) model seeks to reproduce the high clustering coefficient usually found in real networks, which the BA model fails to reproduce [96]. To do so, it describes the growth dynamics of a network in which each node of the network can be in two different states: active or inactive. The model starts with a complete graph of m active nodes. At each time step, a new node j with m outgoing links is added. Each of the m active nodes receives one incoming link from j. The new node j is then activated, while one of the m active nodes is deactivated. The probability Π_i^{deact} that node i is deactivated is given by

$$\Pi_i^{deact} = \frac{1}{k_i + a} \left(\sum_{l \in N_{act}} \frac{1}{k_l + a} \right)^{-1} \tag{2.14}$$

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

where k_i is the in-degree of node i, a is a positive constant and the summation runs over the set N_{act} of the currently active nodes. The procedure is iteratively repeated until the desired network size is reached. The model produces a scale-free network with $\gamma = 2 + a/m$ and with a clustering coefficient $C = 5/6$ when $a = m$. Since the characteristic path length is proportional to the network size ($L \sim N$) in the KE model, additional rewiring of edges is needed to recover the small-world property. Reference [21] thoroughly discusses these and other models.

2.4 The mesoscale level

Research on networks cannot be solely the identification of actual systems that mirror certain properties from formal models. Therefore, the network approach has necessarily come up with other tools that enrich the understanding of the structural properties of graphs. The study of networks (or the methods applied to them) can be classified in three levels:

- The study at the *micro* level attempts to understand the behavior of single nodes. Such level includes degree, clustering coefficient or betweenness and other parameters.
- *Meso* level points at group or community structure. At this level, it is interesting to focus on the interaction between nodes at short distances, or classification of nodes, as we shall see.
- Finally, *macro* level clarifies the general structure of a network. At this level, relevant parameters are average degree $\langle k \rangle$, degree distribution $P(k)$, average path length L, average clustering coefficient C, etc.

The first and third levels of topological description range from the microscopic to the macroscopic description in terms of statistical properties of the whole network. Between these two extremes we find the mesoscopic level of analysis of complex networks. In this level we describe an inhomogeneous connecting structure composed by subsets of nodes which are more densely linked, when compared to the rest of the network.

This mesoscopic scale of organization is commonly referred as *community structure*. It has been observed in many different contexts, including metabolic networks, banking networks or the worldwide flight transportation network [72]. Moreover, it has been proved that nodes belonging to a tight-knit community are more than likely to have other properties in common. For instance, in the world wide web community analysis has uncovered thematic clusters.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Whatever technique applied, the belonging of a node to one or another community cannot depend upon the "meaning" of the node, i.e. it can't rely on the fact that a node represents an agent (sociology), a computer (the internet), a protein (metabolic network) or a word (semantic network). Thus communities must be determined solely by the topological properties of the network: nodes must be more connected within its community than with the rest of the network. Whatever strategy applied, it must be *blind* to content, and only *aware* of structure.

The problem of detection is particularly tricky and has been the subject of discussion in various disciplines. In real complex networks there is no way to find out, *a priori*, how many communities can be discovered, but in general there are more than two, making the process more costly. Furthermore, communities may also be hierarchical, that is communities may be further divided into sub-communities and so on [11, 73, 83]. Summarizing, it is not clear at what point a community detection algorithm must stop its classification, because no prediction can be made about the right level of analysis.

2.4.1 Community detection

A simple approach to quantify a given configuration into communities that has become widely accepted was proposed in [125]. It rests on the intuitive idea that random networks do not exhibit community structure. Let us imagine that we have an arbitrary network, and an arbitrary partition of that network into N_c communities. It is then possible to define a N_c x N_c size matrix e where the elements e_{ij} represent the fraction of total links starting at a node in partition i and ending at a node in partition j. Then, the sum of any row (or column) of e, $a_i = \sum_j e_{ij}$ corresponds to the fraction of links connected to i. If the network does not exhibit community structure, or if the partitions are allocated without any regard to the underlying structure, the expected value of the fraction of links within partitions can be estimated. It is simply the probability that a link begins at a node in i, a_i, multiplied by the fraction of links that end at a node in i, a_i. So the expected number of intra-community links is just $a_i a_i$. On the other hand we know that the real fraction of links exclusively within a community is e_{ii}. So, we can compare the two directly and sum over all the communities in the graph.

$$Q = \sum_j (e_{ii} - a_i)^2 \tag{2.15}$$

This is a measure known as *modularity*. Equation 2.15 has been extended to a directed and weighted framework, and even to one that admits negative weights [76]. Designing algorithms which optimize this value yields good community structure

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

compared to a null (random) model. The problem is that the partition space of any graph (even relatively small ones) is huge (the search for the optimal modularity value seems to be a NP-hard problem due to the fact that the space of possible partitions grows faster than any power of the system size), and one needs a guide to navigate through this space and find maximum values. Some of the most successful heuristics are outlined in [53, 124]. The first one relies on a genetic algorithm method (Extremal Optimization), while the second takes a greedy optimization (hill climbing) approach. Also, there exist methods to decrease the search space and partially relieve the cost of the optimization [10]. In [49] a comparison of different methods is developed, see also [75].

Modularity-based methods have been extended to analyze the community structure at different resolution levels, thus uncovering the possible hierarchical organization of the mesoscale [9, 11, 12].

2.4.2 Node functional role

It is important to keep in mind that one of the interesting applications of community detection is a better understanding of the position (importance) of a node, at a local, modular and global level. Betweenness, degree, etc. (see Section above) already point at this question. However, developing techniques that give more accurate information about it is important to enrich the understanding of the network structure.

When considering modular networks, it is plausible to surmise that the nodes in a network are connected according to the role they fulfill. For example, in a classical hierarchical organization, the chief executive is not directly connected to plant employees but is connected to the members of the board of directors. Such statement holds for virtually any organization; that is, the role of chief executive is defined irrespective of the particular organization considered.

Recently, Guimerà et al. [82] advanced on this issue proposing two descriptors to characterize the modular structure: the z-score (a measure of the number of standard deviations a data point is from the mean of a data set) of the internal degree of each node in its module, and the participation coefficient (P) defined as how the node is positioned in its own module and with respect to other modules. Given a certain partition, the plot of nodes in the z–P plane admit an heuristic tagging of nodes' role. The success of this representation relies on a consistent interpretation of topological roles of nodes, i.e. GA approach is based on the idea that nodes with the same role should have similar topological properties.

The *within-module degree* and the *participation coefficient* are easily computed once the modules of a network are known.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

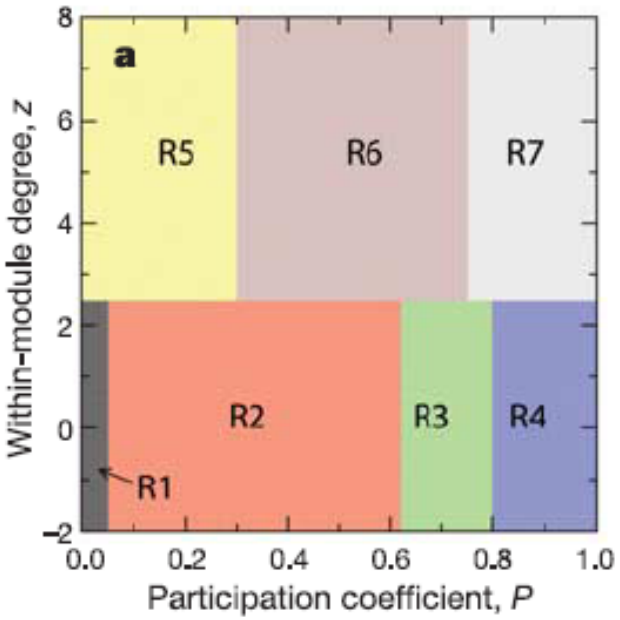

Figure 2.5: Each node in a network can be characterized by its within-module degree and its participation coefficient. GA classify nodes with $z \geq 2.5$ as module hubs and nodes with $z \leq 2.5$ as non-hubs. Non-hub nodes can be assigned into four different roles: (R1) ultra-peripheral nodes; (R2) peripheral nodes; (R3) non-hub connector nodes; and (R4) non-hub kinless nodes. Hub nodes can be naturally assigned into three different roles: (R5) provincial hubs; (R6) connector hubs; and (R7) kinless hubs. After [82].

Given a node i, high values of z_i indicate high within-module degrees and vice versa. Being N_c the number of modules in a network, the participation coefficient P_i reaches a maximum value of $P_i = 1 - \frac{1}{N_c}$ when its links are uniformly distributed among all the modules, and 0 if all its links are within its own module.

If κ_i is the number of links of node i to other nodes in its module s_i, $\overline{\kappa}_{s_i}$ is the average of κ over all the nodes in s_i, and $\sigma_{\kappa_{s_i}}$ is the standard deviation of κ in s_i, then:

$$z_i = \frac{\kappa_i - \overline{\kappa}_{s_i}}{\sigma_{\kappa_{s_i}}} \tag{2.16}$$

is the so-called *z-score*. The participation coefficient P_i of node i is defined as:

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

$$P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{\kappa_{is}}{k_i} \right)^2 \tag{2.17}$$

where κ_{is} is the number of links of node i to nodes in module s, and k_i is the total degree of node i. From these definitions, GA define seven different universal roles, each defined by a different region in the z–P parameter space (see Figure 2.5). According to the within-module degree, nodes are classified in the next way:

- Non-hubs:
  1. (R1) ultra-peripheral nodes; that is, nodes with all their links within their module ($P \leq 0.05$);
  2. (R2) peripheral nodes; that is, nodes with most links within their module ($0.05 < P \leq 0.62$);
  3. (R3) non-hub connector nodes; that is, nodes with many links to other modules ($0.62 < P \leq 0.80$);
  4. (R4) non-hub kinless nodes; that is, nodes with links homogeneously distributed among all modules ($P > 0.80$).
- Hubs:
  1. (R5) provincial hubs; that is, hub nodes with the vast majority of links within their module ($P \leq 0.30$);
  2. (R6) connector hubs; that is, hubs with many links to most of the other modules ($0.30 < P \leq 0.75$);
  3. (R7) kinless hubs; that is, hubs with links homogeneously distributed among all modules ($P > 0.75$).

The analysis of nodes' role by GA can be complemented with the study of at the modular level, taking into account both their internal structure and their interrelations. Such analysis is developed in the next section.

2.4.3 An optimal map of the modular structure of complex networks

This Section describes a set of related tools to screen the modular structure. The comprehension of modular structure in networks necessarily demands the analysis of the contribution of each one of its constituents (nodes) to the modules.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

The method is based on linear projection theory, to study the modular structure in networks that enables a systematic analysis and elucidation of its skeleton. The method yields an optimal mapping of the information of the modular structure (in the sense of least squares) in a two-dimensional space. As an example, the method has been applied to two empirical networks. The statistical analysis of the geometrical projections allow to characterize the structure of individual modules and their interrelations in a unified framework. Application to both synthetic and empirical networks can be seen in [7].

a. Projection of the modular structure A complex network (weighted or unweighted, directed or undirected) can be represented by its graph matrix $\boldsymbol{W}$, whose elements W_{ij} are the weights of the connections from any node i to any node j. Assuming that a certain partition of the network into modules is available, this coarse grained structure can be analyzed (the partition can be obtained by any method). The main object of the analysis is the *Contribution matrix* $\boldsymbol{C}$, of N nodes to M modules. The rows of $\boldsymbol{C}$ correspond to nodes, and the columns to modules. The elements $C_{i\alpha}$ are the number of links that node i dedicates to module α, and can be easily obtained as the matrix multiplication between W_{ij} and the *partition matrix* $\boldsymbol{S}$:

$$C_{i\alpha} = \sum_{j=1}^{N} W_{ij} S_{j\alpha} \tag{2.18}$$

where $S_{j\alpha} = 1$ if node j belongs to module α, and $S_{j\alpha} = 0$ otherwise. The goal is to reveal the structure of individual modules, and their interrelations, from the matrix $\boldsymbol{C}$. To this end, the method deals with the high dimensionality of the original data by constructing a two-dimensional map of the contribution matrix, minimizing the loss of information in the dimensional reduction, and making it more amenable to further investigation.

b. Singular Value Decomposition of the modular structure The approach consists in the analysis of $\boldsymbol{C}$ using Singular Value Decomposition [74] (SVD). It stands for the factorization of a rectangular N-by-M real (or complex) matrix as follows:

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\dagger} \tag{2.19}$$

where $\boldsymbol{U}$ is an unitary N-by-N matrix, $\boldsymbol{\Sigma}$ is a diagonal N-by-M matrix and $\boldsymbol{V}^{\dagger}$ denotes the conjugate transpose of $\boldsymbol{V}$, an M-by-M unitary matrix. This decomposition corresponds to a rotation or reflection around the origin, a non-uniform scale represented by the *singular values* (diagonal elements of $\boldsymbol{\Sigma}$) and (possibly) change in

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

the number of dimensions, and finally again a rotation or reflection around the origin. This approach and its variants have been extraordinarily successful in many applications [74], in particular for the analysis of relationships between a set of documents and the words they contain. In this case, the decomposition yields information between word-word, word-document, and document-document semantic associations, the technique is known as Latent Semantic Indexing [17], and Latent Semantic Analysis [99], see Chapter 4. The scenario here is quite similar to this, where nodes resemble words, and modules resemble documents.

c. An optimal 2D map of the modular structure of networks A practical use of SVD is dimensional reduction approximation, also known as Truncated Singular Value Decomposition (TSVD). It consists in keeping only some of the largest singular values to produce a least squares optimal, lower rank order approximation (see Appendix). In the following we will consider the best approximation of $\boldsymbol{C}$ by a matrix of rank $r = 2$.

The main idea is to compute the projection of the contribution of nodes to a certain partition (rows of $\boldsymbol{C}$, namely $\boldsymbol{n}_i$ for the i-th node) into the space spanned by the first two left singular vectors, the projection space $\mathcal{U}_2$ (see Appendix). The projected contribution of the i-th node is denoted as $\tilde{\boldsymbol{n}}_i$. Given that the transformation is information preserving [36], the map obtained gives an accurate representation of the main characteristics of the original data, visualizable and, in principle, easier to scrutinize. Note that the proposed approach has essential differences with classical pattern recognition techniques based on TSVD such as Principal Components Analysis (PCA) or, equivalently, Karhunen-Loeve expansions. In this case data (columns of $\boldsymbol{C}$) can not be independently shifted to mean zero without loosing its original meaning, this restriction prevents the straightforward application of the mentioned techniques.

To obtain and correctly interpret the outcome of SVD it is worth pointing out the following geometrical properties of the projection of the rows of $\boldsymbol{C}$ we have defined (see Appendix for a mathematical description):

1. Every module α has an intrinsic direction $\tilde{\boldsymbol{e}}_\alpha$ in the projection space $\mathcal{U}_2$ corresponding to the line of the projection of its internal nodes (those that have links exclusively inside the module). These directions are *intramodular projections*. This property is essential to discern among modules that are cohesive, in the sense that the majority of nodes project in this direction, from those modules which are not.

2. Every module α has a distinguished direction $\tilde{\boldsymbol{m}}_\alpha$ in the projection space $\mathcal{U}_2$ corresponding to the vector sum of the contributions of all its nodes. These

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

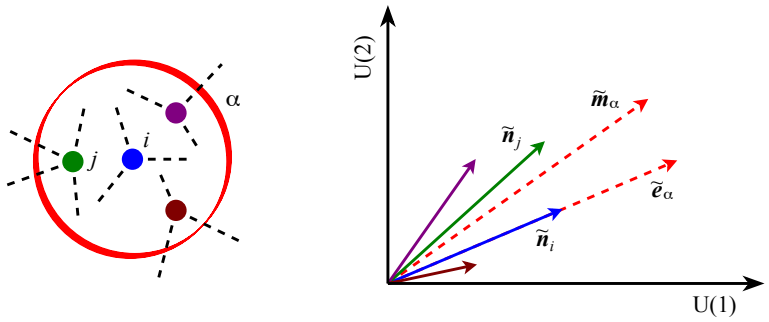

Figure 2.6: Geometrical scheme of the TSVD. The intramodular projection of module α, $\tilde{\boldsymbol{e}}_\alpha$ is the direction where all internal nodes lay (in the plot node i). The node contribution projections $\tilde{\boldsymbol{n}}$ are represented by vectors in different colors. Finally, the modular projection $\tilde{\boldsymbol{m}}_\alpha$ is computed as the vector sum of all the node contribution projections belonging to it. Note that the intramodular projection and the modular projection do not coincide, the differences between both inform about the cohesiveness of the module.

directions are named *modular projections*. The modular projection is relevant when compared to the intramodular projection because their deviations inform about the tendency to connect with other modules. Note that e_α and m_α are equal only if the module is disconnected from the rest of the network.

3. Any node contribution projection $\tilde{\boldsymbol{n}}_i$ is a linear combination of intramodular projections, being the coefficient of each one proportional to the original contribution $C_{i\alpha}$ of links of the node i to each module α. This property comes from the linearity of the projection, and expresses the contribution of nodes to the modules to which they are connected to.

Consequently, from (i) and (iii), we can classify nodes. Nodes with only internal links have a distance to the origin proportional to its degree (or strength). Nodes with internal and external links, separate from the intramodular projection proportionally to their contributions to other modules. From (ii) we can classify modules. Modules that have close modular projections are more interrelated. These geometrical facts are the key to relate the outcome of TSVD and the original data in our problem, see Figure 2.6.

d. Structure of individual modules Keeping in mind the geometrical properties (i) and (iii) exposed above, the structural information relative to each module is

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

extracted by comparing the map of nodes' contributions to the intramodular projection directions. To this end it is convenient to change to polar coordinates, where for each node i the radius R_i measures the length of its contribution projection vector $\tilde{\boldsymbol{n}}_i$, and θ_i the angle between $\tilde{\boldsymbol{n}}_i$ and the horizontal axis. We also define ϕ_i as the absolute distance in angle between $\tilde{\boldsymbol{n}}_i$ and the intramodular projection $\tilde{\boldsymbol{e}}_\alpha$ corresponding to its module α, i.e. $\phi_i = |\theta_i - \theta_{\tilde{\boldsymbol{e}}_\alpha}|$.

Using these coordinates R–ϕ we find a way to interpret correctly the map of the contribution matrix in $\mathcal{U}_2$: i) $R_{int} = R\cos\phi$ informs about the internal contribution of nodes to its corresponding module, as well as to the contribution to its own module by connecting to others. To clarify the latter assertion, let us assume a node i belonging to a module β has connections with the rest of modules in the network. Given that this connectivity pattern is a linear combination of intramodular directions $\tilde{\boldsymbol{e}}_\alpha$, the vector sum implies that connecting with modules α having $|\theta_{\tilde{\boldsymbol{e}}_\beta} - \theta_{\tilde{\boldsymbol{e}}_\alpha}| > \pi/2$ decreases the module R, and vice versa. ii) $R_{ext} = R\sin\phi$ informs about the deviation (as the orthogonal distance) of each node to the contribution to its own module, see Figure 2.7.

Following [7], the internal structure of modules can be explored using the values of R_{int}, and the boundary structure of modules using R_{ext}. Using descriptive statistics one can reveal and compare the structure of individual modules. Provided that the distribution of contributions is not necessarily Gaussian, an exploration in terms of z-scores is not convenient. Instead box-and-whisker charts are used, depicting the principal quartiles and the outliers (defined as having a value more than 1.5 IQR lower than the first quartile or 1.5 IQR higher than the third quartile, where IQR is the Inter-Quartile Range).

The boxplots for the data of each module in the variable R_{int} allow for a visualization of the heterogeneity in the contribution of nodes building their corresponding modules, and an objective determination of distinguished nodes on its structure (outliers). Consequently, the boxplots in R_{ext} inform about the heterogeneity in the boundary connectivity. Nodes with links in only one module are not considered in this statistics because they do not provide relevant information about the boundaries (they have $\phi = 0$), only nodes that act as bridges between modules are taken into account. Assuming that every module devotes some external links (otherwise they would be disconnected), the width of the boxes in this plot is proportional to the heterogeneity of such efforts. If only one node makes external connections, then the boxplot has zero width. Moreover, given two boxes equally wide, their position (median) determines which module contributes more to keeping the whole network connected.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

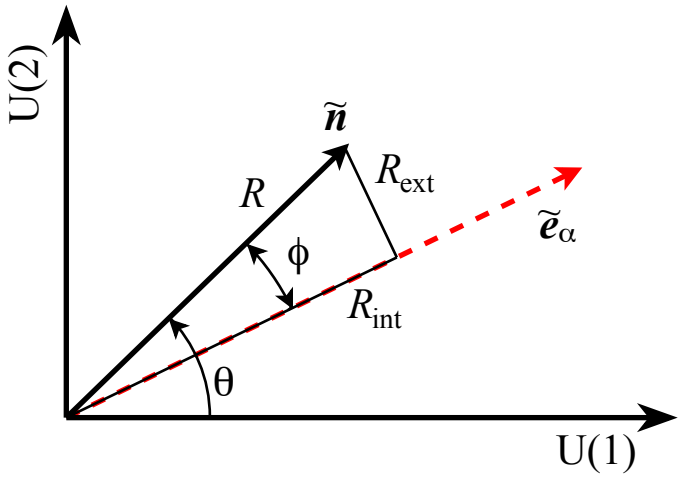

Figure 2.7: Schematic plot of the coordinates proposed to study the structure of individual modules. The relative distance of a node from its module is captured by the angle ϕ. The respective components R_{int} and R_{ext} are depicted.

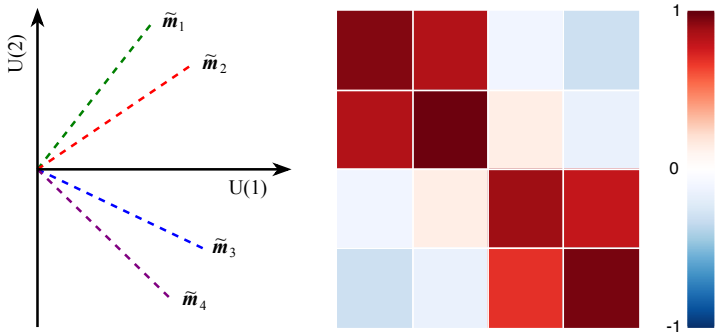

Figure 2.8: Schematic plot of the interrelation between the modular projections of 4 modules. The matrix represents the overlap computed as the scalar product between directions.

e. Interrelations between modules The analysis of the interrelations between modules is performed at the coarse grained level of its modular projections. The modular projections $\tilde{m}_\alpha$ are aggregated measures of the nodes' contribution to their particular module. The normalized scalar product of modular projections provide

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

a measure of the interrelations (overlapping) between different modules. A representation of these data in form of a matrix ordered by the values of $\theta_{\tilde{m}_\alpha}$ reveals the actual skeleton of the network at the topological mesoscale, see Figure 2.8.

The method allows the access to a tractable map of the empirical complex network data according to a biological, functional or topological partition. The analysis of this map can help to anticipate the scope of dynamic emergent phenomena that depends on the structure and relations between modules. Furthermore, the method can be extended to graph bipartitioning by adaptively changing nodes between two modules while maximizing the angle in the R–θ plane between them. Further studies of the similarities between nodes' contribution projections can also help to classify networks according to the role profiles of nodes [148] and/or modules.

Appendix

Properties of TSVD

Let us assume that we preserve only the r largest singular values and neglect the remaining substituting their value by zero, then the reduced matrix $\boldsymbol{C_r} = \boldsymbol{U}\boldsymbol{\Sigma_r}\boldsymbol{V}^\dagger$ has several mathematical properties worth to mention: first, it minimizes the Frobenius norm ($\|\boldsymbol{A}\|_F = \sqrt{\text{trace}(\boldsymbol{A}\boldsymbol{A}^\dagger)}$) of the difference $\|\boldsymbol{C} - \boldsymbol{C_r}\|_F$, that means that among all possible matrices of rank r, $\boldsymbol{C_r}$ is the best approximation in a least squares sense; second, $\boldsymbol{C_r}$ is also the best approximation in the sense of statistics, it maintains the most significant information portion of the original matrix [36]. The left and right singular vectors (from matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively) capture invariant distributions of values of the contribution of nodes to the different modules. In particular the larger the singular value the more information represented by their corresponding left and right singular vectors. We have used the LAPACK-based implementation of SVD in MATLAB. We warn that some numerical implementations of SVD suffer from a sign indeterminacy, in particular the one provided by MATLAB is such that the first singular vectors from an all-positive matrix always have all-negative elements, whose sign obviously should be switched to positive [30].

Projection using TSVD of rank 2

In the case of a rank $r = 2$ approximation, the unicity of the two-ranked decomposition is ensured [74] if the ordered singular values σ_i of the matrix $\boldsymbol{\Sigma}$, satisfy $\sigma_1 > \sigma_2 > \sigma_3$. This dimensional reduction is particularly interesting to depict results in a two-dimensional plot for visualization purposes. In the new space there are two different sets of singular vectors: the left singular vectors (columns of matrix

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

$\boldsymbol{U}$), and the right singular vectors (rows of matrix $\boldsymbol{V}^{\dagger}$). Given that we truncate at $r = 2$, we fix our analysis on the two first columns of $\boldsymbol{U}$, we call this the projection space $\mathcal{U}_2$. The coordinates $\tilde{\boldsymbol{n}}_i$ of the projection of the contributions $\boldsymbol{n}_i$ of node i are computed as follows:

$$\tilde{\boldsymbol{n}}_i = {\boldsymbol{\Sigma_2}}^{-1}\boldsymbol{V}^{\dagger}\boldsymbol{n}_i \tag{2.20}$$

Here ${\boldsymbol{\Sigma_2}}^{-1}$ denotes the pseudo-inverse of the diagonal rectangular matrix $\boldsymbol{\Sigma_2}$ (singular values matrix truncated in 2 rows), simply obtained by inverting the values of the diagonal elements. It is possible to assess the loss of information of this projection compared to the initial data by computing the relative difference between the Frobenius norms:

$$E_r = \frac{\|\boldsymbol{C}\|_F - \|\boldsymbol{C_r}\|_F}{\|\boldsymbol{C}\|_F} = \frac{\sum_{\alpha=1}^{M}\sigma_\alpha^2 - \sum_{\alpha=1}^{r}\sigma_\alpha^2}{\sum_{\alpha=1}^{M}\sigma_\alpha^2} \tag{2.21}$$

Geometrical properties of the projection of $\boldsymbol{C}$

The intramodular projection $\tilde{\boldsymbol{e}}_\alpha$ corresponding to module α, is defined as the projection of the cartesian unit vector $\boldsymbol{e}_\alpha = (0, \ldots, 0, 1, 0, \ldots, 0)$ (the α-th component is 1, the rest are zero), i.e.

$$\tilde{\boldsymbol{e}}_\alpha = {\boldsymbol{\Sigma_2}}^{-1}\boldsymbol{V}^{\dagger}\boldsymbol{e}_\alpha \tag{2.22}$$

Any node in the original contribution matrix can be represented as

$$\boldsymbol{n}_i = \sum_{\alpha=1}^{M} C_{i\alpha}\boldsymbol{e}_\alpha \tag{2.23}$$

Its projection gives the node contribution projection

$$\tilde{\boldsymbol{n}}_i = \sum_{\alpha=1}^{M} C_{i\alpha}({\boldsymbol{\Sigma_2}}^{-1}\boldsymbol{V}^{\dagger}\boldsymbol{e}_\alpha) = \sum_{\alpha=1}^{M} C_{i\alpha}\tilde{\boldsymbol{e}}_\alpha \tag{2.24}$$

a linear combination of intramodular projections. In particular, a node i whose contribution is totally internal to a module α is projected as $\tilde{\boldsymbol{n}}_i = k_i\tilde{\boldsymbol{e}}_\alpha$, where k_i is the node degree. The modular projections $\tilde{\boldsymbol{m}}_\alpha$ are computed as the vector sum of all the projections of nodes contributions, for those nodes belonging to module α, i.e.

$$\tilde{\boldsymbol{m}}_\alpha = \sum_{i=1}^{N} S_{i\alpha}\tilde{\boldsymbol{n}}_i \tag{2.25}$$

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

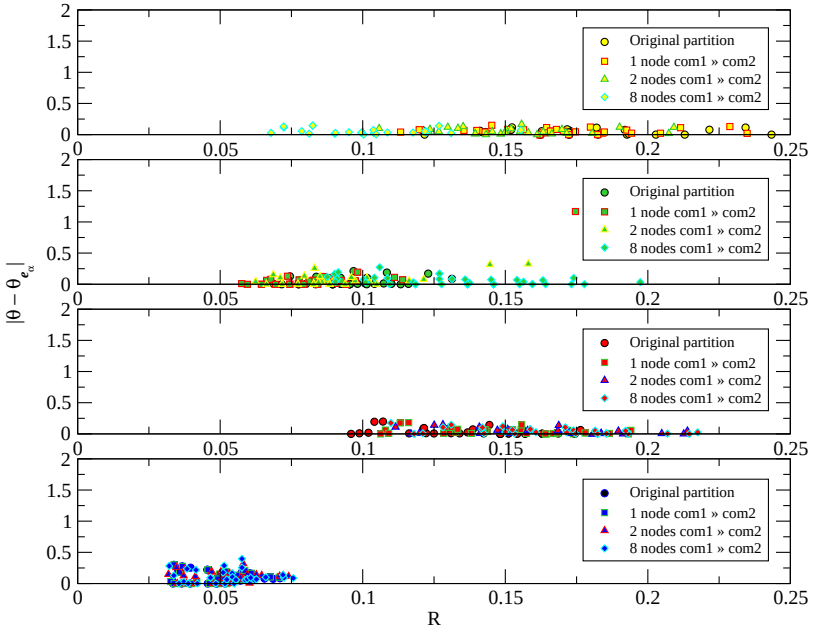

Figure 2.9: Robustness of the method to noise in the partition. We show the separation from the intramodular directions of modules 1 to 4 (top to down) of all nodes, in particular we track the deviation of the nodes when some of them have been assigned to the incorrect module. The nodes that have been moved are those that deviate more from the intramodular projection of module 2.

Effect of noise on C

The method presented is pretty robust to perturbations in the partition or, equivalently, in the contribution matrix $\boldsymbol{C}$. To support the claim we make the following experiment: using the benchmark network proposed by Newman and Girvan [72]. With 128 nodes, $z_{in} = 15$ and $z_{out} = 1$, we perform slight changes in the predefined partition, by moving nodes from module 1 to module 2. First we move only one node, then two nodes, and finally 8 nodes. This changes matrix $\boldsymbol{C}$, which must in turn affect TSVD output. Fig. 2.9 contains the nodes' projection as the mentioned movements take place (squares, triangles and diamonds respectively). Consistently, module 1's nodes projections progressively decrease in R. Module 2 balances this

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

fact, it retains the weight leaving from module 1. Sensitivity to inter-modular connections is also evidenced: when a single new node appears in module 2 (Fig. 2.9, squares), ϕ_i has an outstanding value if compared to the rest; this is also evident when two nodes enter group 2 (Fig. 2.9, triangles). When moving 8 nodes, the effect is less drastic for the deviations in θ and more drastic in R. Unsurprisingly, modules 3 and 4 remain mostly unchanged, the interplay between modules 1 and 2 (nodes leaving from one group towards the other) does not drastically affect their internal characteristics, nor their importance in the whole structure.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Chapter 3

Lexical and Semantic Networks: Structure and Dynamics

Language is surely a strikingly complex object to study. It is built upon many thousands of items (WordNet database [115], for instance, currently gathers over 200,000 words in English), which in turn are often ambiguous. Also, they hold manifold relationships with each other, sometimes for the sake of structure (syntax), sometimes for the sake of their meaning (semantics). Words, and the relationships they hold, evolve across time, new words appear, some of them die, some simply change. As difficult as it may be to study or model these facts, complexity is even greater when language is placed in a cognitive context: linguistic production and comprehension processes occur proficiently in the mind while many other processes are concurrently interacting. Consider, among them, actions from the sensorimotor system (vocalization), the perceptual system (listening, reading) or memory (retrieval, recall and recognition). Finally, a last step to complexity is to consider linguistic performance as a result of neural activity. Language, thus, is a complex object efficiently managed in a complex mental context, which in turn is embodied in the most complex known system, the brain.

Linguistics and psycholinguistics devote much efforts to disentangle the details of the aforementioned facts. However, some fundamental questions can not be addressed from this fine-grained perspective: what is the general structure of language? Is such structure common to every language? Can we describe the general trends of the mechanisms that provide for linguistic efficient performance? Is it possible to describe the principles of language growth (from a child to an adult)? Such questions demand a complementary point of view from that of linguistics and psycholinguistics, one that abstracts and simplifies as much as possible the intricate nature of language. This general view makes the minimum assumptions, in the end language is reduced to a set of entities which are related with each other. Following this line, cognitive processes are characterized as phenomena occurring on top of that structure. These processes are conceived as naïve mechanisms.

As it has been argued in Chapter 1, and implicitly all along the text, the basics of this viewpoint fit naturally in complex systems approach.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Research on language include syntax, prosody, semantics, neuroscience, etc. Some of them deal with physical observables but are not suitably approached from a statistical physics point of view yet (to our best knowledge). That is the case of prosody, which tries to extract useful linguistic information from the loudness, pitch or frequency of language sounds. Others, like syntax, have been subject of study from a network perspective, for example by dealing with syntactic trees understood as graphs. Although this latter line has received much attention [43, 44, 60, 64, 65, 159] (or rather, *because* of it), it constitutes a line of research of its own and lies outside the scope of this work. Then the natural framework of this overview is semantics at the lexical (word) level and some adjacent phenomena (lexicon formation and change). This means that works devoted to linguistic superstructures (phrases and sentences) are not considered; neither are sub-lexical units (lemmas, phonemes, etc.), although there also exists some work on them in the complex systems bibliography [95, 180].

The Chapter is organized as follows: Section 3.1 introduces the question of data acquisition and network construction, pointing some sources that have been used to build up language networks and what interpretation they should receive. The next three Sections are devoted to (i) characterization of language: the organization of language is characterized in terms of general network structural principles (Section 3.2) ; (ii) cognitive growth and development: we attempt to reveal how structural features reflect general processes of language acquisition (Section 3.3); and (iii) cognitive processes: a few models that relate human performance in semantic processing tasks with processes operating on complex networks are presented (Section 3.4).

3.1 Building language networks

An often expressed concern around complex networks is their arbitrary character. When modeling actual, real-world systems using network methodology, the researcher needs to take some decisions: what kind of object must be understood as a vertex, in the first place; and more critical, what must be understood as a link between vertices. In our case, it is not straightforward to define the notion of word interaction in a unique way. For instance, one can connect the nearest neighbors in sentences. Also, one could take into account linguistic standard relations, like synonymy, hyper-or hyponymy, etc. Finally, one can assemble networks out of laboratory data, i.e. data coming from experiments with subjects in psycholinguistics. We detail these three lines in the subsequent paragraphs, closely following the ideas in [111].

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

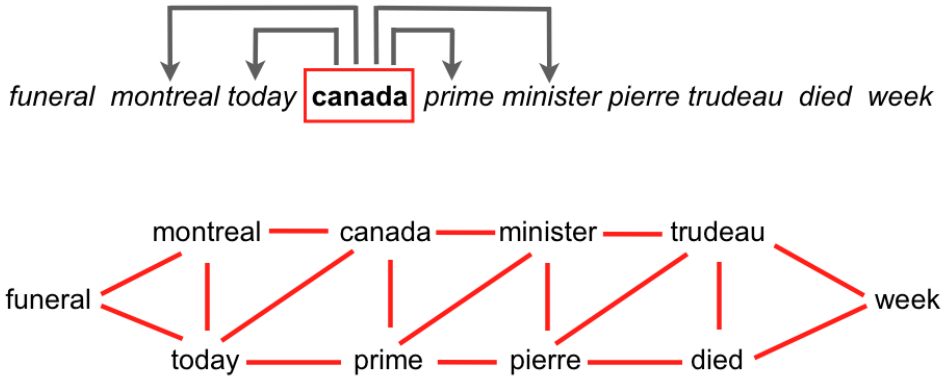

Figure 3.1: Fixed-size window scheme: in this case words are linked to $d = 2$ neighbors, where d is a constant window size. Note that words without semantic content have been removed, i.e. articles, prepositions, etc.

3.1.1 Text analysis: co-occurrence graphs

Intuitively, the simplest strategy to collect relations among entities is to construct a network whose topology reflects the co-occurrence of words. Such intuition is rooted in *collocation analysis*, a well established field of corpus linguistics [157, 172, 173]. It follows a tradition according to which collocations manifest lexical semantic affinities beyond grammatical restrictions [84].

Typically, text co-occurrence networks are obtained with the minimum assumptions and cost, i.e. a fixed adjacency window of width d is predefined, such that two words w_1 and w_2 are connected by an edge (link) if $d_{w_1-w_2} \leq d$. Thus, a *two-word* adjacency network automatically connects a word with any two words before and after it, see Figure 3.1. Often articles and other connecting words are excluded. Their topology quantified by several measurements can provide information on some properties of the text, such as style and authorship [6].

Some limitations must be taken into account under this constructive method: if d is long, the risk of capturing spurious co-occurrences increases. If d is too short, certain strong co-occurrences can be systematically left out [63].

The textual sources for these type of networks can be varied. In some cases a single source is chosen (for example, a book from a particular author). In other cases, collections of newspapers or magazines are used (as in the ACE *corpus* [42]). This subtle difference is important, in the first case the resulting structure reflects (at least partially) the lexical organization of an individual; whereas the latter pro-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

vides an access to the semantic collective system of a language, that is, to the overall organization of its lexical subsystem [156]. This distinction already points in two research poles, the more cognitive- and the more language-oriented, which shall appear later.

3.1.2 Dictionaries and Thesauri

As in the case of multi-source text analysis, again a collective view on language is predominant in the case of dictionaries. *Lexical reference systems* or *terminological ontologies* (e.g. WordNet, [115]), *thesauri* (e.g. Roget's thesaurus, [142]) and related systems build on expert knowledge of lexicographers in order to define *sense relations* (e.g. synonymy, antonymy, hyponymy) between words or conceptual relations between concepts (therefore, they are meaning-based). Following [111], in the case of *thesaurus* graphs based on the expertise of lexicographers and corpus linguists, the characteristics of the network can be interpreted as indicators of *thesaurus* quality or consistency. For instance, a graph representing hyponymy relations within a *thesaurus* should induce a hierarchical structure, whereas polysemy should provide for the small world nature of the semantic system of the language under consideration. Such is the case of Wordnet in the study by Sigman and Cecchi [156].

3.1.3 Semantic features

In many of the most influential theories of word meaning and of concepts and categorization, semantic features have been used as their representational currency. Numerous vector models of memory are based on feature representations. For this reason, the major purpose of collecting semantic feature production norms is to construct empirically derived conceptual representation and computation.

One of the most relevant example of such data collection is that of McRae *et al.* [110] Feature Production Norms (FPN), which will help us illustrate this data gathering technique. FPN were produced by asking subjects to conceptually recognize features when confronted with a certain word. This feature collection is used to build up a vector of characteristics for each word, where each dimension represents a feature. In particular, participants are presented with a set of concept names and are asked to produce features they think are important for each concept. Each feature stands as a vector component, with a value that represents its production frequency across participants. These norms include 541 living and nonliving thing concepts, for which semantic closeness or *similarity* is computed as the cosine (overlap) between pairs of vectors of characteristics. The cosine is obtained as the dot product between two concept vectors, divided by the product of their lengths:

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

$$\cos\theta = \frac{v_1 w_1 + v_2 w_2 + \ldots + v_n w_n}{\| v \| \| w \|} \tag{3.1}$$

As a consequence, words like *banjo* and *accordion* are very similar (i.e. they have a projection close to 1) because their vector representations show a high overlap, essentially provoked by their shared features as musical instruments, while the vectors for *banjo* and *spider* are very different, showing an overlap close to 0 (almost orthogonal vectors).

In terms of network modeling, each node represents a word, and an edge (or link) is set up between a pair of nodes whenever their vectors projection is different from 0 (or above a predefined threshold τ). The meaning of an edge in this network is thus the features similarity between two words. The network is undirected (symmetric relationships) and weighted by the value of the projections. See Figure 3.2 for illustration.

Although these measures are not obtained from an individual, but rather averaged out of many participants in an experiment, this type of data is in the line of cognitive research, in which network modeling is a tool to understand actual mechanisms of human language usage. The same can be said regarding associative networks, in the next subsection.

3.1.4 Associative networks

Association graphs are networks in which vertices denote words, whereas links represent association relations as observed in cognitive-linguistic experiments. Such graphs are considered the most relevant from a psychological point of view. According to the hypothesis that association is one of the principles of memory organization, the question that has to be addressed is which network topologies support an efficient organization in terms of time and space complexity.

The best known Free Association data set in English are University of South Florida Free Association Norms (USF-FA from now on; [122]). Nelson *et al.* produced these norms by asking over 6000 participants to write down the first word (*target*) that came to their mind when confronted with a *cue* (word presented to the subject). The experiment was performed using more than 5000 cues. Among other information, a frequency of coincidence between subjects for each pair of words is obtained. As an example, words *mice* and *cheese* are neighbors in this database, because a large fraction of the subjects related this target to this cue. Note, however, that the association of these two words is not directly represented by similar features but other relationships (in this case mice eat cheese). The network empirically obtained is directed and weighted. Weights represent the frequency of association

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

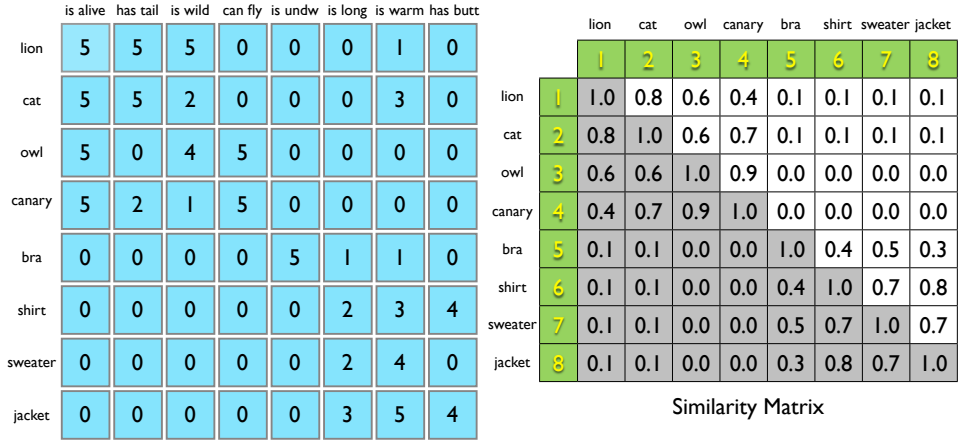

Figure 3.2: A network structure out of semantic features data. Left: each subject assigns semantic features to given nouns, and features build up a semantic vector. In the example, features are *is alive*, *has tail*, *is wild*, *can fly*, *is underwear*, *is long*, *is warm* and *has buttons*. The number in each cell reflects the number of participants who assigned that feature to the corresponding item. Right: cosine overlapping between each pair of vectors from the left matrix. This new similarity matrix can be suitably interpreted as a semantic network. Note that values in both matrices do not represent actual results, and have been put merely for illustrative purposes.

in the sample. These same data exist in Spanish [33, 58], German [112] or French [59].

Generally speaking, Free-Association Norms represent a more complex scenario than Feature Production Norms when considering the semantics of edges. Free-Association Norms are heterogeneous by construction, they may grasp any relation between words e.g. a causal-temporal relation (*fire* and *smoke*), an instrumental relation (*broom* and *floor*) or a conceptual relation (*bus* and *train*), among others.

From this data set, two networks can be created. A directed network, where two

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

word nodes i and j are joined by an arc (from i to j) if the cue i evoked j as an associative response for at least two of the participants in the database. In an undirected version, word nodes are joined by an edge if the words were associatively related regardless of associative direction. Although the directed network is clearly a more natural representation of word associations, most of the literature on small-world and scale-free networks has focused on undirected networks.

Next Sections attempt to review some works centered on network modeling of language. We will move gradually from the language-oriented pole, which is concerned with general structural patterns and dynamics of language as an object *per se*; towards the cognitive-oriented one, which is confronted with a greater degree of detail and complexity (language as an object-in-minds).

3.2 Lexical networks: topology, function, evolution

Soon after the seminal works by Watts and Strogatz, and Barabási and Albert in the late '90s, network scientists focused upon language as an interesting subject. Unsurprisingly, the insights were general in this initial stage, and became deeper from then on.

	N	$\langle k \rangle$	C	L
Moby Thesaurus	30244	59.9	0.53	3.16
Randomized MT	30244	59.9	0.002	2.5

Table 3.1: Results for the conceptual network defined by the Thesaurus dictionary, and a comparison with a corresponding random network with the same parameters. N is the total number of nodes, $\langle k \rangle$ is the average number of links per node, C is the clustering coefficient, and L is the average shortest path. After [120].

One of the first approaches is that of Motter *et al.* in [120], where the network structure of language is studied. The author presents results for the English language, which are expected to hold for any other language. A conceptual network is built from the entries in the Moby Thesaurus, and considers two words connected if they express similar concepts. Motter *et al.*'s resulting network includes over 30,000 nodes (words). Table 3.1 summarizes the results of the analysis. The random counterpart of the Moby Thesaurus [175] highlights the small-world features of the original, i.e. high clustering coefficient and low average path length.

Similarly, Sigman and Cecchi thoroughly characterized the WordNet database [156], with similar results. Their analysis of the degree distribution results in power-law distributions, the fingerprint of self-organizing, evolving systems.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

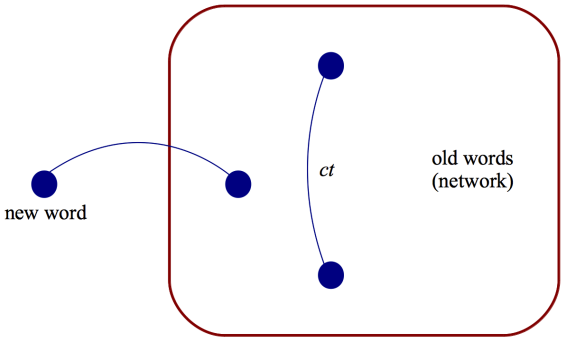

Figure 3.3: Dorogovstev and Mendes' scheme of the language network growth [52]: a new word is connected to some old one i with the probability proportional to its degree k_i (Barabási and Albert's preferential attachment); in addition, at each increment of time, ct new edges emerge between old words, where c is a constant coefficient that characterizes a particular network.

Dorogovstev and Mendes explore the mentioned possibility in [52], namely that language (or, more precisely, lexicon) is a self-organized growing system. Specifically, they discuss whether empirical degree distributions might be the result of some type of preferential attachment dynamics. The authors propose a stochastic theory of evolution of human language based on the treatment of language as an evolving network of interacting words. It is well known that language evolves, then the question is what kind of growth (in the sense of increase of lexical repertoire) leads to a self-organized structure? Although in the general framework of Barabási and Albert's preferential attachment, their proposal adds a second growth mechanism inspired in observations from real collaboration networks. This variation includes, at each time step, the appearance of new edges between already-existing (old) words, besides the birth of new words that link to some old ones (see Figure 3.3).

The model can be described in a precise analytical form. It is possible to detail the evolution of the degree of a certain word "born" at time s and observed at time t:

$$\frac{\partial k(s,t)}{\partial t} = (1+2ct)\frac{k(s,t)}{\int_0^t du\, k(u,t)} \tag{3.2}$$

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

The development of Equation 3.2 leads to a description of the evolution of the degree distribution $P(k,t)$, which matches the empirical findings in [63], i.e. a two-regime power-law with different exponents, see comments below.

Different language networks display as well similar small-world characteristics, see Table 3.2. Also, their degree distribution corresponds in some cases to scale-free networks, see Figure 3.4 (remarkably, the high interest in scale-free networks might give the impression that all complex networks in nature have power-law degree distributions. As is shown in the mentioned figure, this is far from being the case).

	N	$\langle k \rangle$	C	L	D
USF-FA	5018	22	0.1928	3.04	5
SFA-SV	7759	3.05	0.0867	3.71	5
SFA	2901	4.9	0.1487	4.50	8
GFA-SV	3632	2.05	0.034	4.57	8

Table 3.2: Some parameters obtained from four different data-sets: the University of South Florida word association (USF-FA, [122]), Free-association norms for the Spanish names of the Snodgrass and Vanderwart pictures (SFA-SV, [58]), association norms in Spanish (SFA, [33]) and association norms for the German names of the Snodgrass and Vanderwart pictures (GFA-SV, [112]). As the ones reported on Table 3.1, they all conform sparse structures with very low L (if compared to the size of the network). However, only USF-FA and SFA clearly fit in the small-world definition. Low C in the data sets based on the drawings from Snodgrass and Vanderwart [158] can be explained by the specific experimental setup with this material. N is the total number of nodes, $\langle k \rangle$ is the average number of links per node, C is the clustering coefficient, L is the average shortest path, and D is the diameter. The latter descriptors (L and D) have been measured from the undirected, unweighted networks of the data sets.

Most interestingly, these early results led to the claim that they have deep cognitive implications. From the standpoint of retrieval of information, the small-world property of the network represents a maximization of efficiency: high clustering gathers similar pieces of information, low distances makes fast search and retrieval possible. The expression "mental navigation" arises: irrespective of the specifics of the neuronal implementation, it can be thought that the small-world property is a desirable one in a navigation network (it strikes a balance between the number of active connections and the number of steps required to access any node); and, taking mental navigation for granted, one could also expect that the hubs of the network should display a statistical bias for priming in association and related tasks [156]. Navigation, in this context, corresponds to *retrieval* in semantic memory, understood as intentional recovery of a word. "Mental exploration" would instead correspond to search processes (such as when trying to produce words that begin

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

for a certain letter): there is no topological information to achieve this purpose in the network. In both processes shortcuts and hubs must significantly affect proficiency.

These intuitions probably point at the right direction, but there is a need to focus the attention on some specific phenomena. Then, since linguistic phenomena does not occur outside the boundaries of cognition, research necessarily turned towards the cognitive pole.

The work by Ferrer i Cancho and Solé represents significant steps in this direction. For instance, a difference is settled between single- and multi-author linguistic sources. In [63], a network $N \approx 5 \times 10^5$ words is built out of the British National Corpus (BNC). The degree distribution of such network evidences a two-regime power law, one of them with an average exponent close to the Barabási-Albert model ($\gamma_{BA} = -3$). From this twofold behavior the authors claim that the lexicon is divided into a set of core words (kernel, $\gamma = -2.7$) and a set of peripheral words ($\gamma = -1.5$). The kernel lexicon contains words that are common to the whole community of speakers, while in the periphery a certain word is unknown for one speaker and familiar for another. Results suggest that language has grown under the dynamics of preferential attachment, the core of the network (with $\gamma \approx \gamma_{BA}$) containing at least functional words, i.e. those with low or null semantic content. This approach takes into account not only the features of complex physical systems (self-organization, etc.), but how can they be explained in terms of collective behavior.

This "physical system–cognitive phenomena" mapping is again visible in [61, 62]. The question here is to give account of Zipf's least effort principle [188] using network methodology and information theory [155]. Again, the center of the discussion is a cognitive phenomenon (communication) in which a speaker and a listener are involved. As it is well known, word frequencies in human language obey a universal regularity, the so-called Zipfs law. If $P(f)$ is the proportion of words whose frequency is f in a text, we obtain $P(f) \propto f^{-\beta}$, with $\beta \in [1.6, 2.4]$. Given this interval, the author's claim is that the exponent of Zipf's law depends on a balance between maximizing information transfer and saving the cost of signal use. This trade-off is in close relation to the one reported in [66] according to the expression

$$\Omega = \lambda I(S, R) - (1 - \lambda)H(S) \tag{3.3}$$

where Ω is the energy function that a communication system must minimize, $I(S, R)$ denotes the Shannon information transfer between the set of signals S and the set of stimuli R; and $H(S)$ is the entropy associated to signals, i.e. the cost of signal use present in any communication [61]. In this context, $\lambda \in [0, 1]$ is a parameter regulating the balance between the goals of communication (maximize transfer of information) and its cost. Of course, $\lambda = 1$ results in a completely effective commu-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Table 3.3: Language-oriented complex networks works. Ref. [170] partially contains also genuine cognitive-oriented research.

Graph	Source Network	Vertex	Edge	Orient.	N	$\langle k \rangle$	L	C	Reference
thesaurus graph	Moby's thesaurus	word	sense relation	undir.	30,244	59.9	3.16	0.53	[120]
collocation graph	BNC corpus	word	collocation	undir.	460,902	70.13	2.67	0.44	[63]
co-occurrence graph	BNC corpus	word	co-occurrence	undir.	478,773	74.2	2.63	0,69	[63]
thesaurus graph	Roget's thesaurus	word	sense relation	undir.	29,381	S. (3.3)	5.60	0.87	[170]
concept graph	WordNet	word	sense relation	undir.	122,005	3.3	10.56	0.03	[170]
association graph	free assoc. data	word	association	undir.	5,018	22.0	3.04	0.19	[170]
association graph	free assoc. data	word	association	dir.	5,018	12.7	4.27	0.19	[170]

nication, whereas $\lambda = 0$ leads to a costless (though null) communication.

Given this framework, energy Ω can be minimized for different values of λ. Results show a sudden jump from close to null information transfer (low values of λ) to a maximum information transfer at a critical value $\lambda^* \approx 0.5$. For values $\lambda > \lambda^*$, $I(S, R)$ does not increase. These results are in turn interpreted in the context of networks in [62], by showing that favoring information transfer without regard of the cost (low values of λ) corresponds to a dense, richly interconnected network (information availability); above a threshold, the situation is reversed and the network of signals and stimulus (language) is broken or disconnected (certain parts of language remain unreachable). The change from one to another scenario occurs, again, in the form of a phase transition at a certain critical value.

Up to now, we have been able to assess the existence of certain universal statistical trends (see Table 3.3 and references therein), and we have placed language networks in the framework of information and communication theory, which approaches them to its natural place, i.e. embedded in human cognition.

Thus, we now fully turn to the cognitive-oriented research. As Solé *et al.* [159] point out, some (possibly interacting) factors must be considered for a more comprehensive view on linguistic phenomena, for instance: a common brain architecture and vocalization system, or the need for optimization in communication and learnability. These new considerations have turned the attention of research towards a cognitive-oriented work, where the network is not the object of analysis anymore (or not exclusively, at least); rather it is the object *on top of which* language cognitive

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

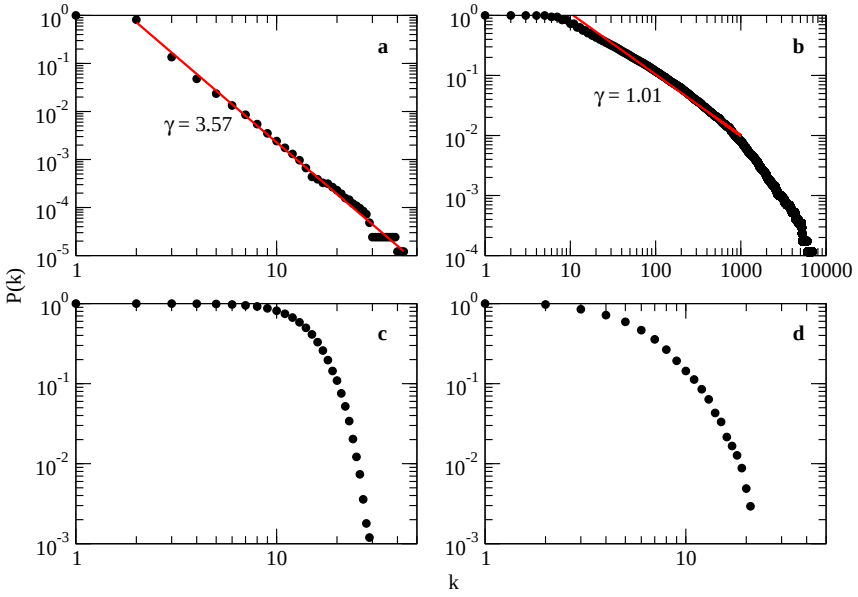

Figure 3.4: Plots of the cumulative degree distribution in four networks. All of them have been converted to unweighted and undirected. (a) WordNet, hypernymy relationships; (b) Co-occurrence networks for variable window size, from the ACE corpus; (c) English Free Association Norms (USF-FA); (d) Roget's thesaurus. Note that the plots are drawn in log-log scale. Only (a) and (b) display a power-law decay, whereas (c) and (d) do not follow a scale-free distribution. All of them, nonetheless, fit in the small-world definition.

mechanisms operate. Furthermore, more attention is put both on the type of data and its original meaning: while a coarse-grain general study on structural principles usually treats with undirected, unweighted networks, the cognitive approach tries to preserve as much as possible the original structures. By doing so, the natural heterogeneity and bias in cognitive phenomena are preserved. For instance, Figure 3.5 illustrates how misleading it can be to oversee the details in data. Summarizing, the study of cognitive processes demands a finer level of detail, where it matters whether a word facilitates another one, but not the other way around; or whether two words are semantically similar up to 0.9, whereas another pair reaches only 0.1.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

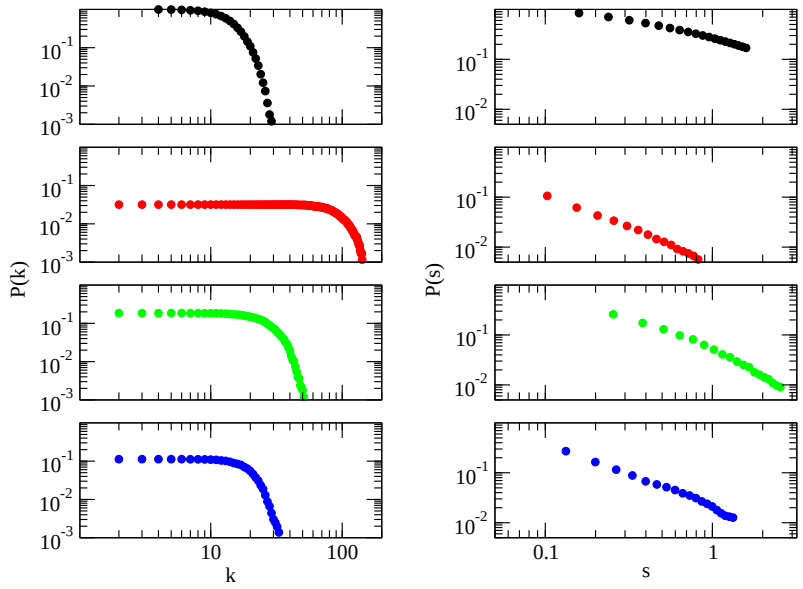

Figure 3.5: Directions and weights matter. Left: log-log plots of the cumulative degree distributions for psycholinguistic data in four languages (from top to bottom: USF, SFA-SV, SFA and GFA). Directions are symmetrized and weights are not taken into account. Right: log-log plots of the cumulative in-strength distribution for the same data without manipulation. Note that there exist striking differences between degree and strength distributions of psycholinguistic data. These differences are also evident in other descriptors, which suggests that comprehension about cognitive-linguistic processes demand attention to such details.

Both situations are treated as symmetric unweighted relationships in most complex network overviews of language.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

3.3 The cognitive pole I: Language and conceptual development

The work by Steyvers and Tenenbaum in 2005 [170] represents, up to date, the most comprehensive effort to join cognitive science with complex systems. As a confluence of these disciplines the authors vindicate the group of theories in psychology of memory which, under the label of *semantic networks*, were developed forty years ago [40, 41]. These classic semantic networks often represent defined relationships between entities, and the topological structure is typically defined by the designer. A classical example of this type of semantic network is Collins and Quillian's groundbreaking work in 1969 [40]. These authors suggested that concepts are represented as nodes in a tree-structured hierarchy, with connections determined by class-inclusion relations. Additional nodes for characteristic attributes or predicates are linked to the most general level of the hierarchy to which they apply, see Figure 3.6. Collins and Quillian proposed algorithms for efficiently searching these inheritance hierarchies to retrieve or verify facts such as *Robins have wings*, and they showed that reaction times of human subjects often seemed to match the qualitative predictions of their model. Word retrieval and recognition processes involve, in this proposal, tracing out the structure in parallel (simulated in the computer by a breadth-first search algorithm) along the links from the node of each concept specified by the input words. Such tracing process is known as "spreading activation". The spread of activation constantly expands, first to all the nodes linked to the first node, then to all the nodes linked to each of these nodes, and so on. At each node reached in this process, an activation tag is left that specifies the starting node and the immediate predecessor. When a tag from another starting node is encountered, an intersection between the two nodes has been found. By following the tags back to both starting nodes, the path that led to the intersection can be reconstructed. Interestingly, the relation between structure and performance is addressed in terms of the *cognitive economy principle*. Such principle, in its weak version, imposes certain constraints on the amount of information stored per node, thus affecting the structure (and its growth) in behalf of better future performance, see [40, 41] for further development.

A tree-structured hierarchy provides a particularly economical system for representing default knowledge about categories, but it places too strong constraints on the possible ways of organizing knowledge. Moreover, it has severe limitations as a general model of semantic structure. Inheritance hierarchies are clearly appropriate only for certain taxonomically organized concepts, such as classes of animals or other natural kinds.

The second classical proposal is that of Collins and Loftus [41] which, although

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

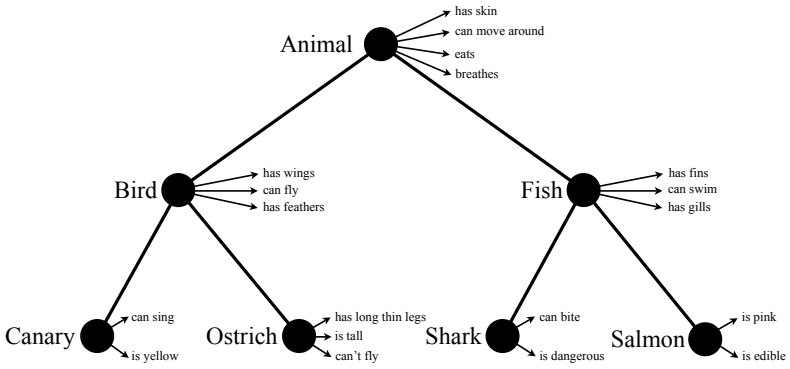

Figure 3.6: The Collins and Quillians tree data structure provides a particularly economical system for representing knowledge about categories. The cognitive economy principle prevents the structure from having redundant information, thus features which belong to one level do not appear in any other. Despite some positive experimental results with humans, the structure is far too rigid to accommodate actual semantic knowledge.

accepting many of Collins and Quillians premises, assumes a quite different data structure: a graph (notice that a graph is a general case of a tree; or, to put it the other way around, a tree is a particular case of a graph). Collins and Loftus model does not differentiate between concepts and their atributes. Therefore, nodes in the graph can either be nouns (such as "apple"), adjectives (such as "red"), or even compounded expressions (such as "fire engine"). Edges connecting them express a semantic relationship between them (not necessarily a category or similarity relationship), and it is assigned a number (a weight). Therefore, Collins and Loftus proposal yields an undirected, weighted graph which formally resembles very much the type of network that has been reviewed along this work.

Note that conceptually there is not much distance between Collins and Loftus graph proposal and complex networks. However, perhaps because of the limited prediction power of these proposals, perhaps because other points of view evidenced higher success at that time, the following decades did not witness a prolon-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

gation of these seminal works. As a consequence, there is relatively small agreement about general principles governing the large-scale structure of semantic memory, or how that structure interacts with processes of memory search or knowledge acquisition.

A complex network approach to language emerges naturally from this tradition, thus the work of Steyvers and Tenenbaum can be thought of as an update, both from the point of view of methodology and data availability. Although this work has a wide scope, part of it reports similar results as those reviewed in the previous Section, for instance a structural characterization of WordNet, Roget's Thesaurus and USF-FA. Our interest is focused now on the genuine cognitive approach to language learning or growth in an individual (*lexical development*).

The first part of the question can be stated: is it possible to find a variation on Barabási and Albert's preferential attachment which guarantees the emergence of a small-world, scale-free network? This question was already tackled by Dorogovstev and Mendes, as we have seen above. The novelty lies on the fact that the goal is to explain the statistics of semantic networks *as the products of a general family of psychologically plausible developmental processes*. In particular, (i) it is assumed that semantic structures grow primarily through a process of differentiation: the meaning of a new word or concept typically consists of some kind of variation on the meaning of an existing word or concept; (ii) it is assumed that the probability of differentiating a particular node at each time step is proportional to its current complexity (how many connections it has); and finally, (iii) nodes are allowed to vary in a "utility" variable, which modulates the probability that they will be the targets of new connections.

These constraints are translated to an algorithm which departs from a clique (fully connected network) of M initial nodes. Then, a node i is chosen to be differentiated at time t with probability $P_i(t)$ to be proportional to the complexity of the corresponding word or concept, as measured by its number of connections:

$$P_i(t) = \frac{k_i(t)}{\sum_{l=1}^{n(t)} k_l(t)} \tag{3.4}$$

where $k_i(t)$ is the degree (number of connections) of node i at time t. Secondly, given that node i has been selected for differentiation, we take the probability $P_{ij}(t)$ of connecting to a particular node j in the neighborhood of node i to be proportional to the utility of the corresponding word or concept:

$$P_{ij}(t) = \frac{u_j}{\sum_{l \in \Gamma_i} u_l} \tag{3.5}$$

where Γ_i stands for the neighborhood of node i. One possibility is to equate a word's

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

utility with its frequency; for a simpler model, one may also take all utilities to be equal, then connection probabilities are simply distributed uniformly over the neighborhood of node i:

$$P_{ij}(t) = \frac{1}{k_i(t)} \tag{3.6}$$

With these equations (Eqs.3.4-3.5, or Eqs.3.4-3.6) each new node is connected to M old nodes. Nodes are added to the network until the desired size N is reached. With these constructive algorithm a synthetic network is obtained, and its statistical features can be compared to the empirical counterparts. Steyvers and Tenenbaum report a significant agreement on degree distribution $P(k)$ match, as well as on some quantities, which are reproduced in Table 3.4.

	N	$\langle k \rangle$	C	L	D	γ
USF-FA	5018	23.5	0.1928	3.04	5	3.01
Synthetic USF	5018	22	0.174	3.00(.012)	5(.000)	2.95(.054)

Table 3.4: Results of model simulations (undirected version). γ is the exponent of the power-law that describes $P(k)$. Standard deviations of 50 simulations given between parentheses.

3.4 The cognitive pole II: Cognitive-linguistic processes on language networks

In the following subsections, we report two examples of the application of complex systems techniques to gain insight on genuine cognitive phenomena.

3.4.1 Google and the mind

The world wide web (WWW) presents at least two resemblances to associative models of language. First, it is organized as a directed network (nodes are web pages and the links between those nodes are hyperlinks, in the case of the WWW); second, its structure is dominated by the contents of its nodes. These factors add up to the fact that both human semantic memory and Internet face a shared computational problem, namely the necessity to retrieve stored pieces of information in an efficient way.

Given this, Griffiths and co-authors point out a very interesting parallelism between the PageRank algorithm [130] (see Figure 3.7) and human performance on certain cognitive processes [80].

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

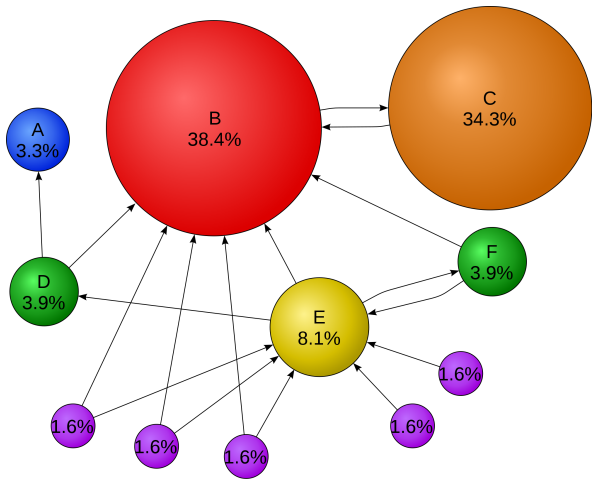

Figure 3.7: An illustration of the output of the PageRank algorithm. A link from an important web page is a better indicator of importance than a link from an unimportant web page. Under such a view, an important web page is one that receives many links from other important web pages. From Wikipedia Commons.

To explore the correspondence between PageRank and human memory, the authors used a task that closely parallels the formal structure of Internet search. In this task, people were shown a letter of the alphabet (the query) and asked them to say the first word beginning with that letter that came to mind. The aim was to mimic the problem solved by Internet search engines, which retrieve all pages containing the set of search terms, and thus to obtain a direct estimate of the prominence of different words in human memory. In memory research, such a task is used to measure *fluency* (the ease with which people retrieve different facts). With this experimental setup, accordance between the word's rank given by the algorithm and by empirical data is measured.

Results evidence that verbal fluency can be predicted, at least partially, attending the prominence (i.e. centrality) of words in memory. Furthermore, PageRank performs better predictions than those obtained attending word usage frequency, see Figure 3.8.

In the context of this review, note that the work of Griffiths and co-authors involves experimental design and direct, detailed comparison between the theoretical hypothesis and empirical results. From this point of view, the mixture of cognitive research and complex network methodology represents a real advance in the comprehension of knowledge organization in humans. Also, this novel orientation

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Beginning letter						
A	B	C	D	P	S	T
Human responses						
Apple (25)	Boy (11)	Cat (26)	Dog (19)	People (5)	Snake (11)	Tea (5)
Alphabet (7)	Bat (6)	Car (8)	Dad (16)	Penguin (3)	Stop (4)	Television (5)
Ant (6)	Banana (5)	Cool (3)	Door (5)	Pizza (3)	Saw (2)	Time (4)
Aardvark (3)	Balloon (4)	Card (2)	Down (4)	Play (3)	Sea (2)	Tree (4)
Ace (2)	Book (4)	Class (2)	Dark (3)	Pop (3)	Sex (2)	Table (3)
Ambulance (2)	Baby (3)	Coke (2)	Dumb (3)	Puppy (3)	Silly (2)	Tall (3)
Animal (2)	Ball (2)	Cookie (2)	Day (2)	Piano (2)	Sister (2)	Tank (3)
Absence (1)	Barn (2)	Crack (2)	Devil (2)	Pie (2)	Sit (2)	Telephone (3)
Acrobat (1)	Bear (2)	Cross (2)	Dinosaur (2)	Pig (2)	Slither (2)	Town (3)
Act (1)	Beef (2)	Cut (2)	Do (2)	Power (2)	South (2)	Train (3)
PageRank						
Animal (2)	Big (0)	Cold (0)	Dog (19)	Pretty (0)	Small (1)	Time (4)
Away (0)	Bad (1)	Car (8)	Dark (3)	People (5)	Sad (1)	Tall (3)
Air (0)	Boy (11)	Cat (26)	Drink (1)	Paper (0)	School (0)	Talk (1)
Alone (0)	Black (0)	Color (0)	Down (4)	Pain (0)	Sun (2)	Tree (4)
Apple (25)	Beautiful (0)	Clothes (0)	Death (1)	Puppy (3)	Smile (0)	Tired (0)
Arm (0)	Blue (2)	Child (1)	Door (5)	Person (1)	Stop (4)	Tiny (0)
Ache (0)	Book (4)	Cute (0)	Day (2)	Play (3)	Soft (1)	Thin (0)
Answer (1)	Body (0)	Clean (0)	Dirty (0)	Place (1)	Sex (2)	Top (1)
Apartment (0)	Bright (0)	Close (0)	Dirt (0)	Party (0)	Sky (0)	Together (0)
Alcohol (0)	Baby (3)	Cry (0)	Dead (0)	Pen (0)	Sleep (0)	Train (3)
Associate frequency						
Animal (2)	Bad (1)	Car (8)	Dog (19)	Paper (0)	School (0)	Time (4)
Air (0)	Book (4)	Clothes (0)	Death (1)	Pain (0)	Small (1)	Tree (4)
Army (0)	Black (0)	Cold (0)	Drink (1)	People (5)	Sex (2)	Talk (1)
Away (0)	Big (0)	Clean (0)	Dirty (0)	Person (1)	Sad (1)	Together (0)
Anger (0)	Baby (3)	Child (1)	Dark (3)	Play (3)	Soft (1)	Test (1)
Answer (1)	Ball (2)	Class (2)	Down (4)	Party (0)	Stop (4)	Television (5)
Art (0)	Body (0)	Church (0)	Dirt (0)	Pretty (0)	Smell (0)	Think (0)
Apple (25)	Bird (0)	Cut (2)	Dead (0)	Problem (0)	Strong (0)	Top (1)
Alcohol (0)	Break (0)	Color (0)	Dance (0)	Police (1)	Smart (0)	Teacher (0)
Arm (0)	Boring (0)	Cat (26)	Danger (1)	Place (1)	Sick (0)	Take (0)
Word frequency						
A (0)	Be (1)	Can (0)	Do (2)	People (5)	She (0)	There (0)
All (0)	Before (0)	Come (0)	Down (4)	Place (1)	Some (0)	Than (0)
After (1)	Back (0)	Course (0)	Day (2)	Part (0)	State (1)	Time (4)
Another (0)	Because (0)	City (0)	Development (0)	Public (1)	Still (0)	Two (1)
Against (0)	Between (0)	Case (0)	Done (1)	Put (2)	See (0)	Through (0)
Again (0)	Being (0)	Children (0)	Different (0)	Point (0)	Same (0)	Take (0)
American (0)	Better (0)	Church (0)	Door (5)	Program (0)	Since (0)	Three (0)
Around (0)	Business (0)	Country (0)	Death (1)	President (0)	Small (1)	Thought (0)
Always (0)	Become (0)	Certain (0)	Department (0)	Present (0)	Say (1)	Think (0)
Away (0)	Big (0)	Company (0)	Dark (3)	Possible (0)	School (0)	Thing (0)

Figure 3.8: Human Subjects Responses in the fluency task and rankings given by different predictors. This table provides a selective list, showing only 10 items for each letter. In the sections of the table corresponding to the three predictors, the order of the words in each column reflects the rankings given by the predictor indicated. Numbers in parentheses are frequencies in the human responses. After [80].

places research on language networks in the general framework of traffic and navigation on complex networks. The hypothesis suggests that search and retrieval

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

are affected by the way information flows, this issue has received much attention during the past years, see for instance [8, 48].

3.4.2 Clustering and switching dynamics

Previous Section deals with a dynamic cognitive search process where subjects' production is independent of meaning, the task depends on the *form* of words, rather than their content. An alternative scenario might be that where subjects are demanded to produce words according to a certain category (for instance, "name any animal you can think of"). This approach has been studied in [78], under the theoretical framework of Troyer's model for optimal fluency [177], in which search and retrieval cognitive processes exist on a double time-scale, a short one regarding local exploration (clustering), and a long one accounting for switch-transitions times.

The authors' proposal shares some aspects with the previous one. However, the issue here is not prominence or availability of words (centrality), but rather the fact that words are organized in communities or modules. Such modules are not only topological clusters, but also thematic groups or topics. From this point of view, the *switching and clustering* mechanism, understood as a double-level navigation process, can be used to predict human performance in such task as it is reported in [28]. The switcher-random-walker model (SRW) is then a cognitive inspired strategy that combines random-walking with switching for random exploration of networks. It is found that the number of steps needed to travel between a pair of nodes decreases when following this strategy, and thus the overall exploration abilities of a SRW within networks improves respect to mere random walkers.

Interestingly, a highly modular organization plus a two-level exploration scheme allows the system to organize information or to evolve without compromising exploration and retrieval efficiency. In this sense, semantic memory might be organizing information in a strongly modular or locally clustered way without compromising retrieval performance of concepts.

Community detection on empirical databases reveals the highly modular structure of word association. Analysis of USF-FA's mesoscale yields a modularity value $Q = 0.6162$, about 150 standard deviations above its randomized counterpart; similar results have been obtained with SFA ($Q = 0.7930$). See an example of detected modular structure for a subset of USF-FA data in Figure 3.9. The partition has been obtained for this review using a combination of algorithms (Extremal Optimization [53], Fast Algorithm [124] and Tabu Search [11]) available at [75]. These values seem a good starting point from which empirical work can be taken ahead.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

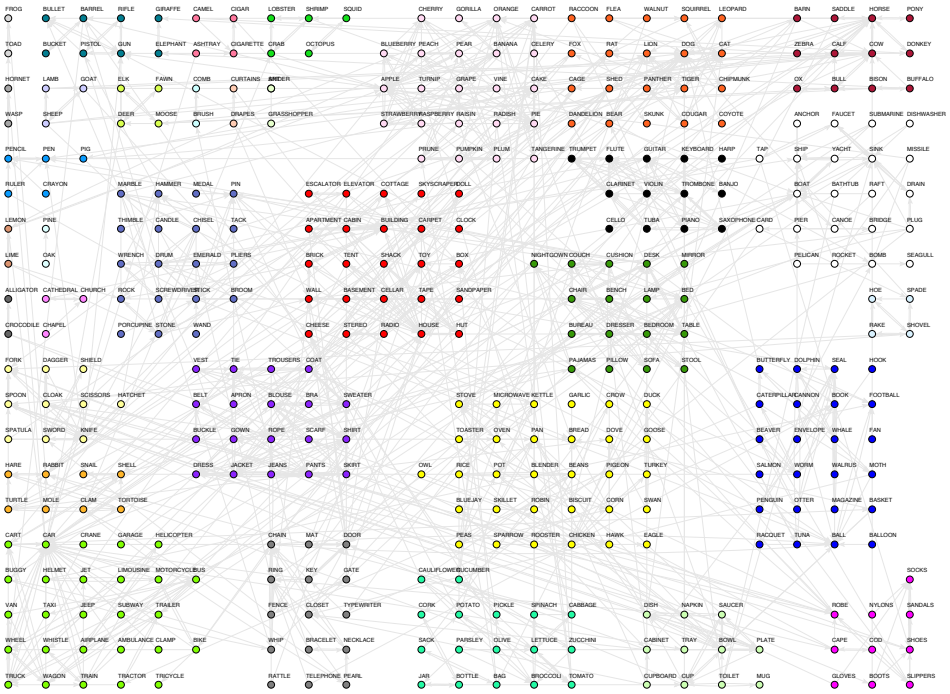

Figure 3.9: Community analysis for a subset of USF-FA with $N = 376$ nodes. The modularity value for this analysis is $Q = 0.8630$. The partition has been obtained for this review using a combination of algorithms (Extremal Optimization [53], Fast Algorithm [124] and Tabu Search [11]) available at [75].

3.5 Conclusions and perspectives

In this Chapter we have reviewed some important work from the last decade on language as a networked system. Work in this area has been strongly motivated by the uprise of a prolific branch of statistical mechanics, complex networks. Its foundations have been outlined in Chapter 2, focusing on a number of *macro* and *micro* statistical properties of networks that have received particular attention, and on some tools to scrutinize the *meso* level.

Section 3.1 elucidates the variety of sources and points of view from which language can be modeled as a network.

In Section 3.2 we have concentrated on the so-called language-oriented works

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

(lexical networks). Inspired by empirical studies of real-world networks ranging from the Internet to citation networks, researchers have approached language so as to propose models of networks that seek to explain either how networks come to have the observed structure, or what the expected effects of that structure will be. Such advances have brought to light two important facts: (i) that language resembles in many aspects other complex systems; and (ii) that different languages are also similar to each other regarding statistical descriptors. These results allow us to talk about the existence of certain universal trends that underlie linguistic structures. Within this Section, we have also seen some incipient efforts to link language topology and linguistic activity in humans.

In the last part of this review (Sections 3.3 and 3.4) we have discussed work on the behavior of processes that take place on networks. This implies a shift from an interest in structures *per se* towards an interest in the mechanisms that operate on them. It also implies a greater transdisciplinary effort, aiming at a convergence with knowledge from cognitive science. We have paid attention to some topics of a cognitive-oriented complex network research, namely lexical development (network growth with a cognitive accent) and mental navigation (dynamical processes on language networks).

The progress in this field is so rapid, that we have failed to discuss and even cite a number of relevant results.

We believe that these results are only the tip of the iceberg. In looking forward to future developments in this area it is clear that there is much to be done. From a methodological point of view, the techniques for analyzing networks are at present no more than a collection of miscellaneous and largely unrelated tools. A systematic program for characterizing network structure is still missing.

On the linguistic side we are just in the first attempts at answering a few questions; this means that almost everything is yet to begin. Some topics that might be important in the future are: are there common mechanisms in the emergence of SF language network structures in artificial communities of agents [15, 165, 166] and language acquisition in children? How can be mental navigation so efficient on a network which displays many different types of links between words? Is it possible to construct a typology of languages where the genealogical relations are reflected in network features? How do semantic categories evolve? Can semantic memory's malfunctions (blocking, persistence, bias, etc.) be explained in terms of topological changes? How are language networks modified through aging and brain damage? If we can gain some understanding for these questions, it will give us new insight into complex and previously poorly understood phenomena.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Chapter 4

The Random Inheritance Model

In the line of the second part of the previous Chapter, we now introduce another model that relates human performance in semantic processing tasks with processes operating on complex networks.

Semantic memory is the cognitive system where conceptual knowledge is stored. Empirical evidence from experiments with subjects and other lexical resources (*thesauri* [142], *corpus* [57], etc.) suggest that this system can be suitably represented as a semantic network, where each node corresponds to a word, and edges stand as pairwise associations. The network reconstructed from semantic information is in contrast with hierarchies created by individuals for computer storage and retrieval –which are trees– [97], the network has an intricate topology of cyclic relationships. Estimations that on average a healthy adult knows from 20,000 to 40,000 words [13] raise challenging questions about storage capacity, organization of the information and verbal performance. Regarding organization, some words are linked by virtue of their semantic similarity (intra-categorical relations, e.g. *car* and *automobile*). Other types of associations fall under the more general semantic relatedness, which includes the former and any kind of functional or frequent association [31], e.g. *car* and *road*. This implies that many types of association exist undistinguished in the network structure. In particular, categorical (similarity) relations are embedded in a much richer structure of superposed relationships.

In this Chapter we follow [26] to propose a computational model to extract semantic similarity information from the track of a dynamical process upon word association data. The main idea is that categorical relations emerge from navigation on the topology of semantic memory. Although we focus on cognitive phenomena and data, our efforts can be more generally interpreted in terms of the extraction of the backbone of a network, which entails that there exist "master relations" between elements (long-lasting similarity relations) and "incidental" (experience-dependent) ones that are entangled with the previous.

We use two empirical data sets to test the model: a general association semantic network as substrate of a dynamic process, and a feature similarity network for comparison purposes. Both have been characterized in the previous Chapter (USF-FA and FPN; simply FA and FP hereafter). After that, the model itself is detailed. We

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

name it the Random Inheritance Model (RIM) because it is based on uncorrelated *random walks* from node to node that propagate an inheritance mechanism among words. The results obtained yield significant success both at the macro- and the microscopic level when compared to actual data. Beyond the practical and applied aspects of RIM, we devise the formal background of the model in Section 4.4. Finally, we discuss that the key to such success is the modular structure of the substrate network, which retains significant meta-similitude relationships.

4.1 Topology of semantic networks

The algorithm that implements our model runs on general word association data, which are typically called Free Association. It is widely accepted that such data offer the most general and realistic insight of the structure of semantic memory, because they are not restricted to a particular kind of association. On the contrary, feature similarity data reports only the amount of features two words have in common, thus displaying strictly pairwise similarity information. See previous Chapter for a deeper comment on such data.

The differences in the nature of edges has drastic effects on the topology of these semantic networks, this can be analyzed in terms of statistical descriptors. In Table 4.1 we highlight some of such descriptors. $\langle s \rangle$ is the average strength per node; L is the average path length, defined as the average of the geodesic paths (minimal distance) between any pair of nodes; D is the diameter of the network; C is the average clustering coefficient. Strength distribution $P(s)$, the cumulative distribution function, which gives the probability that the strength of a node is greater than or equal to s. It is helpful to gain a global vision of a network's connectivity profile, in Figure 4.3 we see FA's and FP's distributions. A complete review of these descriptors can be found in [3, 21, 123].

It is readily understood from table 4.1 that the structures differ largely. The high connectivity in FP gives raise to a dense network, which in turn allows that any node is reachable in less than 2 steps on average. It also has the effect of a highly cohesive structure, i.e. clustering is prominent. In order to avoid size effects (the difference between FA and FP sizes), the same statistics are computed for the common subset of words, the differences between both topologies still hold. Strength distribution, which is plotted for FA's and FP's common subgraphs, also evidences deep structural disagreement, Figure 4.3.

We have analyzed quantities that describe macro and micro levels of networks. Also at the level of groups or communities (mesoscale) differences arise between FA and FP. This is expected, both because reviewed topological features differ largely,

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

	FA (all)	FP (all)	FA (subset)	FP (subset)
N	5018	541	376	376
$\langle s \rangle$	0.77	20.20	0.26	13.43
L	3.04	1.68	4.41	1.68
D	5	5	9	3
C	0.1862	0.6344	0.1926	0.6253

Table 4.1: Main statistical descriptors of the networks FA and FP, and their respective common words' subnetworks. N is the number of nodes; $\langle s \rangle$ is the average strength; L is the average shortest path length; D is the diameter of the network and C is clustering coefficient.

and the semantics of links is different from construction. Modularity optimization methods [49, 53, 124] yield partitions in which groups of words are gathered differently. FA shows a highly modular structure ($Q = 0.6162$), while FP reaches a modularity value $Q = 0.4288$. Lower modularity implies that clear boundaries are harder to define, this fits well with evidence of humans' fuzzy categorical system [143] and with computational models of verbal fluency [78]. Despite this, a close look to the words that conform communities, either in FA or FP, correctly reflect the distinct underlying associations, see Figure 4.1.

4.2 The Random Inheritance Model (RIM)

Up to now we have some clues about the type of topology our algorithm will be run on (FA), and what the output of the model should resemble (FP). From this knowledge we move on to develop the logic steps behind our proposal and describe the mathematical framework behind it. Recent works have pointed out the ability of a random navigation to explore the complexity of networks [45, 129, 185]. Here we propose a random navigation process and an inheritance mechanism to disentangle categorical relationships from a semantic network. Our intuition about the expected success of our approach relies on two facts: the modular structure of the FA network retains significant meta-similitude relationships, and random walks are the simplest dynamical processes capable of revealing the local neighborhoods of nodes when they persistently get trapped into modules. The inheritance mechanism is a simple reinforcement of similarities within these groups. We call this algorithm the Random Inheritance Model (RIM).

RIM proceeds in three steps, (i) initialization, (ii) navigation and inheritance, and (iii) output construction. Step (i) tags every word in the FA network with an initial features vector. The vectors are orthogonal in the canonical basis to avoid initial

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

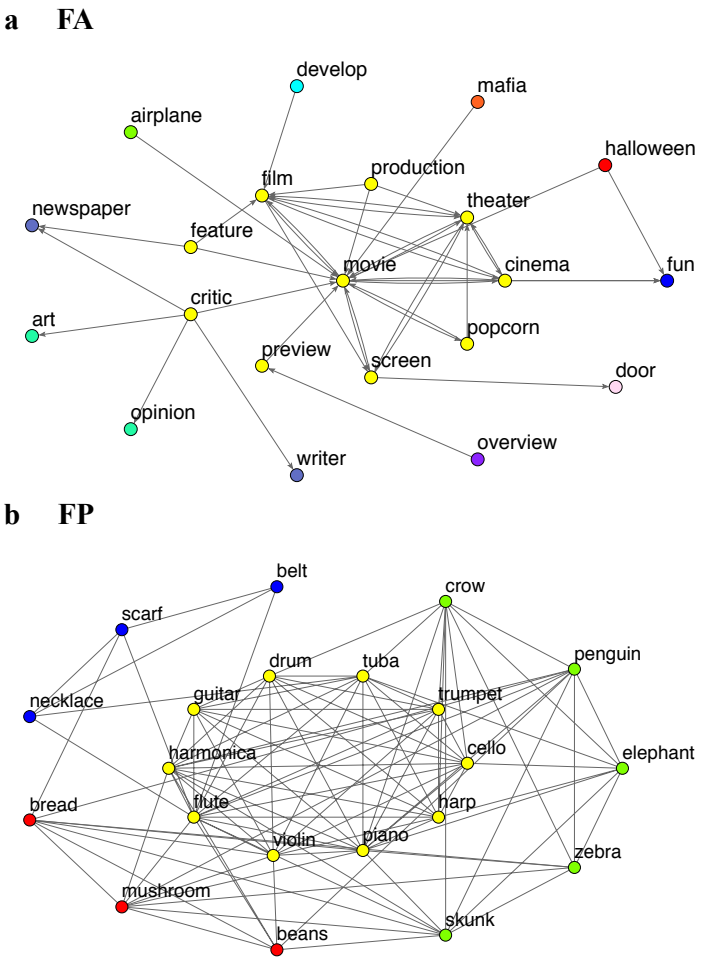

Figure 4.1: A sample of words that conform communities, from partitions obtained through modularity optimization in (a) FA and (b) FP. For the sake of simplicity edges leaving the depicted subgraph have been removed.

bias. That means that every word has associated a vector of N-dimensions, being N the size of the network, with a component at 1 and the rest at zero. The second step consists of launching random walks of length S from every word i in the network. The inheritance mechanism changes the vector of i, v_i depending on the navigation behavior. Let $s = \{s_1, s_2, ..., s_n\}$ the set of visited nodes. Then the new vector for

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

node i is computed as:

$$v_i = \sum_{s_i \in s} v_{s_i} \tag{4.1}$$

Note that (a) update of the feature vectors is synchronized, final values are computed after completion of the inheritance for every word; and (b) a random walk is a time-reversible finite Markov chain, which implies that node i can be itself in the set of visited nodes, see [105] for a survey on the topic. A new (synthetic) network FS is built in step (iii). Nodes in the new structure are those from the substrate network, weights between them are the result of projecting all pairs of updated vectors.
Steps (i)-(iii) are iterated (by simulating several runs) up to convergence of the average of the synthetic feature similarity networks generated at each run. The final average is the synthetic feature similarity network to be compared to FP.

This algorithm can be algebraically described in terms of Markov chains. Before we must define the transition probability of the FA network. The elements of FA (a_{ij}) correspond to frequency of first association reported in [122]. However, note that the 5018 words that appear on the data set are not all the words that appeared in the experiment, but only those that were at the same time cues in the experiment. Therefore data need to be normalized before having a transition probability matrix. We define the transition probability matrix P as:

$$P_{ij} = \frac{a_{ij}}{\sum_j a_{ij}} \tag{4.2}$$

As the original matrix, this one is also asymmetric. Once the matrix P is constructed, the random walkers of different lengths are simply represented by powers of P. In practice, this means that if we perform random walks of length S, after averaging over many realizations we will converge to the transition matrix P^S, every element $(P^S)_{ij}$ represents the probability of reaching j, from i, in S steps. The inheritance process corresponds, in this scenario, to a change of basis, from the orthogonal basis of the N-dimensional space, to the new basis in the space of transitions T:

$$T = \lim_{S\to\infty} \sum_{i=0}^{S} P^i = (I - P)^{-1} \tag{4.3}$$

However, in order to collect information about the dynamical behavior of the random walkers, one does not need to measure this limit. Thus, computations were done up to $S = 10$, and $S = 4$ already evidence good results. Results for RIM are expressed for $S = 4$ from now on. Finally, FS is the matrix that will represent in our model the feature similarity network (synthetic features network), where similarity is calculated as the cosine of the vectors in the new space, given by the scalar product of the matrix and its transpose, $FS = TT^{\dagger}$.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

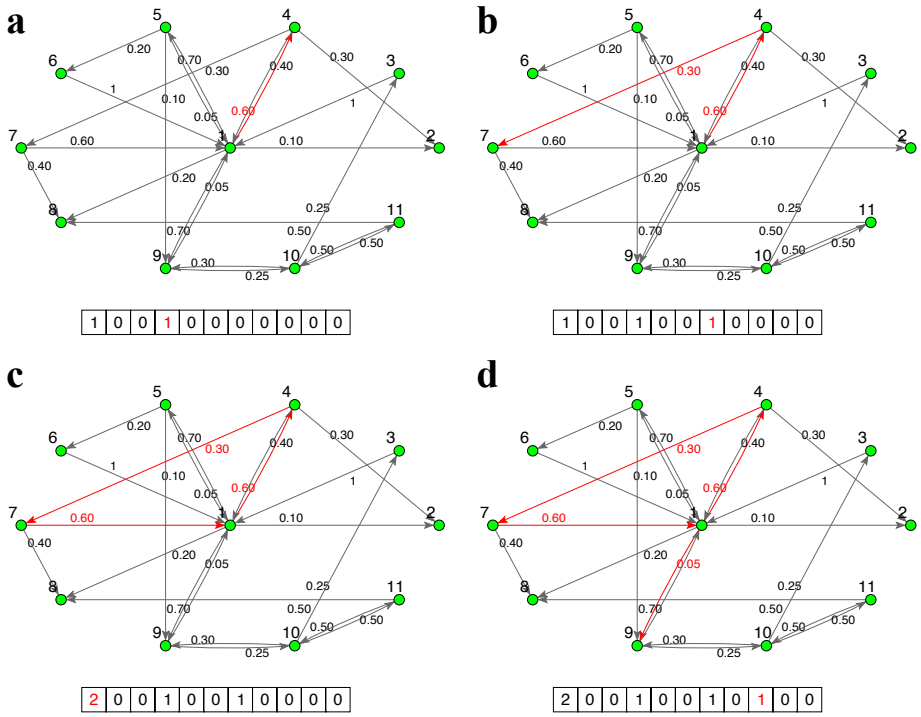

Figure 4.2: In RIM, the visits of a random walker starting at node i trigger the inheritance mechanism, which modifies the features vector of a node i. In the figure, a random walk of 4 steps changes the vector of node 1.

RIM fits naturally in the family of path-based similarity measures [20, 92, 93, 140, 149, 150, 102]. Jaccard index [92], cosine similarity [150] and the like have an inherent constraint, they can only account for short range similarities. This limitation is overcome in measures that take into consideration also long-range relationships [20, 93, 102]. However, a subtle distinctive feature of RIM is that similarity between nodes i and j is not a function of the number of paths from i to j, but depends on their navigational characteristics to the whole network, i.e. two nodes are similar if random walkers departing from them behave similarly. Cosine of vectors at the end of the navigation process accounts for random walkers' global performance. We think this particular feature is adequate in a cognitive-inspired dynamical mechanism, where navigation matters.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Table 4.2: Statistical parameters for Free Association norms FA (substrate of the dynamic process), Feature Production norms FP (empirical target), and the synthetic networks obtained using Latent Semantic Analysis LSA and Random Inheritance Model RIM.

Descriptor	FA	FP	LSA-N	FS
N	376	376	376	376
$\langle s \rangle$	0.26	13.43	39.60	15.62
L	4.41	1.68	0.02	1.77
D	9	3	2	3
C	0.1926	0.6253	0.9611	0.5848

4.3 Model performance

The algorithm sketched above yields a new synthetic network, FS. The capacity of RIM to extract similarity information must be tested against the empirical FP. We first check statistical macroscopical resemblance between FS and FP, by direct comparison of network descriptors and $P(s)$. We also point out results from Latent Semantic Analysis, LSA [50, 99]. LSA uses truncated Singular Value Decomposition to infer semantic similarity between pairs of words. We report results for LSA trained on the TASA corpus and truncation at $d = 300$, for the subset of common words in FA and FP. We will refer to this network as LSA-N. This LSA TASA-based representation is an appropriate benchmark because it largely succeeds at predicting human synonym test judgments [100].

In Figure 4.3 we plot the cumulative strength distribution $P(s)$ of the empirical networks FA, FP, and the synthetic ones LSA-N and FS. The statistical agreement between FP and FS is remarkable. Note that all distributions present an exponential decay instead of a power-law decay, being the cutoff of the distribution in FA more pronounced due to its original sparseness. Random homogeneous networks typically show this specific form of the distributions. Main descriptors of the four networks are presented in table 4.2. Again, the agreement between FP and FS is remarkable, the model reproduces with significant accuracy average strength, average path length, diameter, and clustering of the FP target network. The descriptors indicate that LSA-N is even denser than FP, close to complete connectivity.

Though informative and important, agreement on average or global descriptors does not determine to state the validity of RIM to extract actual categorical information from the original substrate. The reason for this is that nodes are tagged, conformity must be sought down to the local level. In practice, we intend to test

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

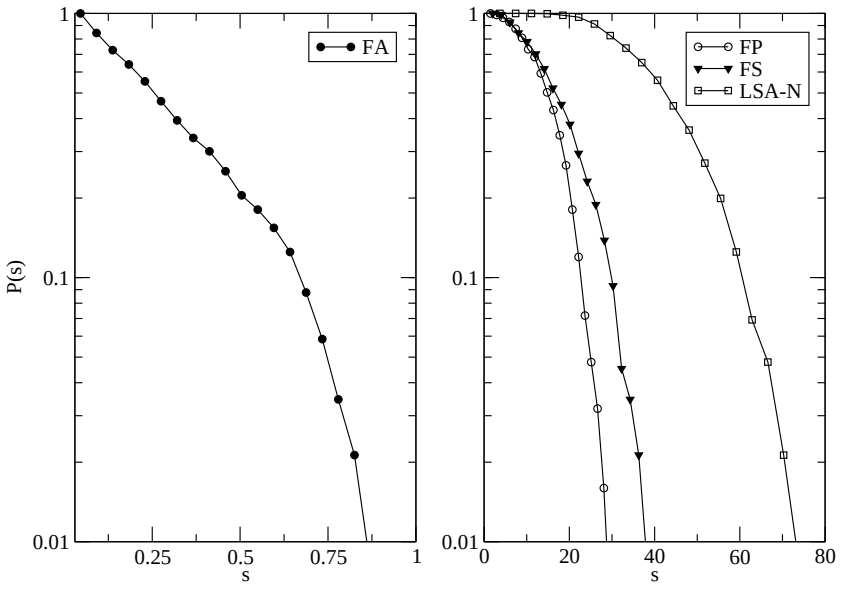

Figure 4.3: Log-linear plots of the cumulative strength distribution of the networks. Left: Free Association norms FA (substrate of the dynamic process). Right: Feature Production norms FP (empirical target) , and the synthetic networks obtained using Latent Semantic Analysis (LSA-N) and Random Inheritance Model (FS).

whether the specific neighborhood of a word in FP is replicated for the same word in FS (and LSA-N). We proceed as follows: given a specific word i, we start sorting its neighbors according to their linking weight. We apply this for each word in our data sets forming lists. The list of each word in FP is the empirical reference, and the lists we want to compare with, are those obtained for each word in the synthetic data sets, FS and LSA-N. We restrict our analysis up to the first 15 ordered neighbors, assuming that these are the most significant ones.

We now need a convenient measure to compare pairs of lists. To this end, we design a restrictive expression that assigns an error score between a list and its reference. Error depends on the number of mismatches between both lists, and also on the number of misplacements in them. A mismatch (M) corresponds to a word that

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

exist in the reference list and not in the synthetic list and vice versa. A misplacement (O) is an error in the order of appearance of both words in each list. The error score E is then defined as:

$$E = E_M + \frac{E_O}{l - E_M} \tag{4.4}$$

where E_M stands for the number of mismatches, E_O the number of displacements and l the length of the list. This quantity is inspired in Levenshtein edit distance [103] and its generalization, Damerau-Levenshtein distance [47]. In them, similarity between two strings depends on the amount of insertions/deletions and transpositions that one has to perform on a string in order to completely match another one. Notice that E is strongly increased when a mismatch appears, movements are less punished. Note also that $E = 0$ when lists match perfectly, we prescribe $E = l + 1$ for two completely different lists.

Besides a proper measure, we also define a suitable micro null case. To this end, we check whether categorical information is available just by listing a word's closest neighbors in the original FA. This implies the calculation of all-to-all shortest paths, weighting links as $d_{ij} = \frac{1}{p_{ij}}$, stronger relatedness is equivalent to shorter distance. Note that a direct neighbor of word i, i.e. a word with an edge from i, might lie at a longer distance than a second-level word. Success with this strategy would imply that RIM's retrieval capacity is merely due to topological closeness.

Success, i.e. $100(1 - E)$ with E as defined in Equation 4.4 and normalized, is plotted in Figure 4.4 for FS, LSA-N. Error in the null model is close to 100%, it has been left out in this plot. On average the success of FS is about 10% higher than that of LSA-N, the null model evidences that categorical information demands a stronger model to be disentangled.

4.4 Formal extension of RIM

In the previous sections we have presented the computational aspects of RIM, together with the basic intuitions that conform it as a plausible approach. Also we have focused on the practical aspects of it, testing it against empirical evidence and against another model (LSA).

Instead, we now turn to the mathematical foundations of RIM, which establish a strong connection between a cognitive semantic exploration strategy with a formalism in which semantic categories emerge as sets of recurrent states in a Markov chain.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

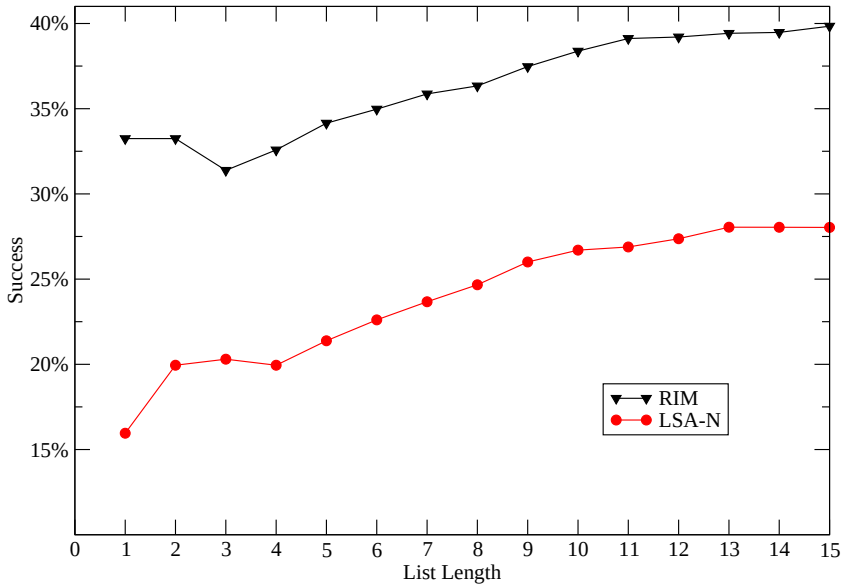

Figure 4.4: Log-linear plots of the cumulative strength distribution of the networks. Left: Free Association norms FA (substrate of the dynamic process). Right: Feature Production norms FP (empirical target) , and the synthetic networks obtained using Latent Semantic Analysis (LSA-N) and Random Inheritance Model (FS).

4.4.1 First approach

Let the process X_t, $t \geq 0$, define a finite Markov chain, where $X_t \in \{1, \ldots, N\}$ is a random variable describing the state of the process at time t. This chain is defined by the transition probability $p_{ij} = \Pr(X_{t+1} = j | X_t = i)$. Let $P = (p_{ij})$ be the *transition matrix* of the Markov chain. This is a *stochastic matrix* because $p_{ij} \geq 0$ for any i, j and $\sum_j p_{ij} = 1$. Let $p_{ij}^{(n)}$ denote the probability that, being in state i at time t, the process is in state j at time $t + n$; thus, $P^n = \left(p_{ij}^{(n)}\right)$. Obviously, P^n is also stochastic for any $n \geq 1$.

To any such Markov chain one can associate a finite directed graph $\mathcal{G}$, whose nodes are labelled $1, \ldots, N$, and such that there is a directed link from node i to node

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

TUBA

FP	LSA	RIM
trombone	**clarinet**	**trombone**
trumpet	violin	**saxophone**
drum	**flute**	**trumpet**
cello	guitar	**flute**
clarinet	**trombone**	**clarinet**
saxophone	fork	**cello**
flute	**trumpet**	violin
harp	cake	**harp**
banjo	**drum**	**banjo**
piano	**piano**	stereo
ERROR	4.83	2.5

ROOSTER

FP	LSA	RIM
chicken	cat	**chicken**
goose	gate	**turkey**
pigeon	donkey	crow
sparrow	barn	robin
penguin	turnip	**sparrow**
pelican	owl	**bluejay**
bluejay	pig	**pigeon**
dove	fence	**pelican**
hawk	lion	**goose**
turkey	strawberry	**hawk**
ERROR	11	2.87

Table 4.3: Some illustrative examples of LSA and RIM's predictive capacity, when compared to our FP (list size $l = 10$).

j if and only if $p_{ij} > 0$. Over this graph, the Markov process can be regarded as a random walk such that if the walker is at node i, it jumps to any of its neighbors j with probability p_{ij} ($p_{ij} = 0$ if j is not a neighbor of i). Thus a diffusion process over a weighted directed graph is equivalent to a finite Markov chain. We will analyze this process in terms of the properties of a finite Markov chain. Accordingly, the words node and state will be used interchangeably.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Assuming the graph $\mathcal{G}$ is not multipartite[1] the states of the Markov chain (hence the nodes of $\mathcal{G}$) can be classified into *transient* and *recurrent*. If we compute the limit

$$a_i \equiv \lim_{n\to\infty} p_{ii}^{(n)}, \tag{4.5}$$

transient states have $a_i = 0$ and recurrent states $a_i > 0$. In other words, the process returns to a transient state only a finite number of times, whereas a recurrent state is visited infinitely often.

In its turn, recurrent states can be partitioned into sets $R_1, \ldots, R_s$, such that $p_{ij}^{(n)} > 0$ for some $n > 0$ if $i, j \in R_\mu$, but $p_{ij}^{(n)} = 0$ for all $n > 0$ if $j \in R_\mu$ and $j \in R_\nu$, with $\mu \neq \nu$. In other words, any node can be reached from any other node within the same set, but nodes in different sets cannot be reached from each other.

According to this classification, starting from any node the process eventually hits a recurrent state. From that moment on, the process remains in the set this recurrent state belongs to. Within each of the sets R_μ the process is ergodic. Thus, if restricted to set R_μ, the process has a stationary probability given by the vector $\left(\pi'_{\mu i}\right)_{i\in R_\mu}$. It can be shown that

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi'_{\mu j}, \qquad i, j \in R_\mu, \tag{4.6}$$

which in turn implies that

$$\sum_{i\in R_\mu} \pi'_{\mu i} p_{ij} = \pi'_{\mu j}, \qquad j \in R_\mu, \tag{4.7}$$

i.e. the stationary probability vector is a left eigenvector with eigenvalue 1 of the submatrix $(p_{ij})_{i,j\in R_\mu}$. Actually, the limit 4.6 provides an iterative procedure to compute this eigenvector.

The partition of nodes that we have just obtained suggests a convenient relabeling of the nodes: the nodes of R_1 are taken as the first nodes, followed by those of R_2, then those of R_3 and so on until R_s, and finally we label the transient nodes (which form the set T). With this relabeling matrix P has the block structure

$$P = \begin{pmatrix} R & 0 \\ U & Q \end{pmatrix}, \tag{4.8}$$

where R is a block diagonal matrix with matrices $P_1, \ldots, P_s$ in the diagonal. Matrix $Q = (q_{ij})$, where $i, j \in T$, is strictly substochastic, i.e. $\sum_j q_{ij} < 1$, because eventually

[1] In other words, there is not a partition of nodes into sets $S_1, \ldots, S_r$ such that nodes from S_i lead only to nodes of S_{i+1}, with $S_{r+1} = S_1$. Such a partition would produce a periodic Markov chain. Anyway, multipartite graphs are the exception rather than the norm in complex networks.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

the process abandons the transient set T. Also, if $U = (u_{ij})$, where $i \in T$ and $j \notin T$, is such that $\sum_j u_{ij} > 0$, for the same reason. The latter matrix describes the probabilities of jumping from a transient state to a recurrent state.

Asymptotic probabilities of a finite, aperiodic Markov chain Our concern is the computation of the limit

$$P_\infty \equiv \lim_{n\to\infty} \frac{I + P + P^2 + \cdots + P^n}{n} = \lim_{n\to\infty} P^n. \tag{4.9}$$

For that we first notice that

$$P^n = \begin{pmatrix} R^n & 0 \\ U^n & Q^n \end{pmatrix}, \tag{4.10}$$

where the identity $PP^n = P^nP$ leads to the following recurrence for matrices U^n:

$$UR^n + QU^n = U^nR + Q^nU. \tag{4.11}$$

First of all,

$$\lim_{n\to\infty} Q^n = 0 \tag{4.12}$$

because of limit 4.5 (equivalently, because Q is substochastic). Secondly, R^n is block diagonal with blocks P_μ^n. Given the limit of these matrices described above, the limit of R^n can be conveniently described by introducing the vectors

$\boldsymbol{\pi}_\mu = (\pi_{\mu i})_{i\notin T}$ and $\mathbf{1}_\mu = (1_{\mu i})_{i\notin T}$ as

$$\pi_{\mu i} = \begin{cases} \pi'_{\mu i} & \text{if } i \in R_\mu, \\ 0 & \text{otherwise,} \end{cases} \qquad 1_{\mu i} = \begin{cases} 1 & \text{if } i \in R_\mu, \\ 0 & \text{otherwise.} \end{cases} \tag{4.13}$$

Regarding R,

$$\Pi = \lim_{n\to\infty} R^n = \sum_{\mu=1}^{s} \mathbf{1}_\mu^{\intercal} \boldsymbol{\pi}_\mu, \tag{4.14}$$

a block diagonal matrix whose μ-th block is a matrix all whose rows contain the stationary distribution for the recurrent set R_μ.

Finally, if we define

$$W = \lim_{n\to\infty} U_n \tag{4.15}$$

and take the limit in Equation 4.11 we obtain

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

$$U\Pi + QW = WR. \tag{4.16}$$

Also taking the limit in $P^{n+1} = P^n P$ we get

$$\Pi R = \Pi, \qquad WR = W \tag{4.17}$$

The first equation summarizes relationships in Equation 4.7, and the second equation allows us to simplify the left-hand side of Equation 4.16 and write

$$U\Pi + QW = W \qquad \Leftrightarrow \qquad W = (I - Q)^{-1}U\Pi. \tag{4.18}$$

The matrix $I - Q$ can be inverted because $I - Q$ is substochastic (which means that the spectral radius of Q is smaller than 1). Let us now define the vectors $\mathbf{v}_\mu(v_{i\mu})_{i\in T}$ as

$$v_{i\mu} = \sum_{j\in R_\mu} w_{ij}, \tag{4.19}$$

i.e. the probability of jumping from the transient state i to any state of R_μ. In matrix form this equation reads

$$\mathbf{v}_\mu^\mathsf{T} = U\mathbf{1}_\mu^\mathsf{T}. \tag{4.20}$$

Then

$$U\Pi = \sum_{\mu=1}^{s} U\mathbf{1}_\mu^\mathsf{T}\boldsymbol{\pi}_\mu = \sum_{\mu=1}^{s} \mathbf{v}_\mu^\mathsf{T}\boldsymbol{\pi}_\mu. \tag{4.21}$$

We define now $\mathbf{w}_\mu^\mathsf{T} = W\mathbf{1}_\mu^\mathsf{T}$; thus, according to Equation 4.18 and taking into account that $\boldsymbol{\pi}_\nu\mathbf{1}_\mu^\mathsf{T} = \delta_{\mu\nu}$,

$$\mathbf{w}_\mu^\mathsf{T} = (I - Q)^{-1}\mathbf{v}_\mu^\mathsf{T} = (I + Q + Q^2 + \cdots + Q^n + \cdots)\mathbf{v}_\mu^\mathsf{T}. \tag{4.22}$$

The meaning of $w_{i\mu}$, the i-th component of $\mathbf{w}_\mu$, is therefore the probability that starting from the transient state i the process gets absorbed into the recurrent set R_μ. Vectors $\mathbf{w}_\mu$ can be obtained as the solution to the linear systems

$$(I - Q)\mathbf{w}_\mu^\mathsf{T} = \mathbf{v}_\mu^\mathsf{T}, \qquad \mu = 1, \ldots, s, \tag{4.23}$$

and in virtue of Equation 4.22, matrix W becomes

$$W = \sum_{\mu=1}^{s} \mathbf{w}_\mu^\mathsf{T}\boldsymbol{\pi}_\mu. \tag{4.24}$$

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Thus, the asymptotic probability of a finite, aperiodic Markov chain is given by

$$P_\infty = \begin{pmatrix} \Pi & 0 \\ W & 0 \end{pmatrix}, \tag{4.25}$$

with Π given by Equation 4.14 and W given by Equation 4.24.

Semantic content of a Markov chain We can interpret the sets R_μ, for $\mu = 1, \ldots, s$ as *semantic categories* contained in the graph $\mathcal{G}$. A random walk starting at any node hits one of these categories with a certain probability. These probabilities, as well as the probability distributions within each category, yield a characterization of the nodes of $\mathcal{G}$, so that we can establish a relationship between these nodes by comparing their respective probability vectors. The rows of P_∞ provide these probability vectors, so that by computing the angle between the vectors corresponding to two different nodes we can establish a semantic relatedness between them. This is what RIM does at finite times (lengths of random walks). Given this formalism, the i, j entry of the matrix $P_\infty P_\infty^\mathsf{T}$ provides the dot product between the vectors associated to nodes i and j. Thus

$$\cos\theta_{ij} = \frac{(P_\infty P_\infty^\mathsf{T})_{ij}}{\sqrt{(P_\infty P_\infty^\mathsf{T})_{ii}}\sqrt{(P_\infty P_\infty^\mathsf{T})_{jj}}}. \tag{4.26}$$

In what follows it will be convenient to introduce a new inner product between vectors of $\mathbb{R}^s$. Given $\mathbf{x} = (x_1, \ldots, x_s)$ and $\mathbf{y} = (y_1, \ldots, y_s)$, we define

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{\mu=1}^{s} x_\mu y_\mu \|\boldsymbol{\pi}_\mu\|^2, \tag{4.27}$$

and its associated norm

$$\|\mathbf{x}\|_R \equiv \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}, \tag{4.28}$$

where $\|\boldsymbol{\pi}_\mu\|$ is the length of the stationary probability vector corresponding to the set R_μ. In order to provide an interpretation for Equation 4.27 let us try to figure out what $\|\boldsymbol{\pi}_\mu\|$ represents. It is very easy to prove that

$$\frac{1}{n_\mu} \leq \|\boldsymbol{\pi}_\mu\|^2 \leq 1, \tag{4.29}$$

$n_\mu \equiv |R_\mu|$ being the cardinal of set R_μ (the number of elements in that set). On the other hand, the lowest bound is reached by a vector $\boldsymbol{\pi}_\mu = (1/n_\mu, \ldots, 1/n_\mu)$ (i.e. when all nodes of R_μ are equiprobable) whereas the upper bound is reached when $\boldsymbol{\pi}_\mu$ has only one nonzero component (which is of course equal to one). Besides, if $\boldsymbol{\pi}_\mu$

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

has $k \leq n_\mu$ nonzero components, all of them equal to $1/k$, then $\|\boldsymbol{\pi}_\mu\|^2 = 1/k$. This suggests an interesting interpretation of $\|\boldsymbol{\pi}_\mu\|^2$ as the inverse of the "effective size" of R_μ. In the light of this interpretation, the inner product (Equation 4.27) can be regarded as a way of measuring the similarity between two vectors of $\mathbb{R}^s$ stressing the importance of the similarity of those components which coincide in narrower semantic categories over those which coincide in wider categories.

Let us now compute

$$P_\infty P_\infty^\mathsf{T} = \begin{pmatrix} \Pi & 0 \\ W & 0 \end{pmatrix} \begin{pmatrix} \Pi^\mathsf{T} & W^\mathsf{T} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \Pi\Pi^\mathsf{T} & \Pi W^\mathsf{T} \\ W\Pi^\mathsf{T} & WW^\mathsf{T} \end{pmatrix}. \tag{4.30}$$

From Equation 4.14 we get

$$\Pi\Pi^\mathsf{T} = \sum_{\mu=1}^{s}\sum_{\nu=1}^{s} \mathbf{1}_\mu^\mathsf{T}\boldsymbol{\pi}_\mu\boldsymbol{\pi}_\nu^\mathsf{T}\mathbf{1}_\nu = \sum_{\mu=1}^{s}\sum_{\nu=1}^{s} \|\boldsymbol{\pi}_\mu\|^2\delta_{\mu\nu}\mathbf{1}_\mu^\mathsf{T}\mathbf{1}_\nu = \sum_{\mu=1}^{s} \|\boldsymbol{\pi}_\mu\|^2\mathbf{1}_\mu^\mathsf{T}\mathbf{1}_\mu. \tag{4.31}$$

The entries of this matrix are

$$\left(\Pi\Pi^\mathsf{T}\right)_{ij} = \|\boldsymbol{\pi}_\mu\|^2\delta_{\mu\nu}, \qquad i \in R_\mu,\ j \in R_\nu, \tag{4.32}$$

from which

$$\cos\theta_{ij} = \delta_{\mu\nu}, \qquad i \in R_\mu,\ j \in R_\nu. \tag{4.33}$$

From Equations 4.14 and 4.24 we get

$$W\Pi^\mathsf{T} = \sum_{\mu=1}^{s}\sum_{\nu=1}^{s} \mathbf{w}_\mu^\mathsf{T}\boldsymbol{\pi}_\mu\boldsymbol{\pi}_\nu^\mathsf{T}\mathbf{1}_\nu = \sum_{\mu=1}^{s} \|\boldsymbol{\pi}_\mu\|^2\mathbf{w}_\mu^\mathsf{T}\mathbf{1}_\mu, \tag{4.34}$$

$$WW^\mathsf{T} = \sum_{\mu=1}^{s}\sum_{\nu=1}^{s} \mathbf{w}_\mu^\mathsf{T}\boldsymbol{\pi}_\mu\boldsymbol{\pi}_\nu^\mathsf{T}\mathbf{w}_\nu = \sum_{\mu=1}^{s} \|\boldsymbol{\pi}_\mu\|^2\mathbf{w}_\mu^\mathsf{T}\mathbf{w}_\mu. \tag{4.35}$$

If we define the vectors $\mathbf{w}_i = (w_{i\nu})_{\nu=1}^s$ and $\mathbf{e}_\mu = (\delta_{\mu\nu})_{\nu=1}^s$, the entries of the two matrices above will be

$$\left(W\Pi^\mathsf{T}\right)_{ij} = \langle \mathbf{w}_i, \mathbf{e}_\mu \rangle, \qquad i \in T,\ j \in R_\mu, \tag{4.36}$$

$$\left(WW^\mathsf{T}\right)_{ij} = \langle \mathbf{w}_i, \mathbf{w}_j \rangle, \qquad i, j \in T, \tag{4.37}$$

from which

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

$$\cos\theta_{ij} = \frac{\langle \mathbf{w}_i, \mathbf{e}_\mu \rangle}{\|\mathbf{w}_i\|_R \|\mathbf{e}_\mu\|_R}, \qquad i \in T,\ j \in R_\mu, \tag{4.38}$$

$$\cos\theta_{ij} = \frac{\langle \mathbf{w}_i, \mathbf{w}_j \rangle}{\|\mathbf{w}_i\|_R \|\mathbf{w}_j\|_R}, \qquad i, j \in T. \tag{4.39}$$

Remarks The meaning of formulae (4.33), (4.38) and (4.39) is very appealing. Every node in the graph $\mathcal{G}$ is linked to one or more of the sets $R_1, \ldots, R_s$ with different probabilities. The semantic relatedness of two nodes is determined by the overlap of this links, with a larger strength put on narrower semantic categories. Semantic categories provide a more precise meaning the narrower they are, because their size is associated to the dispersion of their meaning. In defining the size of categories, not only the number of nodes is important, but also the distribution of weights among these nodes. If a category has a large number of nodes, but their weights are very peaked around one or a few nodes, it means that the meaning is somehow "concentrated" in these few nodes and the effective size of the category is correspondingly small.

Computationally, all that this analysis amounts to determine is the classification of nodes into categories $R_1, \ldots, R_s, T$, and the computation of all vectors $\boldsymbol{\pi}_\mu$, by solving Equation 4.7, and $\mathbf{w}_\mu$, by solving the linear systems in Equations 4.23. There is a recursive algorithm for determining the semantic categories of the graph which is feasible for graphs even with a few million nodes, provided the associated matrix is sparse (i.e. low mean degree). Determining vectors $\boldsymbol{\pi}_\mu$ can be achieved by an iterative method, and solving the linear systems (4.23) can be done by applying an algorithm optimized for sparse matrices.

4.4.2 Markov chains with metastable states

The problem with networks of the type of FA is that most of its nodes can be easily classified as recurrent states. This implies that, in the infinite limit, FA is split into an ergodic subgraph of the order of N, and only a few nodes fall into the transient set. Note that, just as important as semantic categories (recurrent sets), transient states are those that allow effective exploration of the semantic network. For this reason the previous formalism must be re-formulated in terms of quasistationary distribution (or metastable configuration).

Let us assume that we have a Markov chain with metastable configurations among the recurrent states. In other words, suppose that there are two characteristic times: t_{qst} and t_{st}, such that for $t_{\text{qst}} \lesssim t \ll t_{\text{st}}$ the process behaves as if it had reached the stationary state, exhibiting a partition into sets $R_1, \ldots, R_s$, but when $t \lesssim t_{\text{st}}$ this

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

structure disappears and the true stationary state is reached. This stationary state may be formed by a single set or a less fine partition obtained by merging sets of the former partition. If t_{st} is sufficiently large compared with t_{qst}, only the metastable structure may be observable so it make sense to attempt a description in which the above partition plays a relevant role.

In order to achieve this goal we need to find a permutation of the recurrent states such that the block R in Equation 4.8 can be decomposed as $R = R_{\text{qst}} + E$, where R_{qst} is a block diagonal matrix with stochastic matrices $P_\mu^{\text{qst}} = \left(p_{ij}^{(\mu)}\right)_{i,j\in R_\mu}$ $(\mu = 1, \ldots, s)$ in the diagonal, and $E = (e_{ij})_{i,j\notin T}$ is a matrix such that $\sum_{j\notin T} e_{ij} = 0$. This preserves the stochastic character of the original matrix R. The definition of these matrices in terms of the original matrix $R = (p_{ij})_{ij\notin T}$ is

$$p_{ij}^{(\mu)} = \frac{p_{ij}}{\sum_{k\in R_\mu} p_{ik}}, \qquad e_{ij} = \begin{cases} p_{ij} & \text{if } i \in R_\mu \text{ and } j \in R_\nu,\ \mu \neq \nu, \\ -p_{ij}^{(\mu)} \left(1 - \sum_{k\in R_\mu} p_{ik}\right) & \text{if } i, j \in R_\mu. \end{cases} \tag{4.40}$$

The probabilities $p_{ij}^{(\mu)}$ have a well defined meaning: they represent the transition probability from state $i \in R_\mu$ to state $j \in R_\mu$ conditioned to remain within the set R_μ. On the other hand,

$$\sum_{j\notin T} p_{ij} = 1 \qquad \Rightarrow \qquad 1 - \sum_{k\in R_\mu} p_{ik} = \sum_{\nu\neq\mu} \sum_{k\in R_\nu} p_{ik}. \tag{4.41}$$

Metastability occurs only when the right-hand side of the above sentence is *small*. In that case, matrix E becomes a perturbation of the structure described in the previous sections.

The lowest order approximation amounts to simply neglect E. In that case metastable states become true stationary states because

$$\lim_{t\to\infty} R_{\text{qst}}^t = \Pi_{\text{qst}} = \sum_{\mu=1}^{s} \mathbf{1}_\mu^{\mathsf{T}} \boldsymbol{\pi}_\mu \tag{4.42}$$

is a well-defined limit. The distribution $\boldsymbol{\pi}_\mu$ is such that $\boldsymbol{\pi}_\mu R_{\text{qst}} = \boldsymbol{\pi}_\mu$. In the theory of Markov chains, this distribution is known as a *quasistationary* distribution and plays the role, for metastable states, of the stationary probability distribution for true stable states. All the structure discussed above is recovered with the replacement $\Pi \to \Pi_{\text{qst}}$.

Quasistationary time From the previous developments it follows that the challenge is to find the correct quasistationary time, t_{qst}. To do so, we rely on the fact

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

that a correct t_{qst} is one such that the diffusive process has already outlined the quasi-recurrent states (which we interpret as semantic categories), while it is not yet close to t_{st} (it has not yet coarsened semantic differentiation too far). In practical terms, the similarity standard deviation σ_{RIM}, which we measure as

$$\sigma_{RIM} = \frac{1}{N}\sum_i \sigma_i, \tag{4.43}$$

stands as an appropriate quantity to monitor the information diffusion that the Markov process has already performed. In this equation, σ_i is the standard deviation of the similarity values of word i.

Note that the existence of a metastable configuration from which semantic information may be inferred totally depends on the fact that the original substrate, the topology on which the dynamics is performed, already shows some modular structure: otherwise the diffusive process can not encounter a quasistationary state. This is in accordance with works that seek community structure relying on random walker-based dynamics [132].

It is also worth remarking that this framework validates the original approach of RIM, in which similarity was computed at some finite time of the Markov process.

4.5 Summary and Conclusions

We have presented a simple information retrieval algorithm (RIM). This algorithm yields a measure of similarity between all pairs of vertices in a network. RIM is naturally related to a class of path-based similarity measures, but its aim is not the discovery of *structural similarity*. Inspired by cognitive mechanisms of memory search and retrieval, RIM highlights similar words, i.e. words that belong to the same category. From this point of view, the focus is not to spot words with structural similarities, but words with similar meaning.

Along the Chapter we have proposed that RIM is related to open problems in natural language processing and cognitive science, the understanding of conceptual knowledge organization. For this reason empirical data is related to cognitive science, and output interpretation is in terms of semantic knowledge, the capacity of RIM to predict semantic similarity. RIM's results are compared to those of LSA, which has a long history of success in many machine learning linguistic-related tasks.

However we suspect that RIM has a more general interpretation. The meaning of a word (its defining features) is reduced to a dynamic process of probabilistic walks and inheritance, blind to semantic content. Then, semantic similarity is just

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

similarity of the behavior of random walkers: two vertices are highly similar when random walkers departing from them visit, on average, the same nodes. Indeed, topology is a key factor to understand RIM's success. In a highly modular scenario, such as FA, random walkers tend to get trapped [132, 144] reinforcing inheritance among vertices in the same community. Topological communities then enable metasimilitude relationships. While immediate neighborhood does not suffice to infer categorical relationships: mesoscale relationships matter.

The close connection of RIM to random walkers allows its reduction to an algebraic description in terms of Markov chains. All these facts yield an algebraic and topological interpretation of conceptual knowledge. The interest of the developed formalism, either in its stationary or quasistationary version, is that, given a directed graph we can create a new graph undirected graph which gathers the semantic information that emerges from the former. On the one hand, the semantic is characterized by the categories $R_1, \ldots, R_s$. Categories are defined by certain sets of nodes. They may or may not coincide with a priori concepts, but they contain the true information of the semantic of the original graph, in the sense that random walks eventually end up in one of these categories and do not leave them anymore. Nodes of the same category are all linked and links have the maximum weight 1. Nodes of different categories are not linked, though, so categories somehow represent pure concepts.

Transient states are linked to several semantic concepts with different weights. They play a crucial role because they are the connection between different semantic categories. Nodes belonging to two different categories are separated by a two step path through all transient states with nonzero component in those categories. They act as "step ideas" which allow transitions between independent concepts in a navigation through the semantic graph. One such typical navigation which starts within a pure category remains within the category for a while until suddenly it jumps to one "step idea" (transient node) and from there navigation enters a new pure category.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Chapter 5

Semantic Networks' Robustness in Alzheimer's Disease Patients

Following the clear cognitive orientation of the whole work we now turn to cognitive impairments. Network modeling of language is mostly devoted to normal, healthy performance. However, understanding how aging and disease affect proficiency in language production and comprehension is a great concern in the field. Graph theory has proved useful to analyze phenomena related to semantic storage and mechanisms operating on it. Though a simplification, such perspective can grasp the main trends of phenomena under study, opening complementary approaches to them: language growth and child language development [170, 88, 89, 87]; lexical availability [80]; semantic similarity and category formation [25, 27] or verbal fluency [78, 77] stand as good examples of such strategy. Rooted in the influential computer model put forward by Collins and Quillian [134, 40] and elaborated by Collins and Loftus [41], modern theory of complex networks represents a methodological update boosted by massive empirical data availability, while retaining the intuitive character of the framework. It has also provided a relatively simple but powerful quantitative framework to gain insight into brain pathologies. Several studies from cognitive neuroscience report a progressive loss of structural and functional connectivity in brain networks in patients compared with control subjects [164, 104, 174, 163]. Finally, there exists a rich literature regarding network robustness, breakdown and final disintegration, both in the cognitive field [1, 94, 5] and out of it [4, 135, 151, 22, 154].

In this Chapter, we first present theoretical results which analyze how deterioration is related to performance; then we apply the scheme from the first part to test its empirical relevance. In particular, we focus on one of the most intriguing effects of dementia of the Alzheimer type (DAT). The topological substrate used all along the Chapter is again the Free Association Norms dataset from [122] (FA), see Chapter 3. As it has been previously said, the network empirically obtained is directed (asymmetric) and weighted, weights represent the frequency of association in the sample. The normalized frequency yields a probabilistic interpretation: the corresponding matrix is a transition one, see Figure 5.1. We maintain the asymmetry property in

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

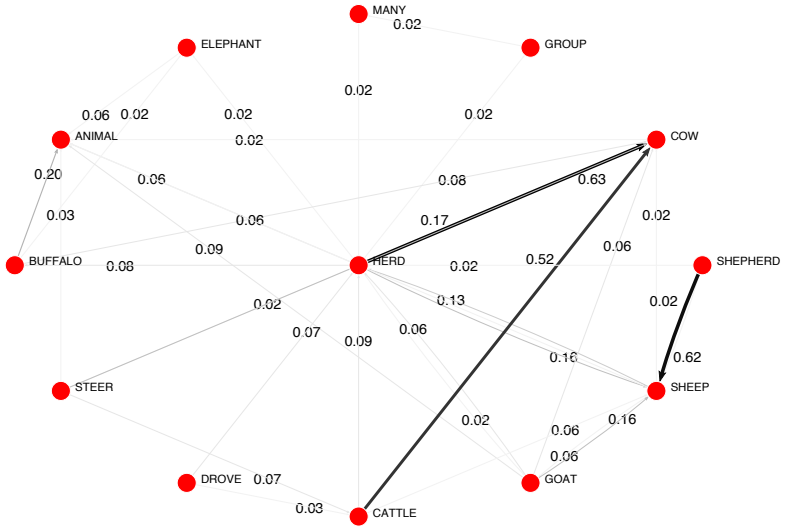

Figure 5.1: Partial representation of the Free Association network topology. Each node has a set of outgoing and incoming links. Because outgoing links correspond to produced frequencies, the resulting graph can be interpreted in terms of probabilities.

our approach to preserve the meaning of the empirical data.

5.1 Modeling dementia: Searching an adequate damage framework

5.1.1 Classical approach

In Chapter 4 we have proposed a mechanism that drives the emergence of category structure. Now we turn to the characteristics of both the original topology and RIM dynamics under error. Literature on error and attack tolerance in complex networks [4, 135, 151, 22, 154] typically model deterioration in two ways: error as the failure (removal) of randomly chosen nodes/edges, and attack as the removal of important nodes/edges ("importance" can be quantified by some descriptor, be it high connectivity, high betweenness, etc.). Using this approach, one typically monitors a suitable network characteristic that signals the moment in which physical disintegration of the structure takes place. Following most of the literature on percolation

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

processes, we follow the evolution of N_{giant}/N_{net}, that is, the proportion of nodes that remain connected to the giant component. As for RIM's dynamics, performance of a word i under error or attack is measured as the proportion of words that remain similar to i in the impoverished structure (we name it $match_i$), compared to the original results. This implies that, each time a node is removed, RIM is applied on the distorted structure and current similarity neighborhoods are compared to the original ones, for each node. The quantity "match" is a global average over the N elements of the network, and of course its value is 1 when no node has yet been removed.

In Figure 5.2 we use this approach by progressively removing nodes. This has been done in three ways: randomly choosing a node (failure), choosing it in terms of highest vertex betweenness (maximum betweenness attack), or eliminating the node with highest ω_{in} (maximum in-strength attack).

At least two conclusions can be drawn from Figure 5.2: in the first place, the relative size of the giant component N_{giant}/N_{net} decays in a similar way to those reported in the literature for scale-free networks [4], i.e. the structure is robust against failures but attacks hinder the integrity of the topology much before, approximately at $f = 0.75$. More interestingly, the RIM dynamics are not as resilient as the structure, and collapses long before the topology is actually disintegrated in the three cases (failure and both attack strategies). That is, before the critical point is reached, the dynamics' performance is much more deteriorated than the topology for any given fraction of removed nodes, f.

5.1.2 New approach

Though informative, we now wonder whether failure or attack correctly grasp the way in which pathologies or aging affects a cognitive structure such as semantic memory. Empirical evidence in the neuroscience literature report on a general decay of the neural structure supporting cognition [164, 104, 163]. Then, realistic modeling demands a different way to approach this problem. Here, we redefine error in the context of cognitive systems. In this framework, it is more useful to consider error in terms of aging or disease, where the whole topology simultaneously decays in some way. By doing so, we capture the degrading behavior of aging and/or disease, which differs from attack (there is no selective action) and from error (which affects only one node/edge at a time). For the sake of clarity, we refer to error in the cognitive framework as degradation.

Degradation assumes that links are increasingly damaged. At a given threshold τ, every link (i, j) in FA with a $\omega_{ij} \leq \tau$ is removed. The surviving links are normalized to preserve a probabilistic interpretation of the structure, see Figure 5.3. This

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

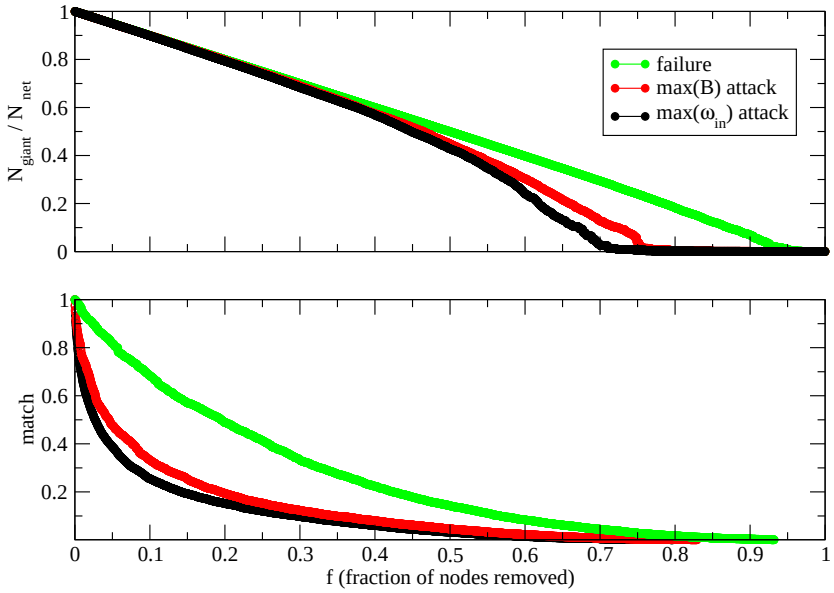

Figure 5.2: Upper: topological deterioration (relative size of the giant component) of FA as a function of f, the fraction of nodes removed from the network. In green, results for error (random failure of nodes); in red, results for attack to vertex betweenness: the nodes with highest B are removed the first. Finally, attack to higher in-strength appears in black. Lower: RIM's resilience for the same strategies.

process is performed with values $0 \leq \tau \leq 1$. As in the case of failure and attack, for each value τ, we monitor both topological and dynamical properties of the resulting network (the size of the giant component of the degraded structure is measured; and RIM is used to find a similarity matrix on the degraded structure, and the result is compared to the non-degraded RIM, i.e. RIM's result at $\tau = 0$ –"match" axis in figures–).

Degradation on the original structure Figure 5.4 shows the results for both topological deterioration (upper panel) and dynamical resilience (lower panel). Focusing on black circles (which correspond to degradation of the FA network), the behavior

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

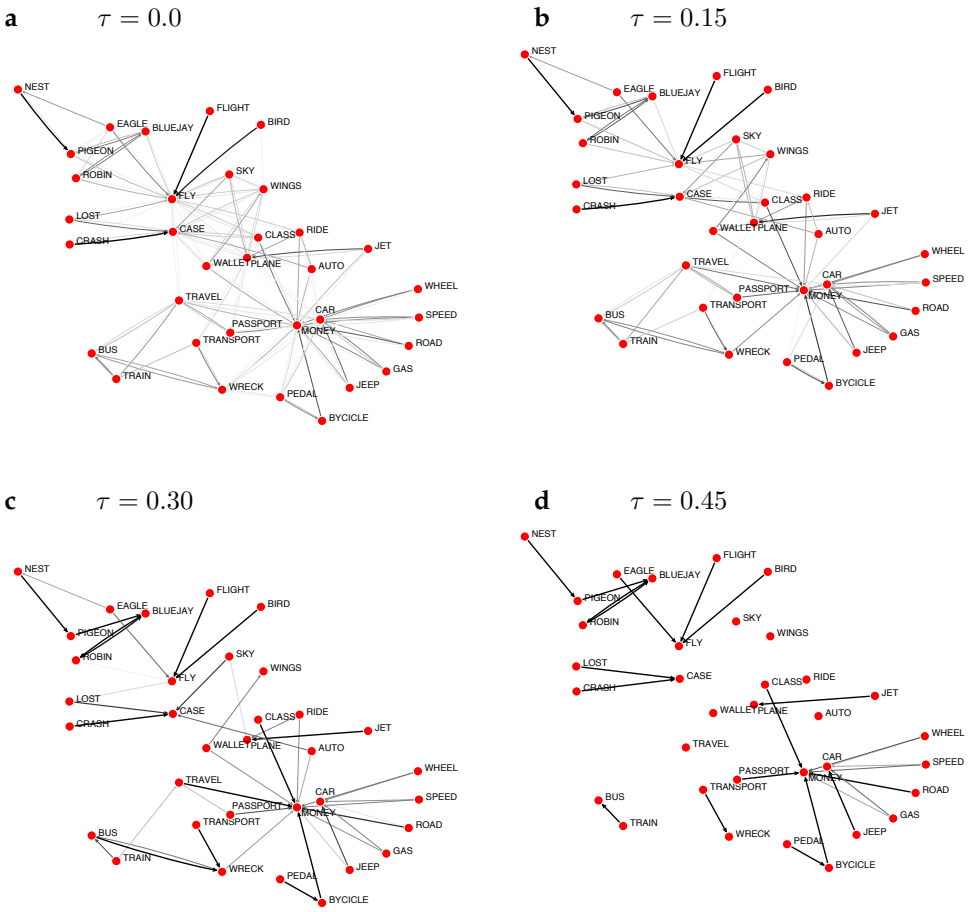

Figure 5.3: Topological global view of the degradation progress: Frames (a), (b), (c) and (d) correspond to different values of the parameter that controls the process τ, 0.0, 0.15, 0.30 and 0.45 respectively. Two main consequences of the process are observed: the topology is impoverished as weaker links (thin, clear lines) disappear, and at the same time some relationships are reinforced (thicker, darker lines) because of the probability normalization. This topology is merely illustrative, it does not correspond to empirical data.

of RIM's dynamics appears to be very sensitive to degradation even at very low values of τ. This suggests that lexical impairment can appear at early stages of semantic memory disease degradation. Interestingly, however, RIM's degradation is much slower than the topological one. At $\tau \approx 0.3$, FA structure is already disin-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

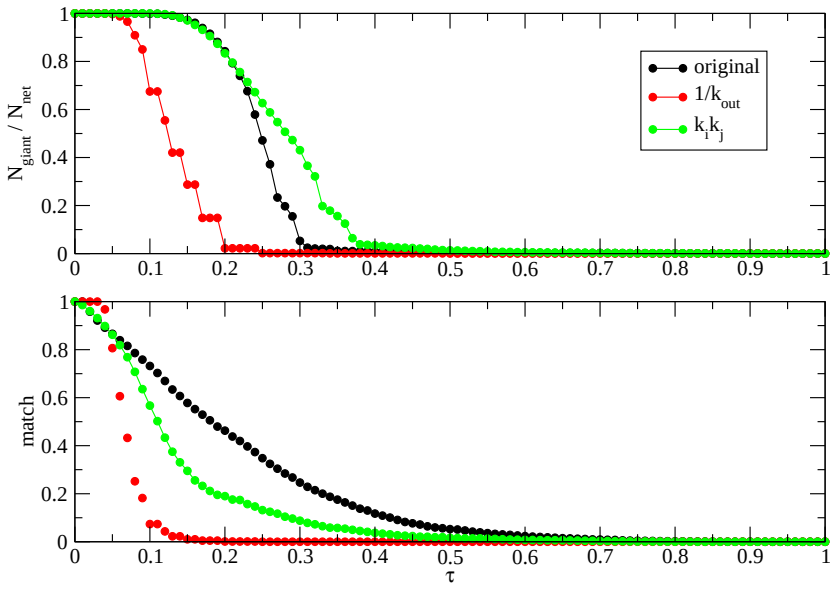

Figure 5.4: Upper: topological deterioration (relative size of the giant component) of FA as a function of τ. In black, results for the original FA structure. The same process of degradation has been applied to an unweighted version of FA understood here as a plausible null model, in red. In green, results for a second null model, which sets the weights of links in FA as $k_i^{out} k_j^{in}$. Lower: RIM's resilience for the same structures.

tegrated, whereas RIM can still recover as much as a 25% of its original content. RIM's results do not vanish up to $\tau \approx 0.6$. Such result indicates that fundamental cognitive capacities as word–word similarity inference and category formation are substantially resilient to structural impoverishment.

Degradation on the null model I Results in black circles (degradation of the FA network) from Figure 5.4 raise the question of which topological aspect of the dynamics' substrate provides for good performance in a disrupted topology. To answer this question we propose to build appropriate null models. The idea is that, by changing the topology that supports the dynamics, we can gain insight on which

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

properties of the original structure provide long-lasting performance. In particular, we devise two null models, both preserve the degree sequence and directions. In the first place, we consider the network FA in which weights are ignored. This implies that each node has the same in- and out-degree distribution, but outgoing links are weighted uniformly, that is $\omega_{ij} = 1/k_{out}$. The topological resilience of this modified FA network is represented as red circles in the top panel of Figure 5.4. The percolation point has moved to the left if compared to the original results, i.e. the network is structurally weaker. This result is not surprising, given that τ affects in the first place weaker links; thus, in an unweighted FA, this means that (i) nodes with higher degree k_{out} have the lowest weights, $1/k_{out}$; and (ii) since weights are distributed uniformly for each node, when $\tau = 1/k_i$ *all* out-going links are removed at once. In this sense, this null model is equivalent to a k_{max}-attack. From the dynamical point of view, it is apparent also that the structure can not support category formation and similarity inference for a long time: RIM's results collapse even before the topology reaches the topological percolation point.

Degradation on the null model II Although the unweighted null model is not an "aggressive" one (it preserves many of the features of the original network, such as degree distribution $P(k)$, average degree $\langle k \rangle$, average clustering coefficient C, etc.), we may devise one in which, furthermore, weight heterogeneity is still present. This can be done by assigning a weight to out-links proportional to the out-degree k_i^{out} of the source node i and the in-degree of the node j receiving that link, k_j^{in}. Then, $\omega_{ij} = k_i^{out} k_j^{in}$. The weights quantified in this fashion are normalized, to replicate the probabilistic interpretation of the original links in FA and in the previous null model. As it is apparent from the green circles in Figure 5.4 (upper panel), the $k_i^{out} k_j^{in}$-weight configuration yields a more resilient structure from a topological point of view, the percolation point is displaced to the right if compared with original FA. This is so because, contrary to the previous case, nodes with high degree are favored, in such a way that both their in- and out-weights are high. Then the threshold parameter τ does not affect hubs until a late degradation stage, the structure is not severely fragmented until $\tau \approx 0.4$. However, RIM decays faster than the original FA counterpart (Figure 5.4, lower panel). Although the value of "match" vanishes approximately at the same time as for the original substrate, dynamic deterioration for early τ values is more rapid.

Up to this point, we have studied how dynamics reacts when confronted to failure and attack, in the first place; and then to progressive degradation of the topological structure. To this end, we follow the line of percolation theory in complex network with some modifications. Results indicate that linguistic performance is severely affected by semantic memory degradation, on the other hand such perfor-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

mance is still significantly effective beyond topological disintegration.

Our study concludes, furthermore, that the specific distribution of weights in the lexical network plays a key role in the resilience both of the topology and the dynamics. Perturbing this weight distribution dramatically changes the capacity of the structure to hold performance of the dynamics on it. The particular value of these weights is just a consequence of contextual diversity (statistical biases of language use), which can soundly be identified as the origin of categorization and the maintainer of semantic integrity. On the other hand, this work raises some questions of interest. From a physical point of view, the new approach to structural damage demands an analytical treatment, in order to predict the topological response to weighted degradation. In this line a reconsideration of current knowledge on percolation theory is necessary.

From the standpoint of neuroscience and psycholinguistics, attention should focus on how physical (neurological) and cognitive degradation are related. Also, it has been reported that pathologies sometimes selectively affect linguistic performance [119, 34]. Then some kind of "selective degradation" should be implemented and studied, attending the modular structure that lexical networks display [26, 27]. Finally, other variables can be taken into account; for instance, at the moment a node is disconnected from the network, its "cognitive load" (the semantic meaning it bears) must be assumed by the remaining connected structure. In this way, degradation could be in interplay with node-breaking avalanches [117, 118], which could explain not only cognitive dysfunction (inexact or impoverished semantic capacities) but also system inefficiency (general performance slowing).

5.2 Degradation against real data: DAT and hyperpriming

The understanding of semantic memory impairments in DAT patients has been an important subject of investigation in the last years. There is converging evidence about the general symptoms, i.e. studies of spontaneous speech, verbal fluency, spelling and numerous other tasks all point to a progressive breakdown of knowledge about words and the objects they represent. However there is not such a consensus when it comes to explain unexpected or paradoxical performance. A situation of this kind is found when DAT patients are confronted with a semantic priming SP task and results are compared with controls. On one hand, unsurprising deficits appear when prime and target hold a object–attribute relationship (*zebra – stripes*), the result is less-than-normal priming effect (hypopriming). However, paradoxical hyperpriming occur for pairs of words which are category-coordinates (*lion – tiger*).

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

The hyperpriming effect is well documented in AD patients under the SP paradigm [121, 35, 106, 71, 70, 98, among others], and it also appears in different circumstances [46, which reports semantic dementia effects in repetition priming]. But the nature of this effect is still unclear. Some researchers hold that AD patients suffer a loss of information in the semantic store, whereas others point to the difficulty to access and process semantic information, see [85]. In this Chapter we use a graph theoretical approach to add some evidence in one way or another.

Semantic priming is associated to a response-to-stimulus time scale. In this work, the rationale is that topologically (structurally) close concepts will be associated in a quicker way than distant ones. Focusing on disease, our approach starts from the hypothesis that the semantic network undergoes degradation in brain pathologies as DAT. As a consequence, the weakest links progressively disappear. In this model, removal of a link relatively reinforces the remaining existing connections. Under this plausible assumption, hyperpriming naturally emerges for certain pairs of words. Furthermore, the severity of degradation can be controlled, then we can observe whether hyperpriming occurs in early or late stages of the disease. Our model is tested against empirical results from a longitudinal study of semantic priming in AD patients [70], the agreement is remarkable. Such agreement holds even for the finer-grained study, which splits category-coordinates pairs in "close" or "distant"; and object–attribute pairs, with the labels "shared" or "distinctive" [98].

5.2.1 Materials and Methods

Semantic Priming In 1971 the article by Meyer and Schvaneveldt [113] introduced the concept of *semantic priming*. Priming is an improvement in performance in a perceptual or cognitive task, relative to an appropriate baseline, produced by context or prior experience. Semantic priming refers to the improvement in speed or accuracy to respond to a stimulus, such as a word or a picture, when it is preceded by a semantically related stimulus (e.g., *cat-dog*) relative to when it is preceded by a semantically unrelated stimulus (e.g., *table-dog*). In other words, the semantic priming effect refers to the phenomenon whereby subjects are faster and more accurate in recognizing a target word when it follows a related word than when it follows an unrelated word.

The stimulus to which responses are made (e.g., *dog*) is the *target* and the preceding stimulus (e.g., *cat* or *table*) is the *prime*. The classical task for investigating semantic priming is the *lexical decision task*. The stimuli consist of correctly spelled words and meaningless strings of letters called "non-words" (e.g., *blit*). On each trial of the experiment, a prime and a target are displayed on a computer screen. Participants are instructed to read the prime silently and then to decide whether the target

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

is word or a non-word. The standard finding is that lexical decision responses are faster and more accurate when the target is semantically related to the prime than when the target is semantically unrelated to the prime.

Semantic priming has captured the attention of cognitive scientists for several years. At least one reason for it is that semantic priming occurs in many cognitive tasks, including lexical decision, naming and semantic categorization. The ubiquity of semantic priming suggests that it is caused by fundamental mechanisms of retrieval from memory. Following the "spreading activation paradigm", a speeded reaction time (RT) –priming effect– would be explained as the spread from the prime to semantically associated concepts, which consequently facilitates lexical access to related target words

In the context of pathology, chronometric measures have been brought to bear on the question of semantic memory loss and anomia. Relative to our concerns in the following sections, the most common findings among impaired subjects (patients with Alzheimer's Disease, schizophrenia, anomia, etc.) is that priming effects are progressively lost, i.e. a prime can not enhance the response to a target anymore. However, under certain conditions and for some early stages in the disease, the priming effect can actually appear stronger than that of healthy (control) subjects.

Topological degradation and semantic priming Placing the degradation framework in the context of DAT and semantic priming demands some comment. First, normalization after each τ thresholding is natural, in the sense that exploration of the semantic network continues, no matter the number of paths available. Besides, this redistribution of associative strength is central to observe the hyperpriming effect. Also, note that in the scope of our definition of degradation, the parameter τ can be understood as the control of the time progression: low values of τ represent early stages of the disease, higher values correspond to later stages.

For each value τ, we calculate the cosine similarity (or *closeness*) between every pair of words on the resulting, degraded structure. These values change across τ. Topologically, cosine similarity between two words i, j on FA expresses the proportion of common neighbors these nodes have. Note that a direct link between such nodes does not add similarity, it rather decreases it given that no self-loops exist in the network. As indicated above, we assume that structurally close concepts would display a stronger priming effect than distant ones, thus the measure of closeness is taken as a surrogate of priming effect: higher closeness corresponds to a higher speeding of response time to target words.

Remarkably, those nodes that are disconnected, i.e. those that no longer have incoming or outgoing links, will display no closeness to any node at all. However, those nodes which remain connected will acknowledge an increase in their close-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

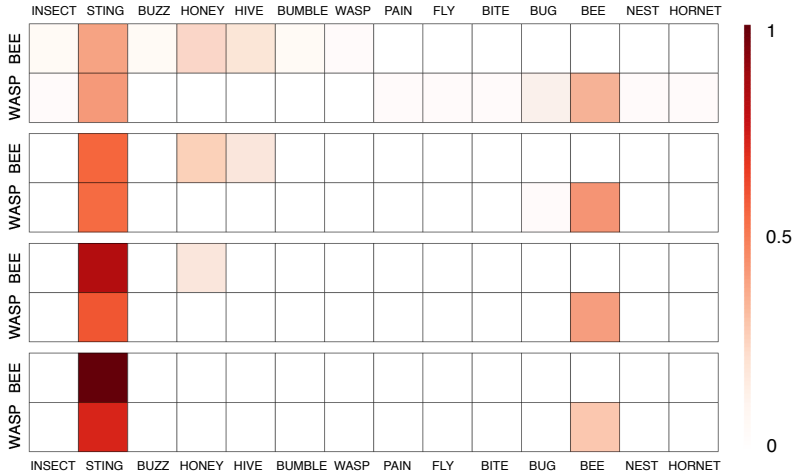

Figure 5.5: Topological local view of the degradation progress: The neighborhoods for the words BEE and WASP are represented as degradation takes place. The value of the threshold τ increases from top to bottom (0.0, 0.1, 0.2 and 0.3 respectively), the weaker connections are progressively removed. Due to normalization, the surviving links increase their value. The paradoxical growth in the words' mutual overlap (cosine) is naturally explained through the described topological change, see text for details.

ness. Such phenomenon is due to a reinforcement of the surviving paths as a consequence of the normalization of the remaining links, which yields higher strengths. Figures 5.3 and 5.5 illustrate the degradation process and subsequent redistribution of associative strength, from a large-scale and local point of view, respectively. Under this probabilistic interpretation, abnormal increased closeness (our priming proxy) between words arises naturally.

Enlarging Statistics: Feature Production Norms In order to decide which pairs of words' closeness should be monitored, we have followed the list in [98] (see Table 5.1). Results for these words are displayed in the corresponding section. However, we understand that the length of such lists suffices in those works where statistics are enlarged with many subjects (both patients and controls). On the contrary, only one topology is available upon which the simulated degradation and dynamics are performed. Thus, reinforcing the statistics is important, to this end we have made

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

use of public psycholinguistic data.

A significant amount of category-coordinate word pairs can be obtained using Feature Production Norms (FP, see Chapter 3), by filtering those pairs of words which display significant similarity. In this fashion we have obtained 246 pairs of category-coordinate words, which correspond to words in FP exhibiting a similarity greater than 0.65. On the other hand, we have filtered produced features in FP to retain the ones under the label "visual-form and surface" and such that the feature's rank is below 3, i.e. the feature has at most rank 3 within the concept according to production frequency. In this way a list with 246 concept-attribute pairs has been obtained.

Statistical errors The FA norms explained above are surely exposed to several sources of experimental error which are unobservable. It is mandatory then, to account for errors in the sample, which are understood as deviations of the sample from the (unobservable) true value. We analyze this using confidence intervals for the whole set of similarity values sampled. Standardizing the similarity values we determine that with a 95% confidence, the error is around 10% in the measure. Taking into account this value, we can generate synthetic samples of free association networks by using the original data and adding gaussian noise to the links with a variance of 10%. This methodology has proved to smooth the behavior of the functions represented without changing its qualitative structure.

5.2.2 Results

Topological disintegration In the context outlined in the previous section, the degradation parameter τ can be understood as a time scale: higher thresholding values correspond to later stages in semantic memory degradation. However, it is not possible to establish an exact mapping between τ and disease progression. Thus it is important to get a picture of how the topology behaves under degradation.

Figure 5.6 plots the topological evolution of FA as the topology is increasingly damaged (it represents the upper panel of Figure 5.4, isolated for the sake of clarity). The *giant component* of a network is the largest connected subnetwork. If a network is connected, the giant component's size N^* spans the size of the system N. At each value of τ, the size of the giant component of the degraded structure is measured. The proportion of nodes in the giant component $\frac{N^*}{N}$ is presented, in order to monitor the disintegration of the structure.

The analysis of the evolution of the giant component indicates that the topology remains mostly undamaged up to $\tau = 0.1$. In the other extreme, FA's giant component is only a 5% of N at the value of $\tau = 0.3$. For this reason, our degradation analy-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

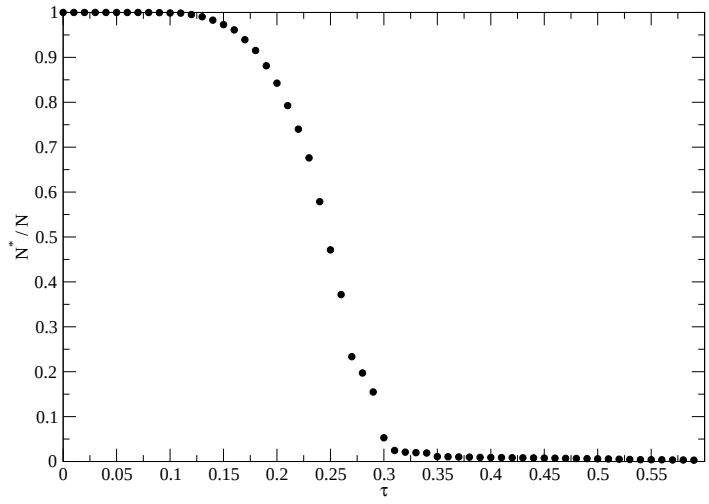

Figure 5.6: Topological degradation of the FA network as a function of τ. The fraction of nodes in the giant component $\frac{N^*}{N}$ is plotted against the degradation parameter (τ), in order to monitor the disintegration of the structure. The evolution of $\frac{N^*}{N}$ provides the range of interest for the study in the subsequent sections, which is $0.05 \leq \tau \leq 0.35$.

sis in the subsequent sections considers only values from $\tau = 0.05$ to $\tau = 0.35$, which correspond to mild-to-severe semantic memory damage. Our model degrades the topological structure by increments of 10^{-3} in τ providing a high resolution of the whole process.

Priming in data-mined word pairs Figure 5.8 depicts the average evolution of SP proxy as explained above, taking as a reference the lists obtained from FP. Such result evidences a qualitative resemblance to the results reported in [70] regarding the coordinate condition, i.e. hyperpriming emerges in the early stage of degradation, and vanishes afterwards (black small circles). The synthetic attribute-condition list evidences similarity to empirical results for low thresholding values, but clearly deviate at the end (small red squares). Nonetheless, it must be taken into account that the longitudinal study in the cited work includes only four experimental sessions. Thus the general tendency in both coordinate and attribute conditions can be recov-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

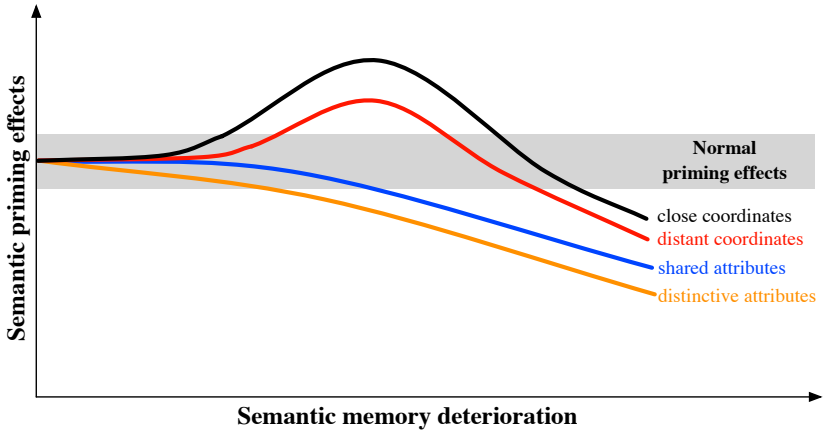

Figure 5.7: Schematic evolution of SP effects in relation to semantic memory degradation in DAT (mild to moderate dementia): adapted from [98]. Lines illustrate the hypothetical evolution of the range of SP effects according to the semantic memory deterioration in different semantic relationship conditions: close and distant category-coordinate, shared and distinctive attribute. Average normal priming effect is represented by the grey area. Under this limit is situated the extinction of priming while hyperpriming is illustrated by lines above it. An extinction of SP is observed in the DAT patients in the distinctive attribute condition, confirming the relative vulnerability of distinctive attributes (see Discussion). Simultaneously a hyperpriming effect was observed in the close category-coordinate condition. The hyperpriming effect was also found in the distant category-coordinate condition albeit to a lesser extent (see Results and Discussion).

ered, see dotted lines in Figure 5.8.

Priming in experimental word pairs For the case of the actual lists used in SP with DAT patients and control subjects we present Figure 5.9, which corresponds both to the category-coordinate and attribute word pairs and respective refinements within. The plot presents six traces: two of them are global averages for the coordinate (black) and attribute (red) conditions. The rest represent the refinements exposed in [98]. Again, the behavior of the curves in the coordinate conditions signal a transient period of hyperpriming with a subsequent decay, finally falling below the initial level of SP. The hyperpriming effect is significantly more remarkable in close coordinates. Despite the lack of significant statistics due to the lists' short length, the statistics have been enriched by means of the production of synthetic samples, see Section 5.2.1. We highlight once again the high level of coincidence with the

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

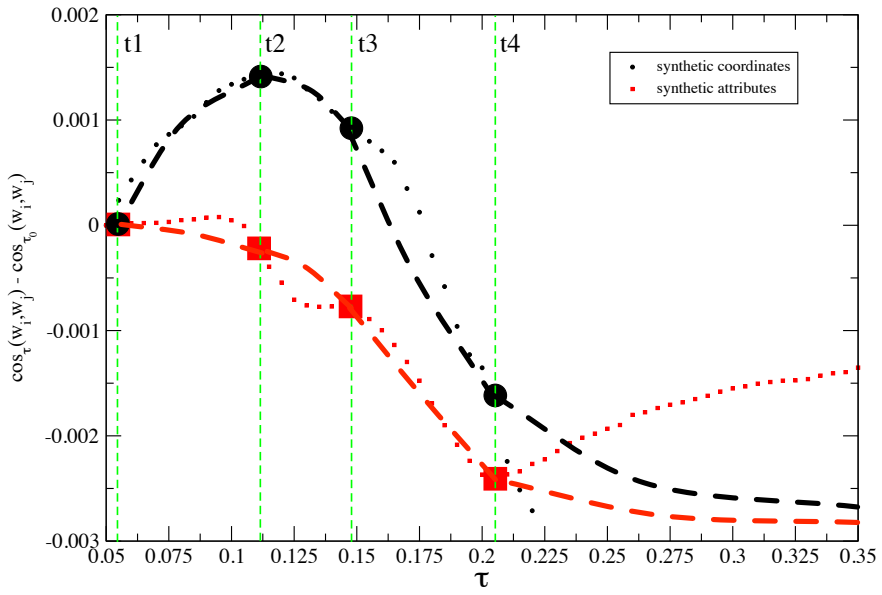

Figure 5.8: Average evolution of the semantic priming effect, both in the coordinate (small black circles) and the attribute (small red squares) condition. The plot begins at $\tau = 0.05$ to emulate DAT's early stages (an already damaged structure); and ends at $\tau = 0.35$, when FA topology is already disintegrated, see Figure 5.4. Discontinuous thick lines (black and red) show the tendency in SP if four single arbitrary points (t1, t2, t3 and t4, green vertical lines) are taken into account, recalling real experimentation in [70] which included only four experimental sessions.

data-mined word pairs and with empirical results for the attribute condition, i.e. a sustained performance in SP with a final decay below the initial state. All in all, results are in full coherence with the *in silico* ones previously offered, and with those empirically obtained.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

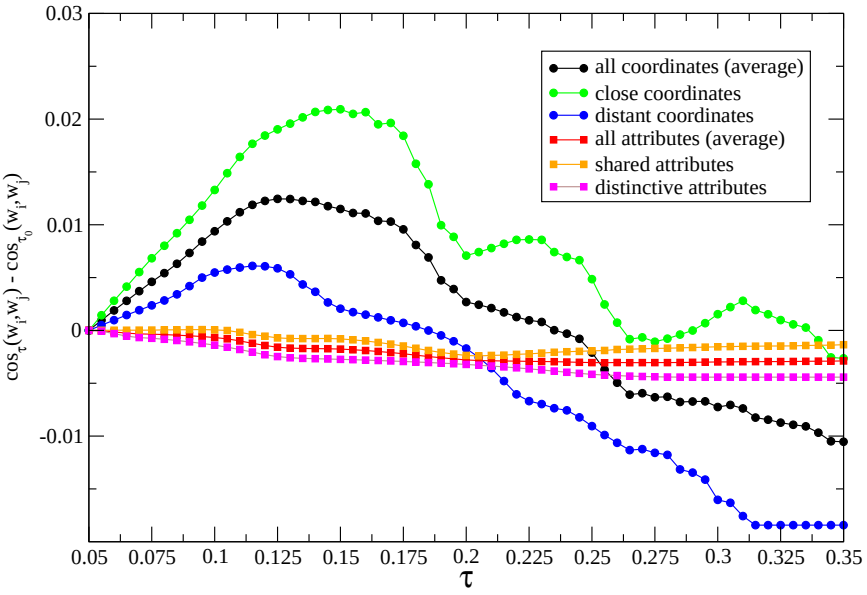

Figure 5.9: Average similarity results for the coordinate condition and attribute condition for word pairs in [98] (note that the lists have been adapted such that the words belong to our empirical dataset –FA–. When not possible, other items have been chosen following the criteria in the original work). The list comprehends 72 pairs (18 under the label "close category-coordinates", 18 in "distant category-coordinates", 18 in "shared attributes" and finally 18 in "distinctive attributes"). Again, the plot shows great resemblance to results reported in [98] (see also Figure 5.7): a transient hyperpriming effect appears before the whole performance decays.

5.2.3 Discussion

So long, we have introduced FA as a plausible representation of the structure of semantic memory. We take cosine similarity between the elements in FA as a good surrogate of SP effects. Once this information is available we study to what extent a progressive degradation of the topological structure affects performance, quantified in terms of such similarity.

To contrast synthetic results with the empirical ones, it is necessary to fix the

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

values of the degradation parameter, τ, in which the topology still yields signifiant values (that is, before what is known in physics as the percolation threshold, the point at which the whole network disintegrates in insignificant pieces). A general overview (Figure 5.4) allows to establish such limiting cases in which degradation is still likely to be cognitively relevant, i.e. $0.05 \leq \tau \leq 0.35$. At the lower limit of this range we may assume that disease has begun its degrading action, though mildly; while at the higher limit we can assure that the topology can not hold cognitive activity anymore.

The truly interesting observation emerges from the results depicted in Figures 5.8 and 5.9, which evidence striking similarity to the ones reported from experimental works regarding hyper- and hypopriming (Figure 5.7). This level of coincidence provides two strong conclusions: (i) the hypothesis by which semantic deficits in DAT stem from the difficulty to access and process semantic information is supported. The predictive success of our computational model is based on the idea that links are increasingly damaged, which is equivalent to hinder accessibility and proper navigation on the semantic network; and (ii) the so-called "category-coordinate condition" and "attribute condition" can be better understood in terms of structural connectivity.

We believe that assertion (i) is self-evident from the results along the paper. Hindered accessibility, modeled as a degrading process of the connections of a network, stands as a sufficient condition to observe hyperpriming. This is compatible with a scenario in which semantic search and retrieval strategies are qualitatively the same, but occur in a distorted topology. We do not claim, however, that other malfunctions (e.g. cognitive slowing) are not also present in the emergence of abnormal priming effects, this is beyond the scope of our computational model.

As for assertion (ii), in the remainder of this section we offer a detailed account of results for close and distant coordinates, as well as for shared and distinct attributes. These labels, at the light of such explanations, can be defined in terms of *topological patterns*, overcoming merely intuitive definitions, or one based on formal oversimplifications.

Following the sketch of SP effects as a function of semantic memory damage in [98], see Figure 5.7, our synthetic model offers an explanation for each case (close and distant coordinates, shared and distinctive attributes). Figure 5.10 illustrates a structural explanation for both close and distant coordinates. For these specific cases hyperpriming effect is reported, being the one corresponding to close coordinates more acute. Close coordinate pairs typically share many associates in FA, their semantic proximity favors the fact that they are linked to some common attributes and to other coordinates in the semantic network. Topologically speaking, beyond their having a direct, mutual connection, there usually exist many other paths con-

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

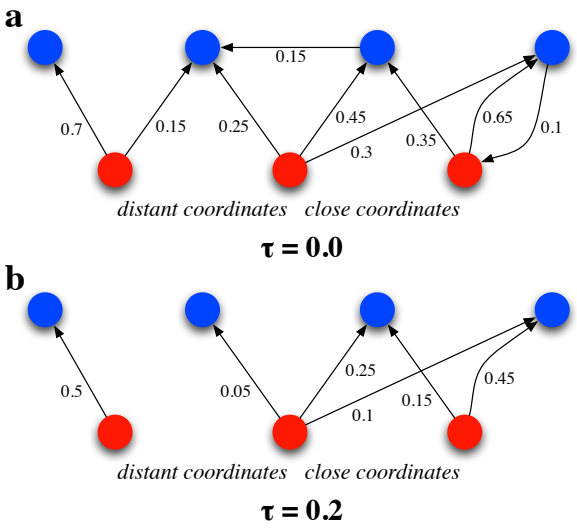

Figure 5.10: Simplified scheme of degradation in the coordinate condition. From a topological perspective, close coordinate words (middle and right red nodes) have a significant proportion of common neighborhood. Moreover, relations in such neighborhood are similarly strong, i.e. capable of surviving at least early degradation. On the other hand, distant coordinates share a lesser amount of neighbors. Note that all weights have been weakened after applying a threshold $\tau = 0.2$; those links weaker than this threshold have been removed.

necting a pair of this kind, which implies a great deal of common neighbors. This entails that the degradation process does not affect such relationships until deterioration is in a late stage. Since weights are normalized after the network has been thresholded, such weights tend to grow up to the moment when they disappear. This re-normalization implies a reinforcement in terms of the cosine similarity, thus the increase in SP is expectable up to mid-values in τ (Figure 5.5). After such values, common relationships are not held anymore, naturally accounting for the transient nature of the hyperpriming effect.

Distant coordinates share most of the characteristics with close coordinates, see Figure 5.10. However, the amount of shared neighbors is not as high as in close coordinates. Also, shared neighbors do not hold as strong relationships as in the coordinates case. This disparity in their connection patterns naturally yields a limited hyperpriming effect on such type of word pairs, not as marked as in close coordinates. Note that there is not a sharp distinction between close and distant coordinates, at the light of this topological characterization their differences are graded.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

The explanation for distinctive attributes is specially simple and elegant. By definition, these attributes are connected to very few concepts, because they are almost unique to those concepts. Being this so, degradation affects them enormously: as soon as the threshold achieves a certain value, the corresponding attribute's node becomes completely isolated, impeding the implicit spreading activation. This sharp dichotomy between existence/non-existence of a link is smoothed by statistics, in which distinctive attribute word pairs exhibit different weights, thus decay does not occur suddenly. Figure 5.11 illustrates this phenomenon. The tendency for such word pairs is a slow decay in early stages of DAT, and similarity dies off as early as $\tau = 0.25$.

Shared attributes show a similar pattern of decay compared to distinctive ones. Their decline, however, is not as fast. As it is apparent from Figure 5.11, the main difference between distinctive and shared attributes is, in topological terms, the creation of triangles (clusters). Unlike distinctive attributes, then, the rupture of a direct connection between a concept and an attribute does not imply the complete disappearance of a SP effect, due to shared connections. Thus degradation affects distinctive attributes first and then shared ones.

The detailed explanation of each case from a structural perspective is compatible with the hypothesis that, not being exactly the same, hyperpriming is close to repetition priming [70], in which prime and target are the same. The loss of distinctive connections at early stages of semantic memory deterioration turns two distinct concepts into very close, almost exact ones regarding their connectivity profile. In terms of the structure of the semantic network, the connection pattern (neighborhood) of a certain node which has lost many of its connections is almost exactly the same as that of another node which has also lost its distinguishing connections. Under this topological perspective, being close-to-synonym is not necessarily being mutually connected; rather two words become synonyms because they share the exact same neighborhood. Beyond the abnormal early stage of disease in which repetition priming appears, it is presumed that impoverishment of conceptual knowledge will prevail, as defining attributes become inaccessible, and the associational strength between related concepts weakened. Topologically, such associative strength actually vanishes. Again, we emphasize that word-word relationships can be redefined in terms of connectivity patterns.

All the same, such structural arguments find their coinciding neurological counterpart in [164, 104, 174, 163]. These works report the fragmentation of neural networks in DAT and other neuropathologies patients (which agree with the relative weight reinforcement due to link degradation and removal) and the disappearance of long range connections within such networks. The study of how changes at the physical layer are mirrored at the cognitive one is increasingly becoming a most

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

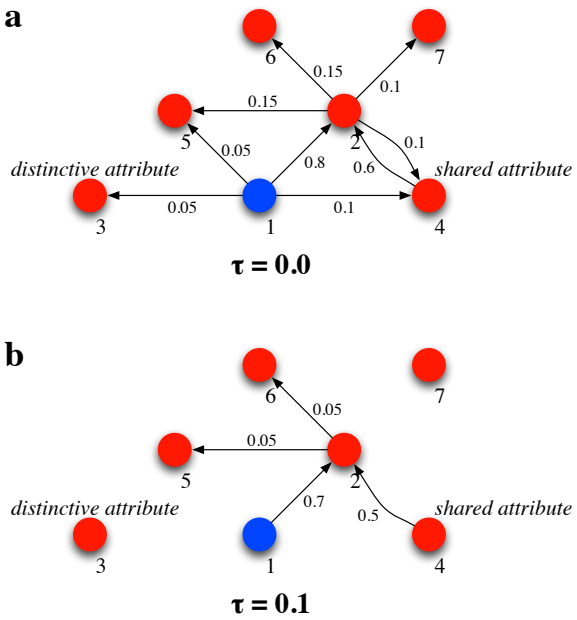

Figure 5.11: Topological characterization of the attribute condition, in an illustrative manner. Shared attributes exhibit a richer connectivity than distinctive attributes. This implies that complete disconnection typically appears at higher thresholding values. This fact explains the sustained evolution of closeness in the shared attribute condition, in contrast to the early collapse of the distinctive attribute condition.

relevant research issue.

The model can be used further. Following the structural approach developed so far, we can try to detect pairs of words which present an abnormal, transient growth in their closeness. That is, once the degradation process arises as a plausible explanation for hyperpriming, it is possible to use such model as a predictive tool. In this line, we have monitored across τ the whole set of possible pairs in our empirical data, to check for atypical closeness (our SP's proxy) evolutions. Seeking for early increases, we have detected 73,012 pairs (0.29% of the total possible) which display hyperpriming; 2,218,070 (8.8%) pairs match the hypopriming scheme. The rest (over a 90% of the possible pairs) have null closeness (note that the number of possible pairs amounts up to $N \times N$ pairs). These figures match our expectations: (i) most of the word pairs have no SP effect, either at the system's healthy state ($\tau = 0$) or in a distorted context; (ii) hypopriming is the most common evolution. As degradation

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

progresses, the main consequence must be a general impoverished performance; and (iii) hyperpriming is a rather restricted phenomenon, a collateral effect in a general weakened scenario.

Furthermore, a close look to the list of word pairs with predicted hyperpriming expands the range of words which might display this phenomenon beyond category-coordinates (see Table 5.2 for some examples). This would indicate that there exists a class of words which share a similar connectivity profile. Within this class one may find category-coordinates, but just as well other types of relations. This novel, more general insight reveals, as expected, that the complexity in the organization of semantic knowledge is far beyond the *ansatz* in [40].

On more general grounds, our explanation could also be used to design specific complementary therapies at the early stage of the Alzheimer's disease from a neuropsychological perspective. Those would rely on the reinforcement of the semantic network by inducing the formation of structural links between distant coordinates and distinctive attributes, for example. We devise that this can be done exposing patients to a sequential learning process linking these, in principle distant, concepts. The structural improvement of the semantic topology will increase the resilience to degradation.

Finally, topological degradation is here assumed to be uniform, i.e. the threshold parameter acts upon *any* connection. This does not match evidence, from which is known that some parts of semantic knowledge might be deteriorated, whereas others remain undamaged. Although selective damage has not been implemented in this work, it can be easily deduced that word pairs whose connections are not damaged do not yield unexpected phenomena, such as hyperpriming. This fact matches the ideas in [35] and [85], who report that hyperpriming is particularly noticeable for those items that explicit memory tasks had revealed to be degraded, but equivalent priming effects for patients and controls were found for items that were not degraded.

Given the highly modular structure of FA [26], some kind of selective degradation scheme could be designed such that different deterioration scenarios could be put to study. This line of future work demands higher attention to semantic memory degradation patterns under disease.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Distinctive attribute	Shared attribute	Distant c.c	Close c.c
train-wagon	pigeon-wings	spoon-fork	wardrobe-table
bicycle-pedal	gorilla-hair	*comb-brush*	whale-shark
spider-web	boot-heel	bee-wasp	*ant-beetle*
pine-needle	*bicycle-wheel*	*tiger-lion*	plate-bowl
shoe-lace	duck-feather	*jeans-trousers*	lobster-shrimp
stag-woods	cap-fabric	strawberry-raspberry	chair-bench
glasses-frame	cat-whiskers	*pear-apple*	*lettuce-celery*
plane-wings	*crow-wings*	pony-horse	fly-butterfly
dog-bark	pick-handle	sandals-shoe	truck-tractor
zebra-stripe	scissors-blade	*bison-buffalo*	daisy-tulip
hammer-head	tiger-claw	cup-bowl	*nail-screw*
bed-sheet	brush-bristle	garlic-onion	cow-sheep
crab-claw	lizard-tail	desk-table	knife-saw
sheep-wool	palm-trunk	snail-slug	dandelion-daisy
mill-blade	cherry-stem	wolf-dog	trousers-shirt
elephant-trunk	grasshopper-leg	necklace-bracelet	*peach-strawberry*
basket-wicker	pumpkin-seed	corn-wheat	vase-glass
snail-shell	*oak-leaf*	*tie-scarf*	*stag-pig*

Table 5.1: Word pairs of the four related conditions of the SP paradigm. Adapted from [98]: Monitored word pairs from which results in Figure 5.9 have been obtained. "c.c" stand for "category-coordinate". In italics, word pairs that have been adapted to accommodate them to available empirical data (FA).

frequency-television	pasta-sauce	Venus-Saturn	fragile-fix
frequency-speaker	pasta-noodles	Venus-Jupiter	fragile-bend
frequency-frequent	pasta-mafia	Venus-Pluto	fragile-smash
frequency-seldom	pasta-meatballs	Venus-Uranus	fragile-fracture

Table 5.2: Predicted hyperpriming word pairs (sample): Some of the pairs of words with predicted hyperpriming. Note that a notable proportion of pairs are category-coordinates (for instance, all the examples with planets involved), but the phenomenon appears also in a more general context (such as in *pasta-mafia* or the pairs in which *fragile* appears).

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Chapter 6

Conclusions. Current and future directions.

The last part of this long work, including these lines, has been written on a train somewhere between Tarragona and Zaragoza. Somewhere between the Universitat Rovira i Virgili and the Institute for Biocomputation and Physics of Complex Systems. I find this situation a suitable metaphor of my incipient scientific career: laid between different areas, my advisors, collaborators and myself are engaged on an interesting trip, exploring fascinating facts about our natural cognitive capabilities. Such exploration is merely at its initial stage, and yet a great effort has been needed to achieve it.

6.1 Where we come from

Before getting to this point, I had to learn a whole new methodology and see how it could serve our purposes to understand language. All that reading and learning has been condensed in Chapters 2 and 3. The former represents a quick overlook to some of the most outstanding review works in the complex network literature; the latter is an attempt to find a unifying trend in the myriad of articles that have focused on language and linguistic phenomena. On the track of this trend, I have tried to put the accent on the importance of a cognitive-oriented research, otherwise it is difficult to overcome merely descriptive achievements.

Following the argument, it does not suffice to consider complex network theory as a simple toolbox: find some data, build a network, characterize and describe it, unveil its possible modular structure. Iterate from the first step... Rather, we have tried (and hopefully succeeded) to use statistical physics and complex systems as an inspiring source of new approaches. Just as some models of social dynamics *are* themselves instantiations of variations on the Ising spin model, our constructions seek to find the physical foundations of certain cognitive processes. Chapter 4 exemplifies this attitude: although Free Association Norms present themselves very interesting features (regarding their being directed, weighted, modular and so on), we place on top of it a random walker-based dynamics, the Random Inheritance Model. We argument that this dynamics is plausible from a cognitive point of view

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

(in the sense that retrieving words form our semantic network resembles a random aggregated exploration in the long run), and we provide evidence that some interesting phenomena emerge from such process (semantic similarity, categorization). Inspired on physical grounds and with empirical relevance, the circle is complete with a proper formalization.

Chapter 5 is also an example of fruitful discussion between cognitive science, physics and computational modeling. In there we seek a plausible model for semantic memory deterioration in Alzheimer's Disease. This brain pathology has been thoroughly studied and therefore much data are available, both in psycholinguistics and in neuroscience. However, percolation theory on networks typically models damage as random error or targeted attack, which is not satisfactory for a process in which the system is attacked concurrently at many sites. Then, we must come up with new methods. In this way, a problem in cognitive science has led us to propose a new structure degrading scheme, which in turn raises interesting questions in physics (which, of course, we intend to solve in the near future).

All in all, these chapters acknowledge the slow, constant effort that has been committed to gain some understanding in linguistic phenomena.

6.2 Where I (would like to) go

By the time I write this text, the amount of open issues on my to-do stack has grown faster than my capacity to deal with them. These issues are not restricted to cognitive-related questions. My secret hope is that this situation will translate into some kind of job stability (I have only reached non-enduring metastable states regarding this part of my life). In any case, I will try to briefly review some of those "cognitive to-do's".

6.2.1 Semantic navigation

In Chapter 4 we have presented RIM as the mechanism responsible for the emergence of categories. In the end, the rationale behind this idea is that a thorough, uninterrupted exploration of our semantic network naturally produces a way to quantify semantic similarity between words. Focused on the prediction of these similarities, we have overlooked other applications of RIM.

For instance, we may think of the information obtained from RIM as a hidden metric space which enables navigation without a global knowledge of the structure. Such locally-directed, semantically-driven navigation should be:

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

1. Efficient: the cost in terms of "steps" (jumps from node to node) should be close to the pure topological shortest paths.

2. Internal characteristics: probably more interesting than the previous point, the content of "similarity paths" should be semantically smoother than those from a purely topological navigation. This means that, while purely topological navigation obeys the constraints from the weights of links (which, we should recall, represent *any* kind of relation), semantic relations from RIM are restricted to semantic similarity. Then the intermediate nodes necessary to reach a target word from a given source might be very different from one scheme to the other.

6.2.2 Wikipedia

I should be thankful to the experimental psychologists that collected the Free Association Norms (FA): as one may realize at this point, they have enabled the most fruitful results of our research. The question about data availability is always a sensitive one. Do we have larger, richer data sets than FA? Are they available in other languages? The answer is Wikipedia; or, to be precise, the internal linking structure of it. Note that, in a single Wikipedia article, one can find many types of links. For example, in the article devoted to Andrey Markov it is possible to find geographic references (where was he born and raised –Ryazan, Russia–, where he studied, and so on); professional references (a generic link to "Mathematics", but also more specific ones like "random process", "Brownian motion", etc.); personal references (noticeable people he interacted with). In this sense Wikipedia is very similar to Free Association Norms, where relations between concepts are very rich, and such richness is the key to success for RIM. Moreover, Wikipedia shares another characteristic with FA, which is its dynamic (growing and rewiring) nature.

English Wikipedia contains at this moment over 3,000,000 articles, each of them filled with links to other articles. It is also available in at least 20 other languages –those with the largest population speaking them. Presumably, this huge available data will soon offer us new insight about human knowledge organization and dynamic mechanisms operating at the cognitive level.

6.2.3 Cognitive robustness. Topology vs. Dynamics

Part of Chapter 5 was devoted to study the effects of different damaging strategies both on the topology and RIM dynamics. Some surprising effects were found, and these deserve further study.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

On one hand, we have a genuine physics problem in the newly introduced percolation-by-degradation. Note that this question can be addressed in two separate steps: (i) simple degradation, in which links below a threshold are deleted but the remaining ones are left unchanged. This demands monitoring the evolution of $P(k)$ in relation with the particular distribution of weights, $P(w)$; and (ii) degradation plus reinforcement, in which $P(w)$ changes both because of the removal of links and because values are normalized at each step, i.e. the distribution is now time-dependent $P(w, \tau)$. Both approaches can be interpreted in terms of systems where some quantity is preserved (presence of reinforcement) or not.

If this problem is solved, that is, if we attain an analytical understanding of the problem (such as describing the critical percolation point), the next question is to distinguish between topological (or static) and dynamic percolation. Throughout this work (and of course in the literature) it has been implicitly assumed that topology and dynamics are in close interplay. Literature on percolation has mainly focused on the structural collapse of systems, while probably the most interesting aspect to study is precisely dynamic resilience, i.e. how long can a system behave as it is expected before it functionally collapses. Of course, this turns the percolation problem into a more complex one, because typically we can study one topology, but the number of dynamics we can run on top of it are countless.

Within the particular cognitive orientation, we believe that the line of work of Chapter 5 is to be continued, that is, being able to characterize brain and cognitive pathologies, even in a simplified manner, so as to allow specialists to gain insight in their particular disciplines.

6.2.4 Linking the functional and the physical layer

In the Introduction we mentioned the fact that semantic networks are abstract objects, in the sense that they "live" in our minds. Leaving aside certain forms of naïve mentalism, it is generally accepted that all these mental structures find their foundations on a physical layer, the brain. We may introduce an illustrative metaphor to understand this peculiar relation mind–brain, by comparing these levels (cognitive–neural) to those of computer communications: the Internet represents the physical, wired, geo-referenced underlying structure of the logical level, the WWW. There is no way of inferring what the physical connectivity looks like if one merely pays attention to links pointing websites.

Is this really so? Are the physical and the logical layers absolutely uncorrelated? Evidence both from psycholinguistics and neuroscience (the emergent and polemic field of *cognitive neuroscience*) doesn't suggest so. Does this mean that there is a perfect mapping from the neural layer to the cognitive one? Again, the answer is

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

no.

Given this situation, one of the greatest challenges in the long run is to understand how these two layers interact and modify each other. I believe that physics-inspired methods will be of much use to tackle these problems. Complex neural topologies have already been spotted [161, 162, 186, 187], which suggest that complex network methods might be adequate in this area. Furthermore, data indicate that there are several pathways connecting the language-relevant brain areas [68], which suggest a networked structure. Finally, there exists strong evidence about specific localization of semantic categories. Brain imaging studies have shown that different spatial patterns of neural activation are associated with thinking about different semantic categories of pictures and words (for example, tools, buildings, or animals) [18, 86, 108, 116, 131]. These works suggest that the lexico-semantic system's organization we observe at an abstract level, i.e. semantically coherent modular structure (see Figure 3.9), may have a close correlate at the physical level. Chapter 5 also points at this direction.

Although more fine grained resolution of fiber tracts and crossings is necessary and unavailable nowadays, we can envisage some future research issues: what is the direction of the information flow in the fiber tracts connecting language areas? Is there a distinctive area where linguistic information is integrated? Is the modular structure detected in language networks mirrored at the neural level? These key questions open up a whole new and intriguing research scenario.

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

List of Publications

Cognitive modeling and neuroscience

Borge-Holthoefer, J. and Arenas, A. Navigating word association norms to extract semantic information. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 621 (2009)

Borge-Holthoefer, J. and Arenas, A. Categorizing words through semantic memory navigation. *European Physical Journal B*, 74(2), 265 (2010)

Borge-Holthoefer, J. and Arenas, A. Semantic Networks: Structure and Dynamics. *Entropy*, 12(5), 1264-1302 (2010)

Borge-Holthoefer, J., Moreno, Y. and Arenas, A. Topological vs. Dynamical Robustness in a Lexical Network, *International Journal of Bifurcation and Chaos* (in press, 2010)

Borge-Holthoefer, J. and Arenas, A. Navigation and Cognition in Semantic Networks. *International Journal of Complex Systems in Science* 1 (2011)

Eguíluz, V., Pérez, T., Borge-Holthoefer, J. and Arenas, A. Confronting structural and functional networks in complex systems with delay, *submitted* (2010)

Borge-Holthoefer, J. Moreno, Y. and Arenas, A. Degradation of a Free Association graph explains abnormal priming in Alzheimer's patients, *in preparation* (2010)

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Other publications

Gómez, S. , Arenas, A., Borge-Holthoefer, J., Meloni, S. and Moreno, Y. Discrete-time Markov chain approach to contact-based disease spreading in complex networks, *Europhysics Letters*, 89, 38009 (2010)

Arenas, A., Borge-Holthoefer, J., Gómez S. and Zamora-López, G. Optimal map of the modular structure of complex networks. *New Journal of Physics*, 12, 053009 (2010)

Gómez, S. , Arenas, A., Borge-Holthoefer, J., Meloni, S. and Moreno, Y. Probabilistic framework for epidemic spreading in complex networks. *International Journal of Complex Systems in Science* 1 (2011)

Erola, P., Borge-Holthoefer, J., Gómez, S. and Arenas, A. Reliability of optimal Linear Projection of Growing Scale-Free Networks, *International Journal of Bifurcation and Chaos* (in press, 2010)

S. Meloni, A. Arenas, S. Gómez, J. Borge-Holthoefer, and Y. Moreno, Modeling Epidemic Spreading in Complex Networks: Concurrency and Traffic. Chapter contribution to the Handbook of Optimization in Complex Networks (in press, 2010)

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

Bibliography

[1] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63–72, 2006. [cited at p. 85]

[2] L. Adamic. The small world web. *Res Adv Tech Digit Libr*, pages 852–852, 1999. [cited at p. 19]

[3] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Review of Modern Physics*, 74(1):47–97, 2002. [cited at p. 14 and 66]

[4] R. Albert, A. Jeong, and A.L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000. [cited at p. 85, 86 and 87]

[5] J. Alstott, M. Breakspear, P. Hagmann, L. Cammoun, and O. Sporns. Modeling the impact of lesions in the human brain. *PLOS Computational Biology*, 5(6), 2009. [cited at p. 85]

[6] L. Antiqueira, M.G.V. Nunes, O.N. Oliveira, and L.F. Costa. Strong correlations between text quality and complex networks features. *Physica A*, 373:811–820, 2006. [cited at p. 45]

[7] A. Arenas, J. Borge-Holthoefer, S. Gomez, and G. Zamora-López. Optimal map of the modular structure of complex networks. *New Journal of Physics*, 12:053009, 2010. [cited at p. 33 and 36]

[8] A. Arenas, A. Cabrales, L. Danon, A. Díaz-Guilera, R. Guimerà, and F. Vega-Redondo. Optimal information transmission in organizations: search and congestion. *Rev. Econ. Des.*, pages 1–19, 2004. [cited at p. 62]

[9] A. Arenas, A. Díaz-Guilera, and C.J. Pérez-Vicente. Synchronization reveals topological scales in complex networks. *Physical Review Letters*, 96(11):114102, 2006. [cited at p. 30]

[10] A. Arenas, J. Duch, A. Fernández, and S. Gómez. Size reduction of complex networks preserving modularity. *New Journal of Physics*, 9:176, 2007. [cited at p. 30]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[11] A. Arenas, A. Fernandez, and S. Gomez. Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics*, 10(053039):053039, 2008. [cited at p. 29, 30, 62 and 63]

[12] A. Arenas, A. Fernández, and S. Gómez. A complex network approach to the determination of functional groups in the neural system of c. elegans. *Lectures Notes in Computer Science*, 5151:9–18, 2008. [cited at p. 30]

[13] A.D. Baddeley. *Human memory: Theory and practice*. Allyn & Bacon, Boston, MA, 1990. [cited at p. 10 and 65]

[14] A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999. [cited at p. 23]

[15] A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels. Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics*, page P06014, 2006. [cited at p. 64]

[16] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, New York, NY, USA, 2008. [cited at p. 14]

[17] M.W. Berry, S.T. Dumais, and G.W. O'Brien. Using linear algebra for intelligent information retrieval. *SIAM review*, 37(4):573–595, 1995. [cited at p. 34]

[18] J.R. Binder, R.H. Desai, W.W. Graves, and L.L. Conant. Where is the semantic system? *Cerebral Cortex*, 120:2767–2796, 2009. [cited at p. 111]

[19] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003. [cited at p. 8]

[20] V.D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM review*, 46(4):647–666, 2004. [cited at p. 70]

[21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.U. Hwang. Complex networks: structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006. [cited at p. 14, 28 and 66]

[22] M. Boguña and M.A. Serrano. Generalized percolation in random directed networks. *Physical Review E*, 72, 2005. [cited at p. 85 and 86]

[23] B. Bollobás. *Random Graphs*. Academic Press, New York, 2nd edition, 2001. [cited at p. 21]

[24] P. Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972. [cited at p. 18]

[25] J. Borge-Holthoefer and A. Arenas. Navigating word association norms to extract semantic information. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2009. [cited at p. 85]

[26] J. Borge-Holthoefer and A. Arenas. Categorizing words through semantic memory navigation. *European Physical Journal B*, 74(2), 2010. [cited at p. 65, 92 and 105]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[27] J. Borge-Holthoefer and A. Arenas. Semantic Networks: Structure and Dynamics. *Entropy*, 12(5):1264–1302, 2010. [cited at p. 85 and 92]

[28] W.A. Bousfield and C.H.W. Sedgewick. An analysis of sequences of restricted associative responses. *J. Gen. Psychol.*, 30:149–165, 1944. [cited at p. 62]

[29] U. Brandes. A faster algorithm for betweenness centrality. *J. Math. Soc.*, 25(2):163–177, 2001. [cited at p. 18]

[30] R. Bro, E. Acar, and T.G. Kolda. Resolving the sign ambiguity in the singular value decomposition. *Journal of Chemometrics*, 22(2):135–140, 2008. [cited at p. 38]

[31] A. Budanitsky and G. Hirst. Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32(1):13–47, 2006. [cited at p. 65]

[32] G. Caldarelli. *Scale-free networks: Complex webs in nature and technology*. Oxford University Press, USA, 2007. [cited at p. 14]

[33] A. Callejas, A. Correa, J. Lupiáñez, and P. Tudela. Normas asociativas intracategoriales para 612 palabras de seis categorías semánticas en español. *Psicológica*, 24:185–214, 2003. [cited at p. 48 and 51]

[34] A. Caramazza and B.Z. Mahon. The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, 7(8):354–361, 2003. [cited at p. 92]

[35] H. Chertkow, D. Bub, and M. Seidenberg. Priming and semantic memory loss in Alzheimer's disease. *Brain and Language*, 36(3):420–446, 1989. [cited at p. 93 and 105]

[36] M.T.C. Chu and G.H. Golub. *Inverse eigenvalue problems: theory, algorithms, and applications*. Oxford University Press, USA, 2005. [cited at p. 34 and 38]

[37] F.R.K. Chung and L. Lu. *Complex graphs and networks*. American Mathematical Society, 2006. [cited at p. 14]

[38] A. Church. A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1:40–41, 1936. [cited at p. 8]

[39] R. Cohen and S. Havlin. *Complex Networks: Structure, Robustness and Function*. Cambridge University Press, New York, NY, USA, 2010. [cited at p. 14]

[40] A. M. Collins and M. R. Quillian. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8:240–247, 1969. [cited at p. 10, 56, 85 and 105]

[41] A.M. Collins and E.F. Loftus. A spreading activation theory of semantic memory. *Psycholical Review*, 82:407–428, 1975. [cited at p. 10, 56 and 85]

[42] P. Collins and P. Peters. The australian corpus project. In *Corpus linguistics, hard and soft: proceedings of the Eighth International Conference on English Language Research on Computerized Corpora*, page 103. Editions Rodopi, 1988. [cited at p. 45]

[43] B. Corominas-Murtra, S. Valverde, and R. V. Solé. The ontogeny of scale-free syntax networks through language acquisition. *Advances in Complex Systems*, 12:371–392, 2009. [cited at p. 44]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[44] B. Corominas-Murtra, S. Valverde, and R. V. Solé. *Evolution of Communication and Language in Embodied Agents*, pages 83–101. Springer, 2010. [cited at p. 44]

[45] L.F. Costa and G. Travieso. Exploring complex networks through random walks. *Physical Review E*, 75(1):16102, 2007. [cited at p. 67]

[46] TB Cumming, KS Graham, and K. Patterson. Repetition priming and hyperpriming in semantic dementia. *Brain and language*, 98(2):221–234, 2006. [cited at p. 93]

[47] F.J. Damerau. A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 1964. [cited at p. 73]

[48] L. Danon, A. Arenas, and A. Díaz-Guilera. Impact of community structure on information transfer. *Physical Review E*, 77(3):36103, 2008. [cited at p. 62]

[49] L. Danon, J. Duch, A. Arenas, and A. Díaz-Guilera. Comparing community structure identification. *Journal of Statistical Mechanics*, page P09008, 2005. [cited at p. 30 and 67]

[50] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990. [cited at p. 71]

[51] SN Dorogovtsev, AV Goltsev, and JFF Mendes. Critical phenomena in complex networks. *Review of Modern Physics*, 80(4):1275–1335, 2008. [cited at p. 14]

[52] S.N. Dorogovtsev and J.F.F. Mendes. Language as an evolving word web. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 268(1485):2603–2606, 2001. [cited at p. 50]

[53] J. Duch and A. Arenas. Community detection in complex networks using extremal optimization. *Physical Review E*, 72, 2005. [cited at p. 30, 62, 63 and 67]

[54] P. Erdös and A. Rényi. On random graphs. *Publ. Math. (Debrecen)*, 6:290–297, 1959. [cited at p. 21]

[55] P. Erdös and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17, 1960. [cited at p. 21]

[56] P. Erdös and A. Rényi. On the evolution of random graphs. *Bull. Inst. Int. Stat.*, 38:343, 1961. [cited at p. 21]

[57] C. Fellbaum, editor. *WordNet: An electronic lexical database*. MIT press Cambridge, MA, 1998. [cited at p. 65]

[58] A. Fernández, E. Díez, M.A. Alonso, and M. Beato. Free-association norms for the spanish names of the snodgrass and vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, 36(3):577–583, 2004. [cited at p. 48 and 51]

[59] L. Ferrand and F.-X. Alario. Normes d'associations verbales pour 366 noms d'objets concrets. *L'Année Psychologique*, 98:531–552, 1998. [cited at p. 48]

[60] R. Ferrer i Cancho. Euclidean distance between syntactically linked words. *Physical Review E*, 70(5):56135, 2004. [cited at p. 44]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[61] R. Ferrer i Cancho. Zipf's law from a communicative phase transition. *European Physical Journal B*, 47(3):449–457, 2005. [cited at p. 52]

[62] R. Ferrer i Cancho. When language breaks into pieces. a conflict between communication through isolated signals and language. *BioSystems*, 84:242–253, 2006. [cited at p. 52 and 53]

[63] R. Ferrer i Cancho and R. V. Solé. The small-world of human language. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 268:2261–2265, 2001. [cited at p. 45, 51, 52 and 53]

[64] R. Ferrer i Cancho and R. V. Solé. Patterns in syntactic dependency networks. *Physical Review E*, 69, 2004. [cited at p. 44]

[65] R.F. Ferrer i Cancho. Why do syntactic links not cross? *Europhysics Letters*, 76(6):1228–1234, 2006. [cited at p. 44]

[66] R.F. Ferrer i Cancho and R.V. Solé. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3):788, 2003. [cited at p. 52]

[67] L. Freeman. A set of measures of centrality based upon betweenness. *Sociometry*, 40, 1977. [cited at p. 18]

[68] A.D. Friederici. Pathways to language: fiber tracts in the human brain. *Trends in Cognitive Sciences*, 13(4):175–181, 2009. [cited at p. 111]

[69] H. Gardner. *The Mind's New Science*. Basic Books, New York, 1985. [cited at p. 8]

[70] B. Giffard, B. Desgranges, F. Nore-Mary, C. Lalevee, H. Beaunieux, V. De La Sayette, F. Pasquier, and F. Eustache. The dynamic time course of semantic memory impairment in Alzheimer's disease: clues from hyperpriming and hypopriming effects. *Brain*, 125(9):2044, 2002. [cited at p. 93, 97, 99 and 103]

[71] B. Giffard, B. Desgranges, F. Nore-Mary, C. Lalevee, V. De La Sayette, F. Pasquier, and F. Eustache. The nature of semantic memory deficits in Alzheimer's disease: new insights from hyperpriming effects. *Brain*, 124(8):1522, 2001. [cited at p. 93]

[72] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002. [cited at p. 28 and 40]

[73] P. Gleiser and L. Danon. Community structure in jazz. *Advances in Complex Systems*, 6(4):565–573, 2003. [cited at p. 29]

[74] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, MD, 1996. [cited at p. 33, 34 and 38]

[75] S. Gómez, A. Fernández, A. Arenas, and J. Borge-Holthoefer. http://deim.urv.cat/~sgomez/radatools.php. [cited at p. 30, 62 and 63]

[76] S. Gómez, P. Jensen, and A. Arenas. Analysys of community structure in networks of correlated data. *Physical Review E*, 80(016114), 2009. [cited at p. 29]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[77] J. Goñi, G. Arrondo, J. Sepulcre, I. Martincorena, N. Vélez de Mendizábal, B. Corominas-Murtra, B. Bejarano, S. Ardanza-Trevijano, H. Peraita, D.P. Wall, et al. The semantic organization of the animal category: evidence from semantic verbal fluency and network theory. *Cognitive Processing*, pages 1–14, 2010. [cited at p. 85]

[78] J. Goñi, I. Martincorena, B. Corominas-Murtra, G. Arrondo, S. Ardanza-Trevijano, and P. Villoslada. Switcher-random-walks: a cognitive inspired mechanism for network exploration. *International Journal of Bifurcation and Chaos*, 20(3), 2009. [cited at p. 62, 67 and 85]

[79] T. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(1):5228–5235, 2004. [cited at p. 8]

[80] T. Griffiths, M. Steyvers, and A. Firl. Google and the mind: Predicting fluency with pagerank. *Psychological Science*, 18:1069–1076, 2007. [cited at p. 59, 61 and 85]

[81] T. Griffiths, M. Steyvers, and J. Tenenbaum. Topics in semantic representation. *Psycholical Review*, 114(2):211–244, 2007. [cited at p. 8]

[82] R. Guimerà and L. A. N. Amaral. Functional cartography of complex metabolic networks. *Nature*, 433:895–900, 2005. [cited at p. 30 and 31]

[83] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68, 2003. [cited at p. 29]

[84] M.A.K. Halliday et al. Lexis as a linguistic level. *In memory of JR Firth*, pages 148–162, 1966. [cited at p. 45]

[85] J.D. Henry, J.R. Crawford, and L.H. Phillips. Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia*, 42(9):1212–1222, 2004. [cited at p. 93 and 105]

[86] A.E. Hillis and A. Caramazza. Category-specific naming and comprehension impairment: A double dissociation. *Brain*, 114(5):2081–2094, 1991. [cited at p. 111]

[87] T.T. Hills, J. Maouene, B. Riordan, and L.B. Smith. The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, 63(3):259–273, 2010. [cited at p. 85]

[88] T.T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith. Categorical structure among shared features in networks of early-learned nouns. *Cognition*, 112(3):381–396, 2009. [cited at p. 85]

[89] T.T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith. Longitudinal analysis of early semantic networks. *Psychological Science*, 20(6):729–739, 2009. [cited at p. 85]

[90] G.E. Hinton, J.L. McClelland, and D.E. Rumelhart. Distributed representations. *Parallel distributed processing*, 1(3), 1986. [cited at p. 8]

[91] K.A. Hutchison. Is semantic priming due to association strength or feature overlap? a microanalytic review. *Psychonomic Bulletin & Review*, 10(4):785, 2003. [cited at p. 10]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[92] P. Jaccard. Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:241–272, 1901. [cited at p. 70]

[93] G. Jeh and J. Widom. Simrank: a measure of structural-context similarity. In *Proceedings of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002. [cited at p. 70]

[94] M. Kaiser, R. Martin, P. Andras, and M.P. Young. Simulation of robustness against lesions of cortical networks. *European Journal of Neuroscience*, 25(10):3185–3192, 2007. [cited at p. 85]

[95] C.T. Kello and B.C. Beltz. Scale-free networks in phonological and orthographic word-form lexicons. *Approaches Phonol. Complex.*, page 171, 2009. [cited at p. 44]

[96] K. Klemm and V.M. Eguiluz. Highly clustered scale-free networks. *Physical Review E*, 65(3):36123, 2002. [cited at p. 27]

[97] K. Klemm, V.M. Eguíluz, and M. San Miguel. Scaling in the structure of directory trees in a computer cluster. *Physical Review Letters*, 95(12):128701, 2005. [cited at p. 65]

[98] M. Laisney, B. Giffard, S. Belliard, V. de la Sayette, B. Desgranges, and F. Eustache. When the zebra loses its stripes: Semantic priming in early Alzheimer's disease and semantic dementia. *Cortex*, 2009. [cited at p. 93, 95, 98, 100, 101 and 106]

[99] T. Landauer and S.T. Dumais. A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240, 1997. [cited at p. 8, 34 and 71]

[100] T. Landauer, P. W. Foltz, and D. Laham. Introduction to Latent Semantic Analysis. *Discourse Process*, 25:259–284, 1998. [cited at p. 8 and 71]

[101] T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, editors. *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates, Mahwah, N.J., 2007. [cited at p. 8]

[102] E.A. Leicht, P. Holme, and M.E.J. Newman. Vertex similarity in networks. *Physical Review E*, 73(2), 2006. [cited at p. 70]

[103] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklad.*, 10(8):707–710, 1966. [cited at p. 73]

[104] Y. Liu, M. Liang, Y. Zhou, Y. He, Y. Hao, M. Song, C. Yu, H. Liu, Z. Liu, and T. Jiang. Disrupted small-world networks in schizophrenia. *Brain*, 131(4):945, 2008. [cited at p. 85, 87 and 103]

[105] L. Lovàsz. Random walks on graphs: A survey. *Bolyai Soc. Math. Stud.*, 2:353–397, 1996. [cited at p. 69]

[106] D.I. Margolin, D.S. Pate, and F.J. Friedrich. Lexical priming by pictures and words in normal aging and in dementia of the alzheimer's type. *Brain and Language*, 54(2):275–301, 1996. [cited at p. 93]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[107] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, Henry Holt and Co*. W.H. Freeman and Company, New York, NY, 1982. [cited at p. 9]

[108] A. Martin, C.L. Wiggs, L.G. Ungerleider, and J.V. Haxby. Neural correlates of category-specific knowledge. *Nature*, 379(6566):649–652, 1996. [cited at p. 111]

[109] J.L. McClelland and D.E. Rumelhart. *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. The MIT Press, 1989. [cited at p. 8]

[110] K. McRae, G.S. Cree, M.S. Seidenberg, and C. McNorgan. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, 37(4):547–559, 2005. [cited at p. 46]

[111] A. Mehler. *Corpus Linguistics. An International Handbook*, chapter Large Text Networks as an Object of Corpus Linguistic Studies. de Gruyter, Berlin/New York, 2007. [cited at p. 44 and 46]

[112] A. Melinger and A. Weber. Database of Noun Associations for German. *http://www.coli.uni-saarland.de/projects/nag/*, 2006. [cited at p. 48 and 51]

[113] DE Meyer and RW Schvaneveldt. Facilitation in recognizing words: Evidence of a dependence upon retrieval operations. *Journal of Experimental Psychology*, 90:227–234, 1971. [cited at p. 93]

[114] S. Milgram. The small world problem. *Psychology Today*, 2, 1967. [cited at p. 13 and 22]

[115] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. Introduction to wordnet: an on-line lexical database. *Int. J. Lexico.*, 3(4):235–244, 1990. [cited at p. 43 and 46]

[116] T.M. Mitchell, S.V. Shinkareva, A. Carlson, K.M. Chang, V.L. Malave, R.A. Mason, and M.A. Just. Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191, 2008. [cited at p. 111]

[117] Y. Moreno, JB Gómez, and AF Pacheco. Instability of scale-free networks under node-breaking avalanches. *Europhysics Letters*, 58:630, 2002. [cited at p. 92]

[118] Y. Moreno, R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Critical load and congestion instabilities in scale-free networks. *Europhysics Letters*, 62:292, 2003. [cited at p. 92]

[119] HE Moss and LK Tyler. A progressive category-specific semantic deficit for non-living things. *Neuropsychologia*, 38(1):60–82, 2000. [cited at p. 92]

[120] A. E. Motter, A. P. S. de Moura, Y.-C. Lai, and P. Dasgupta. Topology of the conceptual network of language. *Physical Review E*, 65, 2002. [cited at p. 49 and 53]

[121] R.D. Nebes. Semantic memory in Alzheimer's disease. *Psychological Bulletin*, 106(3):377–394, 1989. [cited at p. 93]

[122] D. L. Nelson, C. L. McEvoy, and T. A. Schreiber. The University of South Florida word association, rhyme, and word fragment norms. *http://www.usf.edu/FreeAssociation/*, 1998. [cited at p. 47, 51, 69 and 85]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[123] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev*, 45:167–256, 2003. [cited at p. 14 and 66]

[124] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69, 2004. [cited at p. 30, 62, 63 and 67]

[125] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69, 2004. [cited at p. 29]

[126] M.E.J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002. [cited at p. 20]

[127] M.E.J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):26126, 2003. [cited at p. 20]

[128] M.E.J. Newman, A.L. Barabási, and D.J. Watts. *The structure and dynamics of networks*. Princeton University Press, 2006. [cited at p. 14]

[129] J.D. Noh and H. Rieger. Random walks on complex networks. *Physical Review letters*, 92(11):118701, 2004. [cited at p. 67]

[130] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, Standford, CA, 1998. [cited at p. 19 and 59]

[131] S.M. Polyn, V.S. Natu, J.D. Cohen, and K.A. Norman. Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(5756):1963, 2005. [cited at p. 111]

[132] P. Pons and M. Latapy. Computing communities in large networks using random walks. *Lecture Notes in Computer Science*, 3733:284, 2005. [cited at p. 83 and 84]

[133] Z.W. Pylyshyn. *Computation and cognition: Toward a foundation for cognitive science*. Massachusetts Institute of Technology Cambridge, MA, USA, 1984. [cited at p. 7]

[134] M.R. Quillian. Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, 12(5):410–430, 1967. [cited at p. 85]

[135] Cohen R., K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Physical Review Letters*, 85(21), 2000. [cited at p. 85 and 86]

[136] A. Rapoport. Spread of information through a population with socio- structural bias i. assumption of transitivity. *Bull. Math. Biophys.*, 15:523–533, 1953. [cited at p. 22]

[137] A. Rapoport. Spread of information through a population with socio- structural bias ii. various models with partial transitivity. *Bull. Math. Biophys.*, 15:533–546, 1953. [cited at p. 22]

[138] A. Rapoport. Spread of information through a population with socio- structural bias iii. suggested experimental procedures. *Bull. Math. Biophys.*, 16:75–81, 1954. [cited at p. 22]

[139] A. Rapoport and W. Horvath. A study of a large sociogram. *Behav. Sci.*, 6(2):79–91, 1961. [cited at p. 22]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[140] E. Ravasz, AL Somera, DA Mongru, ZN Oltvai, and A.L. Barabási. Hierarchical organization of modularity in metabolic networks, 2002. [cited at p. 70]

[141] R.C. Read and R.J. Wilson. *An atlas of graphs*. Oxford University Press, 1998. [cited at p. 21]

[142] P.M. Roget. *Roget's Thesaurus of English Words and Phrases*. TY Crowell co, 1911. [cited at p. 46 and 65]

[143] E. Rosch and B.B. Lloyd. *Cognition and categorization*. Erlbaum Hillsdale, NJ, 1978. [cited at p. 67]

[144] M. Rosvall and C.T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118, 2008. [cited at p. 84]

[145] H.D. Rozenfeld. *Structure and Properties of Complex Networks: Models, Dynamics, Applications*. VDM Verlag, 2010. [cited at p. 14]

[146] D.E. Rumelhart and J.L. McClelland. the pdp research group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, 1, 1986. [cited at p. 8]

[147] D.E. Rumelhart, J.L. McClelland, et al. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 2 Vols.* Bradford Book, 1987. [cited at p. 8]

[148] M. Sales-Pardo, R. Guimerà, and L. A. N. Amaral. Classes of complex networks defined by role-to-role connectivity profiles. *Nature Physics*, 3(63), 2007. [cited at p. 38]

[149] G. Salton. *Automatic text processing: the Transformation, Analysis and Retrieval of Information by Computer*. Addison-Wesley, Reading, MA, 1989. [cited at p. 70]

[150] G. Salton and M.J. McGill. *Introduction to modern information retrieval*. McGraw-Hill, Auckland, 1983. [cited at p. 70]

[151] N. Schwartz, Cohen R., D. ben Avraham, A.L. Barabási, and S. Havlin. Percolation in directed scale-free networks. *Physical Review E*, 66, 2002. [cited at p. 85 and 86]

[152] J.R. Searle. The chinese room revisited. *Behavioral and Brain Sciences*, 5:345–348, 1982. [cited at p. 7]

[153] J.R. Searle. Minds, brains, and programs. *The Turing test: verbal behavior as the hallmark of intelligence*, page 201, 2004. [cited at p. 7]

[154] M.A. Serrano and M. Boguña. Clustering in complex networks. ii. percolation properties. *Physical Review E*, 74, 2006. [cited at p. 85 and 86]

[155] C.E. Shannon and W. Weaver. A mathematical theory of communication, 1948. *Bell Syst. Tech. J.*, 27(379):623, 1948. [cited at p. 52]

[156] M. Sigman and G. Cecchi. Global organization of the wordnet lexicon. *Proceedings of the National Academy of Sciences*, 99(3):1742–1747, 2002. [cited at p. 46, 49 and 51]

[157] J. Sinclair. *Corpus, concordance, collocation*. Oxford University Press, 1991. [cited at p. 45]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[158] J.G. Snodgrass and M. Vanderwart. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp. Psychol. Hum. Learn.*, 6:174–215, 1980. [cited at p. 51]

[159] R. V. Solé, B.C. Murtra, S. Valverde, and L. Steels. Language networks: their structure, function and evolution. *Trends in Cognitive Sciences*, 2006. [cited at p. 44 and 53]

[160] R. V. Solé and S. Valverde. *Networks: Structure, Dynamics and Function*, chapter Information Theory of Complex Networks: On Evolution and Architectural Constraints, pages 169–190. Lecture Notes in Physics. Springer-Verlag, 2004. [cited at p. 20]

[161] O. Sporns, D. Chialvo, M. Kaiser, and CC. Hilgetag. Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8:418–425, 2004. [cited at p. 111]

[162] O. Sporns and R. Kötter. Motifs in brain networks. *PLoS Biol.*, 2(11), 2004. [cited at p. 111]

[163] C.J. Stam, W. De Haan, A. Daffertshofer, B.F. Jones, I. Manshanden, A.M. van Cappellen van Walsum, T. Montez, J.P.A. Verbunt, JC de Munck, B.W. van Dijk, et al. Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. *Brain*, 132(1):213–224, 2009. [cited at p. 85, 87 and 103]

[164] CJ Stam, BF Jones, G. Nolte, M. Breakspear, and P. Scheltens. Small-world networks and functional connectivity in Azheimer's disease. *Cerebral Cortex*, 17(1):92–99, 2007. [cited at p. 85, 87 and 103]

[165] L. Steels. Language as a complex adaptive system. *Lecture Notes in Computer Science*, pages 17–28, 2000. [cited at p. 64]

[166] L. Steels. The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Conn. Sci.*, 17(3):213–230, 2005. [cited at p. 64]

[167] L. Steels and T. Belpaeme. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, 28(04):469–489, 2005. [cited at p. 8]

[168] L. Steels and R.A. Brooks. *The artificial life route to artificial intelligence: Building embodied, situated agents*. Lawrence Erlbaum Associates, 1995. [cited at p. 8]

[169] L. Steels and F. Kaplan. Aibo's first words: The social learning of language and meaning. *Evol. Comm.*, 4(1):3–32, 2001. [cited at p. 8]

[170] M. Steyvers and J. B. Tenenbaum. The largescale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78, 2005. [cited at p. 53, 56 and 85]

[171] S.H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001. [cited at p. 13]

[172] M. Stubbs. *Text and corpus analysis*. Blackwell, 1996. [cited at p. 45]

[173] M. Stubbs. *Words and phrases: Corpus studies of lexical semantics*. Blackwell Pub, 2001. [cited at p. 45]

UNIVERSITAT ROVIRA I VIRGILI
COMPLEX NETWORKS THEORY AND ITS APPLICATION TO LANGUAGE.
Javier Borge Holthoefer
ISBN:978-84-694-2176-5/DL:T. 1034-2011

[174] K. Supekar, V. Menon, D. Rubin, M. Musen, and M.D. Greicius. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Computational Biology*, 4(6), 2008. [cited at p. 85 and 103]

[175] Moby Thesaurus. ftp://ibiblio.org/pub/docs/books/gutenberg/etext02/mthes10.zip. [cited at p. 49]

[176] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969. [cited at p. 13 and 22]

[177] A.K. Troyer, M. Moscovitch, and G. Winocur. Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsycholgy*, 11:138–146, 1997. [cited at p. 62]

[178] A.M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42:230–265, 1936. [cited at p. 8]

[179] A.M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950. [cited at p. 8]

[180] M. Vitevitch. The clustering coefficient of phonological neighborhoods influences spoken word recognition. *J Acoust. Soc. Am.*, 120:3252, 2006. [cited at p. 44]

[181] P. Vogt. Language evolution and robotics: Issues in symbol grounding and language acquisition. *Artificial cognition systems. Idea Group*, 2006. [cited at p. 8]

[182] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994. [cited at p. 17]

[183] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440, 1998. [cited at p. 13, 17 and 22]

[184] D. J. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University, Princeton, NJ, 1999. [cited at p. 23]

[185] S.J. Yang. Exploring complex networks by walking on them. *Physical Review E*, 71(1):16107, 2005. [cited at p. 67]

[186] C. Zhou, L. Zemanová, G. Zamora-López, G., C.C. Hilgetag, and J. Kurths. Structure–function relationship in complex brain networks expressed by hierarchical synchronization. *New Journal of Physics*, 9:178, 2007. [cited at p. 111]

[187] C.S Zhou, L. Zemanová, G. Zamora-López, C.C. Hilgetag, and J. Kurths. Hierarchical organization unveiled by functional connectivity in complex brain networks. *Physical Review Letters*, 97, 2006. [cited at p. 111]

[188] H. Zipf. *Human behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Addison Wesley, Cambridge, Mass., 1949. [cited at p. 52]